**Title**

# Olfactive stimulation interventions for managing procedural pain in preterm and full-term neonates: A systematic review and meta-analysis

**Authors**

**Gwenaëlle De Clifford-Faugère**, RN, M.Sc., PhD(c) [1,2,3]
*Affiliations*:
[1]Faculty of Nursing, Université de Montréal, Montreal, Quebec, Canada
[2]CHU Sainte-Justine Research Centre, Montreal, Quebec, Canada
[3]Faculté des Sciences Médicales et Paramédicales, Aix Marseille Université, EA3279-CEReSS, Marseille, France
*Institutional address:* C.P. 6128 Succ. Centre-ville, Montreal, Canada, H3C 3J7
*(514) 343*-6111 poste 51473
*Email:* gwenaelle.de.clifford@umontreal.ca

**Andréane Lavallée**, RN, Ph.D. (c) [1,2]
*Affiliations*:
[1]Faculty of Nursing, Université de Montréal, Montreal, Quebec, Canada
[2]CHU Sainte-Justine Research Centre, Montreal, Quebec, Canada
*Institutional address:* C.P. 6128 Succ. Centre-ville, Montreal, Canada, H3C 3J7
*Email:* andreane.lavallee@umontreal.ca

**Christelle Khadra**, RN, Ph.D. (c) [1,2]
*Affiliations*:
[1]Faculty of Nursing, Université de Montréal, Montreal, Quebec, Canada
[2]CHU Sainte-Justine Research Centre, Montreal, Quebec, Canada
*Institutional address:* C.P. 6128 Succ. Centre-ville, Montreal, Canada, H3C 3J7
*Email:* christelle.khadra@umontreal.ca

**Ariane Ballard**, RN, Ph.D. (c) [1,2]
*Affiliations*:
[1]Faculty of Nursing, Université de Montréal, Montreal, Quebec, Canada
[2]CHU Sainte-Justine Research Centre, Montreal, Quebec, Canada
*Institutional address:* C.P. 6128 Succ. Centre-ville, Montreal, Canada, H3C 3J7
*Email:* ariane.ballard@umontreal.ca

**Sébastien Colson**, RN, Ph.D [3,4]
*Affiliations*:

[3]Faculté des Sciences Médicales et Paramédicales, Aix Marseille Université, EA3279-CEReSS, Marseille, France
[4]École des sciences infirmières, Faculté des Sciences Médicales et Paramédicales, Aix Marseille Université, Marseille, France
*Institutional address:* [4]École des sciences infirmières, Faculté des Sciences Médicales et Paramédicales, 27 Boulevard Jean Moulin, 13385 Marseille Cedex 5, France
*Email:* sebastien.colson@univ-amu.fr

**Marilyn Aita**, RN, Ph.D [1,2].
*Affiliations*:
[1]Faculty of Nursing, Université de Montréal, Montreal, Quebec, Canada
[2]CHU Sainte-Justine Research Centre, Montreal, Quebec, Canada
*Institutional address:* C.P. 6128 Succ. Centre-ville, Montreal, Canada, H3C 3J7
*Email:* marilyn.aita@umontreal.ca

**Corresponding author**
**Gwenaëlle De Clifford-Faugère**, RN, M.Sc.
Faculty of Nursing, Université de Montréal
C.P. 6128 Succ. Centre-ville,
Montreal, Canada, H3C 3J7
*(514) 343*-6111 poste 51473
Email : gwenaelle.de.clifford@umontreal.ca
Gwenaelle.De-Clifford-Faugere@chuv.ch

**Conflict of interest**
All authors have no conflict of interest to declare.

**Title:** Olfactive stimulation interventions for managing procedural pain in preterm and full-term neonates: A systematic review and meta-analysis


## Abstract

**Background.** Preterm and full-term neonates undergo many painful procedures during their hospitalization in the neonatal intensive care unit. Unrelieved and repeated pain can have important repercussions on their motor and intellectual development. Still, pain management interventions are limited for neonates.

**Objective.** This systematic review aimed to evaluate the effectiveness of olfactive stimulation interventions on the pain response of preterm and full-term infants during painful procedures.

**Design.** Systematic review and meta-analysis.

**Data sources.** An electronic search was conducted from inception to August 2019 in PubMed, MEDLINE, Embase, CINAHL, PsycINFO, Web of Sciences, CENTRAL, Scopus and ProQuest.

**Review methods.** Study selection, data extraction, assessment of risk of bias and quality of evidence were performed by two independent reviewers.

**Results.** 3311 studies were screened. Of the 14 studiesincluded studies (n=1028 infants), results from 10 were combined in meta-analysis. The latter demonstrated that olfactive stimulation interventions using a familiar odor were effective compared to standard care on pain reactivity (SMD -0.69; 95% CI -0.93 to -0.44; $I^2$ = 20%, $p$<0.00001), pain regulation (SMD -0.40; 95% CI -0.66 to -0.14; $I^2$ = 13%, $p$= 0.002), crying duration during (SMD -0.42; 95% CI -0.73 to -0.10; $I^2$ = 47%, $p$= 0.009) and after the procedure (SMD -0.37; 95% CI -0.68 to -0.07; $I^2$ = 0%, $p$= 0.01), heart rate after the procedure (MD -3.87; 95% CI -7.36 to -0.38; $I^2$ = 99%, $p$=0.03), oxygen saturation during (MD -0.47; 95% CI -0.86 to -0.08; $I^2$=91%, $p$=0.02) and after the procedure (MD -0.56; 95% CI -0.99 to -0.13; $I^2$=99%, $p$=0.01). No adverse event was reported.

**Conclusion.** These findings are based on low to very low quality of evidence limiting our confidence in effect estimates. More rigorous trials with a larger sample size are needed to enhance the comprehension of the mechanisms underlying olfactive stimulation interventions and the interventions' efficacy.

**Systematic review registration**

The study protocol was previously registered in the International Prospective Register of Systematic Reviews (PROSPERO) on 24.02.2017 (registration: CRD42017058021) and published in the *BMC Systematic review* journal.

**What is already known about the topic?**

- Repeated and untreated pain leads to important consequences in preterm and full-term infants' development.
- Pain management interventions are very limited with this population.

**What this paper adds**

- This meta-analysis provides evidence that olfactive stimulation interventions using a familiar odor were effective on pain reactivity, pain regulation, crying duration, heart rate and oxygen saturation variations.
- A familiar odor, either natural such as mother milk odor or artificial with a habituation period, could be used in clinical practice to improve pain management.
- Further research with a larger sample sizes is needed to clarify which modality of olfactive stimulation interventions could be the most effective for pain management.

**Keywords**

Pain, odor, non-pharmacological intervention, neonatology, systematic review

## BACKGROUND

In the past decades, it has been recognized that preterm and full-term infants can feel pain and express it with non-verbal specific signs.[1] In the neonatal intensive care unit, each infant undergoes between 7.5 and 17.3 painful procedures per day such as heel pricks, endotracheal suction and venipunctures.[2]

Repeated and untreated pain may engender short-term and long-term consequences for preterm and full-term infants.[3-11] Indeed, the number of painful procedures experienced by preterm infants' from birth to 40 weeks of gestation isnegatively correlated with lower intellectual and motor development at 8 and 18 months of corrected age.[3] Moreover, a hypersensitivity to pain can still be present at 7 years of age in preterm infants.[4] This hypersensitivity to pain is also found in full-term infants who have experienced five or more painful skin breaking procedures in their first two days of life.[5]

Pharmacological pain management interventions may be used for procedural pain management.[12] However, limited evidence exists to support their long-term impacts and safety,[12] so non-pharmacological interventions are an interesting alternative. Systematic reviews have reported that effective non-pharmacological interventions for preterm infants are sucrose administration before the painful procedure, combined or not combined with non-nutritive sucking,[13] or skin-to-skin contact.[14] For full-term infants, sucrose administration before the painful procedure, which is more effective if combined to non-nutritive sucking,[13] breastfeeding,[15] and skin-to-skin contact[14] are considered to be effective interventions. However, these interventions are not always applicable in the clinical practice since breastfeeding requires the mother's presence, and skin-to-skin contact requires one of the two parents. In clinical practice, procedural pain management is still sub-optimal.[16,17] Thus, it is important to consider innovative interventions that could be used anytime during the infant's hospitalization, with or without the presence of parents. As such, olfactive stimulation interventions could be applicable and easily implemented in practice with minimal preparation and without requiring any parent's presence. In addition, the olfactive stimulation could be easily combined with other effective interventions, such as sucrose, non-nutritive sucking and skin-to-skin contact as recommended for an optimal pain management[12]. To date, little is known about the efficacy of olfactive stimulation interventions for procedural pain management in infants.

An olfactive stimulation intervention consists of exposing the infant to an odor during a painful procedure. The odor could either be natural, such as the odor of their mother's

breastmilk, or artificial, such as vanilla. The infants can be exposed to the odor using different ways, such as a gauze placed near the infant's nose or through an odor diffuser.

How olfactive stimulation interventions might work to relieve pain is an important question to address considering that the foetus' first functional olfactive cells appear as early as 11 weeks of gestation.[18,19] The olfactory system starts to be functional at 28 weeks of gestation and then the infants are able to detect, distinguish and recognize a specific odor.[18] Full-term infants can also express odor preferences when they are exposed to different smells by showing attraction or repulsion to some odors.[20,21] In the literature, the underlying mechanisms of action of olfactive stimulation interventions are not clearly identified. To date, we know that skin-to-skin contact is effective for pain management in preterm and full-term infants.[14] Skin-to-skin contact has multiple dimensions such as closeness with the mother, touch and skin contact, as well as auditory and olfactory stimulations. So, while in skin-to-skin contact, infants can smell their own mother's odor. It is the same during breastfeeding, which is also an effective intervention for full-term infants.[15] As these two interventions have an olfactory dimension that could coutribute their efficacy, it suggests that an olfactive stimulation intervention could manage pain in infants. The possible efficacy of olfactive stimulation could be explained by its impact on the emotional and affective pain components or by its distraction component.[22]

To our knowledge, only one recent systematic review assessed the efficacy of olfactive stimulation interventions for procedural pain management of infants.[23] However, this review only focuses on the maternal milk odor. Since other types of odors can be used for olfactive stimulation and pain management in infants, amore comprehensive systematic review is needed to increase our understanding of the efficacy of this type of intervention. Conducting this review will provide better knowledge on the effectiveness of olfactive stimulation interventions in the pain management of preterm and full-term infants and thus prevent its long-term consequences.

**Objective**

The objective of this systematic review is to evaluate the effectiveness of olfactive stimulation interventions on preterm and full-term infants' pain response and other secondary outcomes (crying duration, variations in heart rate, oxygen saturation and cortisol levels as well as the occurrence of adverse events) during a painful procedure compared to standard care.

## MATERIAL and METHODS

### Protocol and registration

We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) recommendations to conduct this systematic review.[24,25] The protocol of this systematic review has been previously published[22] and registered in the PROSPERO database (CRD42017058021). Following are the minor modifications that has been made to the present manuscript compared to the published protocol:

1) In addition to the pre-specified outcomes, we included four secondary outcomes that were subsequently identified in the included studies: crying duration, cortisol levels, and physiological parameters variations, including both heart rate and oxygen saturation;

2) For the search strategy, we did not manually search neonatology or pain journals for relevant studies as the related published articles had already been captured by the electronical search in the various databases.

### Eligibility Criteria

Randomized controlled trials (RCTs) and quasi-experimental studies were considered in this systematic review. However, quasi-experimental studies were included only in the qualitative synthesis and were excluded from the meta-analysis. This systematic review focus only on studies comparing olfactive stimulation interventions (all types of odors, whether natural (e.g. milk), or artificial (e.g. vanilla) and all methods (e.g. gaze, diffuser) of delivering the odor) to one or more comparator groups, on pain response of preterm infants (< than 37 weeks of gestation) and full-term infants undergoing one of the ten most frequent painful procedure.[17] The painful procedures considered were as follow[17]: (1) nasal aspiration, (2) tracheal aspiration, (3) heel prick, (4) adhesive removal, (5) gastric tube insertion, (6) venipuncture, (7) arterial puncture, (8) installation of peripheral intravenous cannula, (9) chest physiotherapy and (10) removal of peripheral intravenous line. Studies examining the effectiveness of multisensory interventions were excluded. Comparators were either a placebo (sterile water) or standard care (any care or pain management intervention carried out in the clinical setting). The primary outcome was pain response, measured by standardized scales at one or more of the two following timepoints: at the beginning of the procedure, such as needle insertion (pain reactivity) or immediately after the procedure (pain regulation).

### Search methods for identification of studies

The electronic search was performed from inception to August 2019 in the following databases: PubMed, MEDLINE, Embase, CINAHL, PsycINFO, Web of Science, the Cochrane Central Register of Controlled Trials (CENTRAL), Scopus and ProQuest. The search strategy was developed with an allied health sciences librarian and adapted for each of the nine databases (see Table S1 supplemental digital). The search was restricted to French and English articles. Three grey literature databases were used: http://www.opengrey.eu/, http://opengrey.org/ and www.greylit.org. Ongoing or unpublished clinical trials were searched in registration sites, including ClinicalTrials.gov (http://www.ClinicalTrials.gov) and the World Health Organization International Clinical Trials Registry Platform (http://www.who.int/ictrp/en/). Only one study was identified (NCT03626974) through these registration sites and the authors were successfully contacted, but the trial could not be included as the recruitment was still ongoing. Conference abstracts were searched in Biosis and Biological abstract without any time restriction, but none met the inclusion criteria of this systematic review. In addition, the reference lists of included studies were manually checked.

**Data collection and analysis**

**Study selection.** All citations were saved in a bibliographic management software (EndNote© X7) and duplicates were removed. Firstly, titles and abstracts were screened by two independent reviewers (AL, DG) according to the prespecified inclusion criteria. Secondly, full-texts of previously selected studies were reviewed independently by the same reviewers (AL, GD). Discrepancies were resolved by consensus and a third reviewer (MA) was consulted in case of disagreement. Reasons for study exclusion were documented.

**Data extraction and management.** According to the Cochrane Handbook recommendations,[26] we developed a specific data extraction form which was pilot-tested with three studies and then refined. Two reviewers (AL, GD) extracted data independently from all the included studies. The extracted data were subsequently compared to ensure consistency. A third researcher (MA) was consulted in case of disagreement. Details regarding the information extracted from each study are specified in the study protocol.[22] Extracted data was then entered in the Review Manager (RevMan) software (version 5.1 Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2014). To avoid errors in quantitative analysis, the data were double-checked by two reviewers (AL, GD) before conducting the analysis.

**Types of outcome measures**. The primary outcome was pain response divided by time of measurement, as suggested by Pillai Riddell:[27,28] 1) pain reactivity, and 2) pain regulation.

Pain reactivity corresponds to a measure of pain during the painful event. For example, for a needle-related procedure, pain reactivity starts at the moment of the needle insertion and lasts until the end of the blood sample. Pain regulation is the measure of pain immediately following the end of the painful procedure. It refers to the ability of the infants to regulate their pain and stress related to the procedure (autoregulation system[29]), which results in less pain immediately after the painful procedure. In other words, it is the ability of the preterm infant to regain a similar state as of before the painful procedure. Pain response had to be evaluated by a standardized validated tool (multidimensional measures), such as the Premature Infant Pain Profile[30,31] (PIPP) suitable for 28 to 40 weeks of gestation (WG) infants, the Neonatal Infant Pain Scale[32] (NIPS) for 26 to 47 WG infants, the Neonatal Facial Coding System[33] (NFCS) for those from 26 to 47 WG, and the *Douleur aiguë du nouveau-né*[34] [acute pain of the newborn] (DAN) from 24 to 41 WG infants. These standardized tools are considered as a multidimensional measure of pain as they included pain behavioural and physiological indicators. As others Cochrane systematic reviews,[27, 66] we combined different tools in the meta-analysis for pain reactivity and pain regulation because all these tools measure similar parameters (named "all pain scores"). Secondary outcomes were unidimensional measures: 1) crying duration (pain behavioural indicator) at two timepoints : during and after the painful procedure, 2) variations in heart rate (pain physiological indicator) at two timepoints: during and after the painful procedure, 3) variations in oxygen saturation (pain physiological indicator) at two timepoints: during and after the painful procedure, 4) variations in cortisol levels between before and after the painful procedure (pain hormonal indicator); 5) occurrence of adverse events.

**Assessment of risk of bias in included studies**

Risk of bias of each included study was independently assessed by two reviewers (AL, GD) using the Cochrane Risk of Bias Assessment Tool.[35] Considering the nature of the olfactive stimulation interventions, we always considered that the risk of group contamination by odor dissemination was another potential risk of bias. A third reviewer (CK) was consulted in case of disagreement and for validation. Each of these sources of bias were rated in terms of unclear risk, low risk and high risk. The overall risk of bias of each included study was taken in consideration for the analysis and the interpretation of the results.

**Quality of Evidence**

The assessment of the quality of evidence across studies was performed using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) working group methodology and was ranked as high, medium, low or very low, by three reviewers (AB, AL and GD).[36] We assessed the quality of evidence for each outcome in accordance with the five domains: risk of bias, inconsistency, indirectness, imprecision and publication bias.[26,37,38] For each domain, we downgraded one or two points according to the following judgment criteria. For the risk of bias, the quality of evidence was not downgraded if the majority of risk of biases judgements were rated as "low", downgraded by one point if the majority of risk of biases judgements were rated as "unclear" or "high" and downgraded by two points if the majority of risk of biases judgements were rated “high”. For the inconsistency, the quality of evidence was not downgraded if the heterogeneity was considered as not important (<40%), downgraded by one point if there was moderate or substantial heterogeneity among studies (40% to 75%) and downgraded by two points if there was considerable heterogeneity among studies (75% to 100%). For the indirectness of the evidence, we did not downgrade for any outcome as we were confident in the four sources of indirectness (population, intervention, outcomes measures and indirect comparisons). For the imprecision, the quality of evidence was downgraded by one point if the total number of participants was less than 400 for the assessed outcome and downgraded by two points if the total number of participants was less than 150. For publication bias, we did not downgrade for any outcome. We generated the “Summary of findings” tables using the GRADE profiler Guideline Development Tool software and the GRADE criteria (2015, McMaster University and Evidence Prime Inc.).

**Summary Measures**

We conducted statistical analyses using the Review Manager 5.1 software. All effect estimates were reported using a random-effect model with a 95% confidence interval (CI). All outcomes were continuous data. Standard deviations (SD) and standardized mean differences (SMD) were used as different scales were used across included study. Data were analyzed using the mean difference (MD) instead of the SMD when there was no inconsistency between scales or when no scales were used. When a trial provided two treatment arms for one meta-analysis, we split the control group’s sample size “n” in two to avoid a “double-counting” bias.[39] The unit of analysis was the infants receiving the olfactive stimulation intervention or standard care. There were no cross-over trials and cluster randomized trials included. Interpretation of the effect sizes are based on Cohens’d classification: 0.2 represents a small effect, 0.5 a moderate effect and 0.8 a large effect. .[39,67]

**Planned methods of analysis**

We conducted a meta-analysis with a random effects model and inverse variance when at least two studies were available for one outcome. A random effects model was chosen because it considers the within-study variability as well as the between study variability.[24,25] Data were summarized by a descriptive synthesis if it was not feasible to perform a meta-analysis (less than two studies, unvalidated tool used for the population, clinical heterogeneity, missing data). Heterogeneity was statistically verified by using the chi-square test ($x^2$), with a p-value < 0.1 considered for statistical significance[40], and by a $I^2$ statistics. We interpreted the $I^2$ according to the Cochrane Handbook recommendation[40]: 0-40% : not an important heterogeneity; 30-60%: moderate heterogeneity; 50-90%: substantial heterogeneity; and 75-100%: considerable heterogeneity. Main results included preterm and full-term infants.Subgroup analysis were conducted for each outcome when possible according to the types of odor used for the olfactive stimulation intervention (i.e. mother's milk odor, amniotic fluid odor) and by population (preterm and full-term infants) in order to guide further clinical recommendations and to use all data available as several RCTs have more than two arms.

**Missing data**

Five authors were contacted for missing data in six studies.[41-46] After many attempts, we did not obtain the data for three of these studies.[44-46] When data were not available and data imputation was not possible,[39] studies were excluded from the meta-analysis and results were reported in the qualitative synthesis.

## RESULTS

**Study selection**

The study selection process is illustrated by the PRISMA Study Flow Diagram (Figure 1). Of the 3311 screened studies, a total of 1852 studies remained after removing 1459 duplicates. We excluded 1817 studies during the first selection (title and abstract) and 38 studies were assessed for eligibility in a final selection (full-texts). Among these, twenty-three full-text articles were excluded for the following reasons: pain was not assessed (n=8), no intervention was provided (n=6), it was a multisensorial intervention (n=5), and the painful procedure was not eligible (n=4). Finally, a total of 14 studies met the inclusion criteria and were included in the qualitative synthesis of this systematic review.[41-55] Moreover, three studies were not included in the meta-analysis, because data was not available even after contacting the authors,[44-46] and one study was not considered because of its quasi-

experimental design.[53] Therefore, the quantitative synthesis included a total of ten studies.[41-43,47-51,54,55]

**Characteristics of included studies**

The 14 included studies involved a total of 1028 preterm and full-term infants. Two papers referred to the same study but they reported different outcomes.[51,52] Among the included studies, five were conducted with preterm infants ,[41,48-52] and nine with full-term infants [42-47,53-55] (see Table 1). The 14 studies were published between 1997 and 2007. Four studies were conducted in Iran,[49,51,54,55] two in the United States of America,[41,42] two in Japan,[45,46] two in Turkey,[47,48] two in France,[50,53] one in Switzerland,[43] and one in Estonia.[44] Designs of these studies consist of five RCTs with two arms,[44,46,49,50,54] one RCT with three arms,[51] six RCTs with four arms,[42,43,45,47,48,55] one RCT with six arms[41] and one quasi-experimental design with three arms.[53]

Included studies evaluated different painful procedures consisting of heel-pricks (n=8),[42-49] venipunctures (n=4),[50,51,53,54], both heel-pricks and venipunctures (n=1),[41] and arterial punctures (n=1).[55]

The olfactive stimulation interventions focused on different odors, either natural or artificial. More than half of the studies considered two odors or more.[43,45-49,51,53] Eight studies considered natural odors, including newborn own mother's breast milk,[43,45,47-51,53] amniotic fluid odor,[47,48] mother’s odor,[48] and breast milk of other mothers.[45] Five studies were conducted using natural and artificial odors in separate groups.[43,45,49,51,53] The twelve studies on artificial odors considered vanilla odor,[41-44,51,55] lavender odor,[46,47,54] and formula milk.[45,46,49,53]

Different methods of delivering the odor were used in the studies. In four studies a gaze was used, placed at either 10 centimetres,[54] 7 centimetres,[44] 1 centimetres,[55]from the infant’s nose or at a non-indicated distance.[53] In three studies a cotton pad immersed with the odor was used, placed near the infant's nose, [42,43] only specified in Jebreilli et al.’s study[51] at one millimeter. Two studies used an odor diffuser,[45,50] while two other studies used a flask containing the odor deposited inside.[46,47] Four studies administered the odors using different means including a filter paper,[49] a scarf ,[41] a doll,[48] or the mother's chest (vanilla).[42]

Odor habituation, which refers to the preliminary exposure to an odor, was performed in six studies[41-43,51,54,55] for a period of between 8 hours to 18 hours prior to the painful procedure to improve the effectiveness of the olfactory stimulation intervention. However, this habituation was only performed for artificial lavender odors[54] or vanilla odors.[41-43,51,55] In

this review, we considered that an odor was familiar to the infant when it is a natural odor such as mother's breast milk or an artificial odor with an habituation.

Standard care were described in three studies as non-nutritive sucking,[50] swaddling,[46] touch and talk softly.[55] In three studies, standard care was no interventions for pain management.[44,48,54] No information was provided on pain management in the control group in the others studies.[41-43,45,47,51,53]

Various outcomes were measured in the included studies and most of them considered more than one outcome. Pain, the primary outcome of this review, was assessed by a reliable and valid pain scale in seven studies: four chose the Premature Infant Pain Profile,[48-51] two the *Douleur aiguë du nouveau-né*,[50,54] one the Neonatal Infant Pain Scale,[47] and one the Neonatal Facial Coding Scale.[44] Crying duration was measured in 11 studies.[41-45,48-50,53-55] Physiological parameters were measured during the painful procedure in three different studies: heart rate and oxygen saturation.[47,52,55] Cortisol level was only measured in two study.[45,46]

**Risk of bias in included studies**

The risk on bias of each included study is detailed in Figure 2. Across the 14 included studies, six reported adequately the random-sequence generation,[42,43,50,51,54,55] whereas the other studies did not mention the process used. One study was quoted as high risk,[53] because there was no randomization (quasi-experimental design). Only two studies did not provide an adequate allocation concealment description[54,55]. Considering the nature of the interventions, blinding of personnel was especially difficult and was only performed adequately in three studies.[45,50,55] Infants were blinded in all studies. As pain is often measured by video recordings, blinding of outcome assessment was adequately done in 11 studies.[41-47,50,51,53,55] Incomplete data outcome was properly reported in four studies.[44,47,50,54] All studies were quoted as low risk of bias for selective reporting excepted one[43] because the study's hypotheses were modified after the results. Eight studies were quoted as having a high risk of bias for the other bias category since the odor used for the olfactive stimulation in those studies were very strong, such as the vanilla odor, which could have been smelled by an infant in the control group if near of an infant in the experimental group, thus causing contamination between groups.[41-44,46,48,51,53] A detailed explanation of the risk of bias judgment is presented in Table S2 of the supplemental digital content.

**Risk of Bias Across Studies**

Heterogeneity ($I^2$) varied between 0% and 99%. The heterogeneity could be explained essentially by the type of odor used as confirmed by subgroup analysis as well as by the differences in the selected population, differences in the neonatal unit setting and control interventions or comparators. We could not perform a funnel plot analysis to assess publication bias as initially planned in the protocol[22] because less than ten studies were included in the meta-analysis.[56] Only one included study was quote "high risk" for selective reporting bias.[43]

**Effects of olfactive stimulation interventions: Synthesis of results**

**Familiar odor vs Standard care**

*Pain reactivity (all pain scores).* Five studies,[47,48,50,51,54] including 390 participants (treatment=239, control=151), investigated the effects of an olfactive stimulation intervention, using a familiar odor (natural odor and artificial odor with habituation) compared to standard care, on pain reactivity. The effect was found to be significant to reduce infants' pain reactivity (SMD -0.69; 95% CI -0.93 to -0.44; $I^2$ = 20%, $p$<0.00001). Subgroup analysis by types of odor[22] suggest that mother's milk odor (n=171), (SMD -0.82; 95% CI -1.26 to -0.39; $I^2$ = 42%, $p$=0.0002), and artificial odor with habituation (n=148), (SMD -0.67; 95% CI -1.01 to -0.33; $I^2$ = 0%, $p$=0.0001) were both effective interventions to reduce pain reactivity compared to standard care. However, amniotic fluid odor was not effective (n=71) (SMD -0.38; 95% CI -0.88 to 0.12; $I^2$ = 0%, $p$=0.13). Forest plots for all are available in the supplemental digital content (Figures S1-S239).

*Pain regulation (all pain scores).* A total of four studies,[47,48,50,51] involving 310 participants (treatment=199, control=111), found that pain regulation was significantly lower for those receiving the olfactive stimulation intervention using a familiar odor than for those receiving standard care (SMD -0.40; 95% CI -0.66 to -0.14; $I^2$ = 13%, $p$= 0.002). Subgroup analyses by the types of odor did not reveal a significant effect for mother's milk odor (SMD -0.43; 95% CI -0.89 to 0.04; $I^2$ = 51%, $p$= 0.07), amniotic fluid odor (SMD -0.40; 95% CI -0.90 to 0.10; $I^2$ = 0%, $p$= 0.11), and artificial odor with habituation compared to standard care (SMD -0.24; 95% CI -0.75 to 0.26; $p$= 0.35).

*Crying duration* d*uring the procedure.* The effect of a familiar odor on time of crying during a painful procedure was reported by seven studies[41-43,48,50,54,55] including 357 participants (treatment= 193, control=164). The results showed a significant difference between groups favoring the experimental group (SMD -0.42; 95% CI -0.73 to -0.10; $I^2$ =

47%, *p*= 0.009. Subgroup analyses suggested that mother's milk odor (SMD -0.25; 95% CI -0.86 to 0.36; $I^2$=43%, *p*=0.43), vanilla odor with habituation (SMD -0.53; 95% CI -1.12 to 0.05; $I^2$=63%, *p*=0.07), and amniotic fluid odor was not statistically significant when compared to standard care (SMD -0.11; 95% CI -0.84 to 0.62; *p*=0.78). However, lavender odor with habituation seemed effective for reducing the duration of crying (SMD -0.50; 95% CI -0.94 to -0.05; *p*=0.003) compared to standard care, but this result was only based on one study. There was no significant difference between subgroups (*p*=0.74; $I^2$ = 0%).

*Crying duration after the procedure.* We conducted a meta-analysis including five studies[41-43,48,50] (n=187) to investigate the effect of a familiar odor on crying duration after the procedure. There was a significant difference in favor of the familiar odor group compared to the standard care group (SMD -0.37; 95% CI -0.68 to -0.07; $I^2$ = 0%, *p*= 0.01). Mother's milk odor was significantly superior to standard care (SMD -0.48; 95% CI -0.94 to -0.03; $I^2$ = 0%, *p*= 0.04), while vanilla with habituation (SMD -0.33; 95% CI -0.97 to 0.30; $I^2$ = 41%, *p*= 0.30) and amniotic fluid (SMD -0.18; 95% CI -0.91 to 0.55; *p*= 0.63) were not significantly different than standard care for reducing crying duration after the procedure. There was no significant difference between subgroups (*p*=0.78; $I^2$=0%).

*Heart rate variation during the procedure*. Three studies[47,52,55] (n=302) assessed heart rate variation during the painful procedure and did not find a significant difference between the familiar odor group and the standard care group (MD -2.29; 95% CI -4.88 to 0.31; $I^2$ = 97%, *p*=0.08). Subgroup analyses revealed that mother's milk odor (two studies) (MD -4.46; 95% CI -7.39 to -1.54; $I^2$=40%, *p*=0.003) was effective, while vanilla odor with habituation (two studies) (MD -0.42; 95% CI -1.77 to 0.93; $I^2$=20%, *p*=0.54) and amniotic fluid odor (one study) were not (MD 1.24; 95% CI -4.15 to 6.63; *p*=0.65). There was a significant difference between subgroups (*p*=0.03; $I^2$ = 70.4%).

*Heart rate variation after the painful procedure*. Three studies[47,52,55] (n=212) investigated the heart rate variation and analyses showed a significant evidence in favor of the familiar odor group compared to standard care (MD -3.87; 95% CI -7.36 to -0.38; $I^2$ = 99%, *p*=0.03), with a considerable heterogeneity. Mother's milk odor was significant (two studies) (MD -6.26; 95% CI -11.40 to -1.11; $I^2$ = 80%, *p*=0.02), but vanilla with habituation (one study) (MD 0.17; 95% CI -0.22 to 0.56; *p*=0.39) and amniotic fluid (one study) (MD -4.87; 95% CI -13.36. to 3.62; *p*=0.26) were not. There was a significant difference between subgroups (*p*=0.03; $I^2$=72.6%).

*Oxygen saturation variation during the procedure*. Three studies,[47,52,55] including 302 participants, investigated the effects of olfactive stimulation using a familiar odor

intervention, compared to the standard care, on the variation in oxygen saturation during the painful procedure and a significant effect was found favoring the experimental group with a considerable heterogeneity (MD -0.47; 95% CI -0.86 to -0.08; $I^2$=91%, *p*=0.02). Subgroup analyses suggest that mother's milk odor (two studies) (MD -0.79; 95% CI -0.90 to -0.67; $I^2$=0%, *p*<0.00001) and vanilla odor with habituation (two studies) (MD -0.23; 95% CI -0.37 to -0.10; $I^2$=0%, p=0.0008) were effective, while amniotic fluid odor was not (one study) (MD 1.15; 95% CI -0.59 to 0.89; *p*=0.69).

*Oxygen saturation variation after the procedure. R*esults of two studies[47,52,55] with 211 participants, revealed a significant effect favoring the experimental group on the variation in oxygen saturation with a considerable heterogeneity (MD -0.56; 95% CI -0.99 to -0.13; $I^2$=99%, *p*=0.01). Mother's milk odor (two studies) (MD -0.98; 95% CI -1.03 to -0.92; $I^2$=0%, *p*<0.00001) and vanilla odor with habituation (one study) (MD -0.42; 95% CI -0.47 to -0.37; *p*<0.00001) were effective, but amniotic fluid odor (one study) was found to not be an effective intervention (MD 0.02; 95% CI -0.57 to 0.61; *p*=0.95).

*Cortisol variation*. Two studies[45,46] (n= 131 full-term infants) measured salivary cortisol as a hormonal indicator of pain, but data were not available to perform a meta-analysis (missing data and not suitable for data imputation).

*Adverse events*. No study reported any adverse event or no adverse event was observed. Only three studies mentioned the safety of the intervention and data were not available to perform a meta-analysis. Across the included studies, two studies mentioned the safety of the intervention without providing details[49,54] and one study specified that familiar odor did not generate distress for the infants during blood sampling.[51] Other studies did not mention the safety of the intervention.

**Familiar odor vs Artificial odor without habituation**

*Pain reactivity (all pain scores)***.** Two studies[47,49] examined the effect of a familiar odor on infants' pain reactivity compared to an artificial odor without habituation, so an unfamiliar odor. The meta-analysis, including 128 participants (treatment=75, control=53), did not show a significant effect with a considerable heterogeneity (SMD -0.28; 95% CI -1.39 to 0.83; $I^2$ = 89%, *p*=0.62). Results are presented in a forest plot. For subgroup analyses, two studies[47,49] (n=88) showed no evidence of mother's milk odor effects (SMD -0.61; 95% CI -2.12 to 0.90; $I^2$ = 91%, *p*=0.43) and amniotic fluid odor (one study), (SMD 0.39; 95% CI -0.26 to 1.05; *p*=0.24), on pain reactivity.

*Pain regulation*. Results from one study[47] (n=78 full-term infants) contributing with two experimental groups, comparing familiar odors (mother's milk and amniotic fluid) to an unfamiliar odor (lavender without habituation), revealed no significant difference between groups for the pain regulation outcome (SMD -0.17; 95% CI -.063 to 0.29; $I^2$ = 0%, *p*=0.47). No subgroup analysis was performed as only one study was included.

*Crying duration during the procedure.* The effect of a familiar odor compared to an artificial odor without habituation on crying duration during the procedure was reported in five studies[41-43,49,55] including 230 participants (treatment=119, control=111), and showed a significant difference between groups in favor of the experimental group (SMD -0.49; 95% CI -0.92 to -0.06; $I^2$ = 54%, *p*=0.03). Subgroup analyses showed no evidence of mother's milk odor (two studies, n=67), (SMD -0.44; 95% CI -1.26 to 0.37; $I^2$ = 53%, *p*=0.29) or vanilla odor with habituation (four studies, n=163), (SMD -0.48; 95% CI -1.09 to 0.13; $I^2$ = 66%, *p*=0.12) on crying duration.

*Crying duration after the procedure.* A meta-analysis with three studies[41-43] (n=90) showed no evidence of effect on crying duration after the painful procedure (SMD -0.11; 95% CI -0.53 to 0.31; $I^2$=0%, *p*=0.51). No effect was found with subgroup analyses of mother's milk odor (one study, n=17), (SMD -0.35; 95% CI -1.35 to 0.66; *p*=0.50) and vanilla odor with habituation (two studies, n=73), (SMD -0.06; 95% CI -0.53 to 0.41; $I^2$ = 0%, *p*=0.80).

*Heart rate variation during and after the procedure*. For heart rate variation during the procedure, two studies[47,55] (n=168 full-term infants) showed no evidence of effects of a familiar odor compared to an artificial odor without habituation (MD 0.16; 95% CI -4.54 to 4.87; $I^2$=47%, *p*=0.95). After the procedure, one study[47] including two experimental arms (n=78 full-term infants) showed evidence of effects on the heart rate variation (MD -6.75; 95% CI -12.51 to -0.99; $I^2$=0%, *p*=0.02).

*Oxygen saturation variation during and after the procedure*. Based on two studies[47,55] (n=168 full-term infants), there was no significant difference between groups on the variation in oxygen saturation during the procedure (MD 0.10; 95% CI -0.95 to 1.14; $I^2$=86%, *p*=0.86) nor after the procedure based on one study (n=78 full-term infants), (MD 0.38; 95% CI -0.35 to 1.11; $I^2$=56%, *p*=0.30).

It was not possible to perform subgroup analyses for heart rate and oxygen saturation variation during and after the painful procedure considering that only two studies were included.

*Cortisol variation*. One study[49] involving 50 full-term infants showed that the salivary cortisol level was significantly higher in those receiving the olfactive stimulation intervention

using mother's milk odor compared to those receiving a formula milk odor (MD -8.70; 95% CI -11.49 to -5.91; $p$<0.00001).

*Adverse events***.** Data were not available to perform a meta-analysis. Five studies[41-43,47,55] did not mention if adverse events were observed or not. Only one study mentioned the safety of the intervention without providing details.[49]

**Additional analyses**

We performed subgroups analyses according to the population (preterm and full-term infants). All these results are presented in Table 2. An olfactive stimulation intervention using a familiar odor compared to standard care was found to significantly reduce pain reactivity in both preterm infants (n=234) (SMD -0.64; 95% CI -1.05 to -0.23; I2 = 52%, p= 0.002) and full-term infants (n=156) (SMD -0.73; 95% CI -1.07 to -0.40; I2 = 0%, p<0.0001); whereas pain regulation was only significantly lower in preterm infants (SMD -0.42; 95% CI -0.78 to -0.06; I2 = 41%, p= 0.02). A familiar odor compared to an artificial odor without habituation was only found to significantly reduce pain reactivity in preterm infants (n=50) (SMD -1.38; 95% CI -2.00 to -0.76; p<0.0001).

**Quality of evidence summary**

In this systematic review, the overall quality of evidence was considered as low to very low. Each comparison is reported in a separate "summary of findings" table. For the comparison between familiar odor interventions and standard care, the overall quality of evidence was rated as low for pain reactivity, pain regulation and crying duration after the procedure and very low for crying duration during the procedure, heart rate variation (during and after the procedure) and oxygen saturation variation (during and after the procedure) (Table 3). For the comparison of familiar odor and artificial odor without habituation, the overall quality of evidence was rated as low for heart rate variation during the procedure and very low for the other outcomes (supplemental digital content, Table S3). For the two comparisons, the main reasons for downgrading scores were high risk of bias, high heterogeneity and small sample size.

## DISCUSSION

**Summary of main results**

This systematic review aimed to evaluate the effectiveness of olfactive stimulation interventions on preterm and full-term infants' pain response including pain reactivity and

pain regulation, as well as secondary outcomes such as crying duration, heart rate and oxygen variations, cortisol level and occurrence of adverse events. This review summarized the evidence of 14 trials including a total of 1028 preterm and full-term infants. Ten RCTs were included in the meta-analysis to evaluate the effectiveness of olfactive stimulation interventions, using a familiar odor, compared to standard care or to unfamiliar odor on pain reactivity, pain regulation, crying duration (during and after the procedure), heart rate and oxygen saturation variations (during and after the procedure), as well as cortisol level.

According to our results, olfactive stimulation interventions using a familiar odor compared to standard care were statistically effective to manage procedural pain. More specifically, we found a significant difference in favor of the olfactive stimulation interventions for pain reactivity, pain regulation, crying duration (during and after the procedure), heart rate variation after the procedure and oxygen saturation variation during and after the procedure. However, it did not seem effective on heart rate variations during the procedure. Following the Cochrane Handbook recommendations' to convert SMD, our results on pain reactivity represent a difference of 1.9 on the PIPP, which is clinically significative, and 1.1 for pain regulation.[26] Subgroup analyses showed that mother's milk odor had a significant effect on all outcomes except pain regulation and crying duration during the procedure. Moreover, an artificial odor with a habituation was effective for pain reactivity, as well as oxygen saturation variation during and after the procedure. Regarding the comparison of olfactive stimulation interventions using a familiar odor and an artificial odor without habituation, a significant effect was found only on crying duration during the procedure, heart rate variation after the procedure and cortisol levels based on a small number of participants. Subgroup analyses did not reveal significant differences. Subgroup analyses performed by population revealed that olfactive stimulation interventions compared to standard care are effective on preterm and full-term infant pain reactivity, on preterm infants' pain regulation and crying duration after the procedure as well as on full-term infants' crying duration during the procedure and heart rate variations afterthe procedure. For adverse events, we were not able to perform a meta-analysis as no adverse events were reported in included studies. Only the safety of the intervention was mentioned by few authors and no major nor minor issues were reported.

**Quality of evidence**

This meta-analysis showed a significant effect of olfactive stimulation interventions compared to standard care on almost all outcomes. However, these results are based on low-

to very low quality of evidence, which means that our confidence in the estimated effect is limited[37] due to high risk of bias, high statistical heterogeneity and small sample size in included studies. The risk of bias was affected by four main reasons. First, the allocation concealment was not clearly described, except in two studies.[54,55] Second, considering the nature of the interventions, adequate blinding of the personnel might be questionable since they could have been able to smell the odors. Third, the risk of contamination between groups was present in the majority of studies,[41-44,46,48,51,53] especially when a strong odor was used such as vanilla or lavender. Fourth, pain response was sometimes not measured by a valid and reliable pain scale.[41-43,55] We did not downgraded for publication bias and we minimize it by following a systematic process for study identification by conducting a comprehensive search in nine scientific databases, three grey literature databases and contacting authors of potential studies for missing data.

The quality of evidence was also affected by a considerable heterogeneity in six of the 17 comparisons. It was nonetheless important to conduct these comparisons, while using a random effects model, as it provided interesting and preliminary results on the trends of the effects. In addition, most of the statistical heterogeneity was explained by subgroups analysis by the type of odors. The remaining heterogeneity could be explained by differences in the setting, the type of painful procedure performed (i.e. heel prick, venipuncture) or the standard care, which were often not described in the studies. Noteworthy, standard care usually varies in different clinical settings and therefore should be carefully described to show proper control in studies.

One of the major concerns regarding the quality of evidence in this systematic review was the small sample size of included studies. These studies might have lacked power to detect the effect of the interventions. These findings highlight the necessity of conducting more rigorous RCTs with larger sample size.

**Limitations**

First, changes were made following the protocol publication[22] and those were clearly reported in the method section. Second, the search strategy was limited to French and English articles and the grey literature search was only performed in English. Third, the absence of response from three authors regarding missing data, despite several attempts to reach them, precluded us from adding data to the statistical analyses which could have affected the pooled estimate for the two comparisons on crying duration and cortisol level. Finally, standard care interventions may have differed between studies related to different neonatal unit practices.

However, as the standard care were often not described, it was not possible to integrate these differences into our data interpretation.

**Agreements and disagreements with other studies or reviews**

One systematic review has recently been published on mother's milk odor to manage pain in infants.[23] Results found by the authors are congruent with our results on the subgroup analyses with mother's milk odor : a significant effect was found on pain scores during the procedure, crying duration after the procedure, heart rate and oxygen variations during and after the procedure. Besides considering all types of odor, another important distinction between this review and ours lies in the methods to retrieve potential studies as well as the evaluation of the quality of evidence. Thus, our systematic review contributes to a more comprehensive review on olfactive stimulation interventions, as there are few studies published on this topic.

**Clinical Implications**

Procedural pain management interventions in preterm an full-term infants are crucial to prevent consequences of untreated pain, such the impact on motor and intellectual skills as well as a hypersensitivity to pain up to 7 years of age.[3-6] However, in included studies, standard care was often not described or the control group did not receive any intervention, which is a major ethical concern considering the untreated pain consequences. Non-pharmacological interventions have gained popularity because many of them can be easily applied in practice with minimal preparation or cost and it is specially the case for olfactive stimulation interventions. In addition, olfactive stimulation interventions could be used in combination with other non-pharmacological interventions, such as sucrose. The American Academy of Pediatrics[12] recommends the combination of different non-pharmacological interventions to enhance their effectiveness. Two studies[57,58] have evaluated the effectiveness of olfactive stimulation interventions in combination with other sensory interventions (i.e. tactile, auditive) and revealed a significant effect on pain response. However, multisensory interventions test the effectiveness of a combination of three interventions, so it is impossible to know which intervention is more effective or how it works as the evaluated outcome is the result of their interaction.

This systematic review helps us to enhance our understanding about mechanisms underlying olfactive stimulation interventions as these could have an impact on the emotional

and affective pain components. Subgroup analyses in this review found that mothers' milk odor and an artificial odor with a previous period of habituation are the most effective to reduce pain in infants. Moreover, the comparison of olfactive stimulation interventions using a familiar odor compared to an unfamiliar odor (artificial odor without habituation) revealed a significant effect on crying duration during the procedure, heart rate variation after the procedure and cortisol levels. These results mean that a familiar odor, either natural or artificial with habituation, has the potential to reduce the pain response. Current knowledge on the central nervous system also support the potential impact of olfactive stimulation interventions on emotional and affective pain components. In fact, the olfactory system is anatomically linked to the limbic system (emotions), which is involve in the perception of pain.[59,60] Thus, the mechanism of action would be at the cortical level as corroborated by animal studies.[61] The olfactive stimulation intervention with a habituation period would have a greater effect, this being explained in animal studies by a systemic effect during prolonged exposure to an odor.[62,63] Interestingly, the period of habituation was only done for artificial odors in included studies. Only one pilot study investigated a previous period of habituation with mothers' milk odor.[64] The results indicated that is was feasible and acceptable for both mother and nurses to perform the period of habituation with mothers' milk odor.[64] In light of the meta-analysis results, health professionals could use a pad with mother's milk or an artificial odor (with an habituation period) near the infant's nose when performing routine blood sampling.

## CONCLUSION

Evidence of effectiveness of olfactive stimulation interventions were found in this review, based on low to very low quality of evidence. Adverse events were not reported nor assessed by authors. Further studies should always report adverse events to guide clinical practice and ensure the safety of olfactive stimulation intervention. Moreover, even if non-pharmacological interventions are not always applicable in neonatal unit practice or performed by healthcare professionals, especially by nurses,[65] standard care should always implied an intervention considering the consequences of untreated pain. Included studies focused mainly on heel pricks and venipunctures, but the effect of the intervention on other painful procedures remains unknown. Future research should investigate the effectiveness of olfactive stimulation interventions during other frequent painful procedure such as nasal or tracheal aspiration, installation or removal of peripheral intravenous line.[17]. Furthermore, habituation with artificial odors lasted in the identified studies between 8 and 18 hours

representing an important difference time interval. As for the intervention, different methods of administering the odor were used by the authors (gaze, odor diffuser, odor deposited inside the incubator, filter paper, a scarf or a doll) with distances from one millimeter to ten centimetres. Therefore, future research should examine which modality could be the most effective for pain management. To conclude, more rigorous studies with larger sample size are required.


## ACKNOWLEDGEMENTS

The authors would like to thank the following persons for their valuable implication: Ms. Assia Mourid, librarian at the Faculty of Nursing of the University of Montreal and Ms. Josianne Dupuis, librarian, for their help in designing and performing the search strategy; Dre Sylvie Le May, Professor at the Faculty of Nursing of the Université de Montréal, for her precious advices; Dre Nathalie Goubet for her precious collaboration to obtain missing data and the Applied Clinical Research Unit (URCA) of the CHU Sainte-Justine for their statistical support. GD would like to thank the Fonds de Recherche en Santé du Québec (FRQ-S) for doctoral scholarship.

## FUNDING SOURCES

GD received funding for communication and transfer of knowledge from the Faculty of Nursing of the Université de Montréal to support the publication of this original article. This review is conducted as part of the doctoral studies of GD.


## CONFLICT OF INTEREST

All authors have nothing to declare.

# REFERENCES


1. Buonocore G, Bellieni CV, SpringerLink. *Neonatal Pain : Suffering, Pain, and Risk of Brain Damage in the Fetus and Newborn.* 2 ed: Cham : Springer International Publishing : Imprint: Springer; 2017.
2. Cruz MD, Fernandes AM, Oliveira CR. Epidemiology of painful procedures performed in neonates: A systematic review of observational studies. *European journal of pain.* 2016;20(4):489-498.
3. Grunau RE, Whitfield MF, Petrie-Thomas J, et al. Neonatal pain, parenting stress and interaction, in relation to cognitive and motor development at 8 and 18 months in preterm infants. *Pain.* 2009;143(1-2):138-146.
4. Valeri BO, Ranger M, Chau CM, et al. Neonatal Invasive Procedures Predict Pain Intensity at School Age in Children Born Very Preterm. *Clin J Pain.* 2016.
5. Taddio A, Shah V, Atenafu E, Katz J. Influence of repeated painful procedures and sucrose analgesia on the development of hyperalgesia in newborn infants. *PAIN®.* 2009;144(1):43-48.
6. Brummelte S, Grunau RE, Chau V, et al. Procedural pain and brain development in premature newborns. *Annals of neurology.* 2012;71(3):385-396.
7. Zwicker JG, Grunau RE, Adams E, et al. Score for neonatal acute physiology-II and neonatal pain predict corticospinal tract development in premature newborns. *Pediatric neurology.* 2013;48(2):123-129.e121.
8. Ranger M, Zwicker JG, Chau CMY, et al. Neonatal Pain and Infection Relate to Smaller Cerebellum in Very Preterm Children at School Age. *The Journal of Pediatrics.* 2015;167(2):292-298.e291.
9. Ranger M, Chau CM, Garg A, et al. Neonatal pain-related stress predicts cortical thickness at age 7 years in children born very preterm. *PloS one.* 2013;8(10):e76702.
10. Vinall J, Miller SP, Bjornson BH, et al. Invasive procedures in preterm children: brain and cognitive development at school age. *Pediatrics.* 2014;133(3):412-421.
11. Doesburg SM, Chau CM, Cheung TP, et al. Neonatal pain-related stress, functional cortical activity and visual-perceptual abilities in school-age children born at extremely low gestational age. *Pain.* 2013;154(10):1946-1952.
12. American Academy of Pediatrics. Prevention and Management of Procedural Pain in the Neonate: An Update. *Pediatrics.* 2016.
13. Stevens B, Yamada J, Ohlsson A, Haliburton S, Shorkey A. Sucrose for analgesia in newborn infants undergoing painful procedures. *Cochrane Database of Systematic Reviews.* 2016(7).
14. Johnston C, Campbell-Yeo M, Disher T, et al. Skin-to-skin care for procedural pain in neonates. *Cochrane Database of Systematic Reviews.* 2017(2).
15. Shah PS, Herbozo C, Aliwalas LL, Shah VS. Breastfeeding or breast milk for procedural pain in neonates. *Cochrane Database of Systematic Reviews.* 2012(12):N.PAG-N.PAG 1p.
16. Johnston C, Barrington KJ, Taddio A, Carbajal R, Filion F. Pain in Canadian NICUs: Have We Improved Over the Past 12 Years? *The Clinical Journal of Pain.* 2011;27(3):225-232.
17. Carbajal R, Rousset A, Danan C, et al. Epidemiology and treatment of painful procedures in neonates in intensive care units. *Jama.* 2008;300(1):60-70.
18. Marlier L, Gaugler C, Astruc D, Messer J. La sensibilité olfactive du nouveau-né prématuré. *Archives de pediatrie.* 2007;14.

19. Browne JV. Chemosensory Development in the Fetus and Newborn. *Newborn and Infant Nursing Reviews.* 2008;8(4):180-186.
20. Allam MD-E, Marlier L, Schaal B. Learning at the breast: Preference formation for an artificial scent and its attraction against the odor of maternal milk. *Infant Behavior and Development.* 2006;29(3):308-321.
21. Marlier L, Schaal B. Human Newborns Prefer Human Milk: Conspecific Milk Odor Is Attractive Without Postnatal Exposure. *Child development.* 2005;76(1):155-168.
22. De Clifford-Faugère G, Lavallée A, Aita M. Olfactive stimulation interventions for managing procedural pain in preterm and full-term neonates: a systematic review protocol. *BMC Systematic reviews.* 2017;6(1):203.
23. Zhang S, Su F, Li J, Chen W. The Analgesic Effects of Maternal Milk Odor on Newborns: A Meta-Analysis. *Breastfeeding Medicine.* 2018;13(5):327-334.
24. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J. Clin. Epidemiol.* 2009;62(10):1006-1012.
25. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS medicine.* 2009;6(7):e1000100.
26. Higgins J, Green S. Cochrane handbook for systematic reviews of interventions. The Cochrane Collaboration; 2011.
27. Pillai Riddell RR, Racine NM, Gennis HG, et al. Non-pharmacological management of infant and young child procedural pain. *The Cochrane database of systematic reviews.* 2015;12:Cd006275.
28. Pillai Riddell R, Racine NM, Craig KD, Campbell L. Psychological theories and biopsychosocial models in paediatric pain. In: McGrath PJ, Stevens B, Walker CD, Zempsky WT, eds. *Oxford Textbook of Paediatric Pain*. Oxford, UK: Oxford University Press; 2013.
29. Als H. Toward a synactive theory of development: Promise for the assessment and support of infant individuality. *Infant mental health journal.* 1982;3(4):229-243.
30. Stevens B, Johnston C, Petryshen P, Taddio A. Premature Infant Pain Profile: development and initial validation. *The Clinical journal of pain.* 1996;12(1):13.
31. Stevens BJ, Gibbins S, Yamada J, et al. The premature infant pain profile-revised (PIPP-R): initial validation and feasibility. *Clin J Pain.* 2014;30(3):238-243.
32. Lawrence J, Alcock D, McGrath P, Kay J, MacMurray SB, Dulberg C. The development of a tool to assess neonatal pain. *Neonatal network: NN.* 1993;12(6):59-66.
33. Grunau RV, Craig KD. Pain expression in neonates: facial action and cry. *Pain.* 1987;28(3):395-410.
34. Carbajal R, Paupe A, Hoenn E, Lenclen R, Olivier-Martin M. DAN : une échelle comportementale d'évaluation de la douleur aiguë du nouveau-né. *Archives de Pédiatrie.* 1997;4(7):623-628.
35. Higgins JP, Altman DG, Gotzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ.* 2011;343:d5928.
36. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ (Clin Res Ed).* 2008;336.
37. Schünemann H, Bro ek J, Guyatt G, Oxman A (editors). *GRADE handbook for grading quality of evidence and strength of recommendations*. 2013. http://gdt.guidelinedevelopment.org/centralprod/design/client/handbook/handbook.html.

38. GRADE working group. Grading quality of evidence and strength of recommendations. *BMJ.* 2004;328(7454):1490.
39. Higgins J, Li T, Deeks JJ. Chapter 6: Choosing effect measures and computing estimates of effect. In: Higgins J, Thomas J, Chandler J, et al., eds. *Cochrane Handbook for Systematic Reviews of Interventions*. London: Cochrane; 2019.
40. Deeks JJ, Higgins J, Altman DG, Cochrane Statistical Methods Groups. Chapitre 9: Analysing data and undertaking meta-analyses. In: Higgins J, Green Se, eds. *Cochrane Handbook for Systematic Reviews of Interventions*.2011.
41. Goubet N, Rattaz C, Pierrat V, Bullinger A, Lequien P. Olfactory experience mediates response to pain in preterm newborns. *Developmental psychobiology.* 2003;42(2):171-180.
42. Goubet N, Strasbaugh K, Chesney J. Familiarity breeds content? Soothing effect of a familiar odor on full-term newborns. *Journal of developmental and behavioral pediatrics : JDBP.* 2007;28(3):189.
43. Rattaz C, Goubet N, Bullinger A. The calming effect of a familiar odor on full-term newborns. *J Dev Behav Pediatr.* 2005;26(2):86-92.
44. Romantsik O, Porter RH, Varendi H. The effects of olfactory stimulation and gender differences on pain responses in full-term infants. *Acta paediatrica..* 2014;103(11):1130-1135.
45. Nishitani S, Miyamura T, Tagawa M, et al. The calming effect of a maternal breast milk odor on the human newborn infant. *Neuroscience Research.* 2009;63(1):66-71.
46. Kawakami K, Takai-Kawakami K, Okazaki Y, Kurihara H, Shimizu Y, Yanaihara T. The effect of odors on human new born infants under stress. *Infant Behavior and Development.* 1997;20(4):531-535.
47. Akcan E, Polat S. Comparative Effect of the Smells of Amniotic Fluid, Breast Milk, and Lavender on Newborns' Pain During Heel Lance. *Breastfeeding medicine.* 2016.
48. Kucuk Alemdar D, Kardas Ozdemir F. Effects of Having Preterm Infants Smell Amniotic Fluid, Mother's Milk, and Mother's Odor During Heel Stick Procedure on Pain, Physiological Parameters, and Crying Duration. *Breastfeeding medicine.* 2017;12:297-304.
49. Badiee Z, Asghari M, Mohammadizadeh M. The Calming Effect of Maternal Breast Milk Odor on Premature Infants. *Pediatrics & Neonatology.* 2013;54(5):322-325.
50. Baudesson de Chanville A, Brevaut-Malaty V, Garbi A, Tosello B, Baumstarck K, Gire C. Analgesic Effect of Maternal Human Milk Odor on Premature Neonates: A Randomized Controlled Trial. *Journal of human lactation.* 2017;33(2):300-308.
51. Jebreili M, Neshat H, Seyyedrasouli A, Ghojazade M, Hosseini MB, Hamishehkar H. Comparison of breastmilk odor and vanilla odor on mitigating premature infants' response to pain during and after venipuncture. *Breastfeeding Medicine.* 2015;10(7):362-365.
52. Neshat H, Jebreili M, Seyyedrasouli A, Ghojazade M, Hosseini MB, Hamishehkar H. Effects of Breast Milk and Vanilla Odors on Premature Neonate's Heart Rate and Blood Oxygen Saturation During and After Venipuncture. *Pediatrics & Neonatology.* 2015.
53. Mellier D, Bezard S, Caston J. Études exploratoires des relations intersensorielles olfaction–douleur. [Exploratory studies of intersensory olfaction–pain relationships]. *Enfance.* 1997;1:98-111.
54. Razaghi N, Hoseini ASS, Aemmi SZ, Mohebbi T, Boskabadi H. The Effect of Lavender Scent on Pain of Blood Sampling in Term Neonates. *International Journal of Pediatrics.* 2015;3(2):535-541.

55. Sadathosseini AS, Negarandeh R, Movahedi Z. The Effect of a Familiar Scent on the Behavioral and Physiological Pain Responses in Neonates. *Pain Management Nursing.* 2013;14(4):e196-203 191p.
56. Sterne JA, Egger M, Moher D, Cochrane Bias Methods Group. Chapter 10: Adressing reporting bias. In: Higgins J, Green Se, eds. *Cochrane Handbook for Systematic Reviews of Interventions*2011.
57. Bellieni CV, Cordelli DM, Marchi S, et al. Sensorial saturation for neonatal analgesia. *The Clinical journal of pain.* 2007;23(3):219.
58. Bellieni CV, Bagnoli F, Perrone S, et al. Effect of multisensory stimulation on analgesia in term neonates: a randomized controlled trial. *Pediatric research.* 2002;51(4):460.
59. Purves D, Augustine G, Fitzpatrick D, Hall WC, LaMantia A-S, White LE. *Neurosciences.* 5e ed: Louvain-la-Neuve : De Boeck Supérieur; 2015.
60. Marchand S. *Le phénomène de la douleur.* 2 ed. Montréal: Montréal : Chenelière-éducation; 2009.
61. Shipley MT, Ennis M, Puche AC. The Olfactory System. In: Conn PM, ed. *Neuroscience in Medicine*. Totowa, NJ: Humana Press; 2003:579-593.
62. Zhou L, Ohata M, Owashi C, Nagai K, Yokoyama I, Arihara K. Odors generated from the Maillard reaction affect autonomic nervous activity and decrease blood pressure through the olfactory system. *Journal of the Science of Food and Agriculture.* 2018;98(3):923-927.
63. Zhao F, Wang X, Zariwala HA, et al. fMRI study of olfaction in the olfactory bulb and high olfactory structures of rats: Insight into their roles in habituation. *NeuroImage.* 2016;127:445-455.
64. De Clifford-Faugère G, Aita M, Héon M, Le May S. Management of procedural pain in preterm infants through olfactive stimulation with mothers' milk: A pilot study. *Science of Nursing and Health Practices-Science infirmière et pratiques en santé.* 2019;2(1):3.
65. De Clifford-Faugere G, Aita M, Le May S. Nurses' practices regarding pain management of preterm infants in a context of research. *Applied Nursing Research.* 2018.
66. Harrison D, Reszel J, Bueno M, et al. Breastfeeding for procedural pain in infants beyond the neonatal period. *Cochrane Database of Systematic Review*. 2016, 10, CD011248. DOI: 10.1002/14651858.CD011248.pub2.
67. Cohen J. *Statistical power analysis for the behavioral sciences*. 2nd Edition. Routledge, 1988.

## ABBREVIATIONS

CI : confidence interval

GRADE: Grading of Recommendations Assessment, Development and Evaluation working group methodology

MD: mean difference

PRISMA: Preferred Reporting Items for Systematic review and Meta-Analysis

RCT: Randomized controlled trials

SD: standard deviation

SMD: standardized mean difference

WG: weeks of gestation

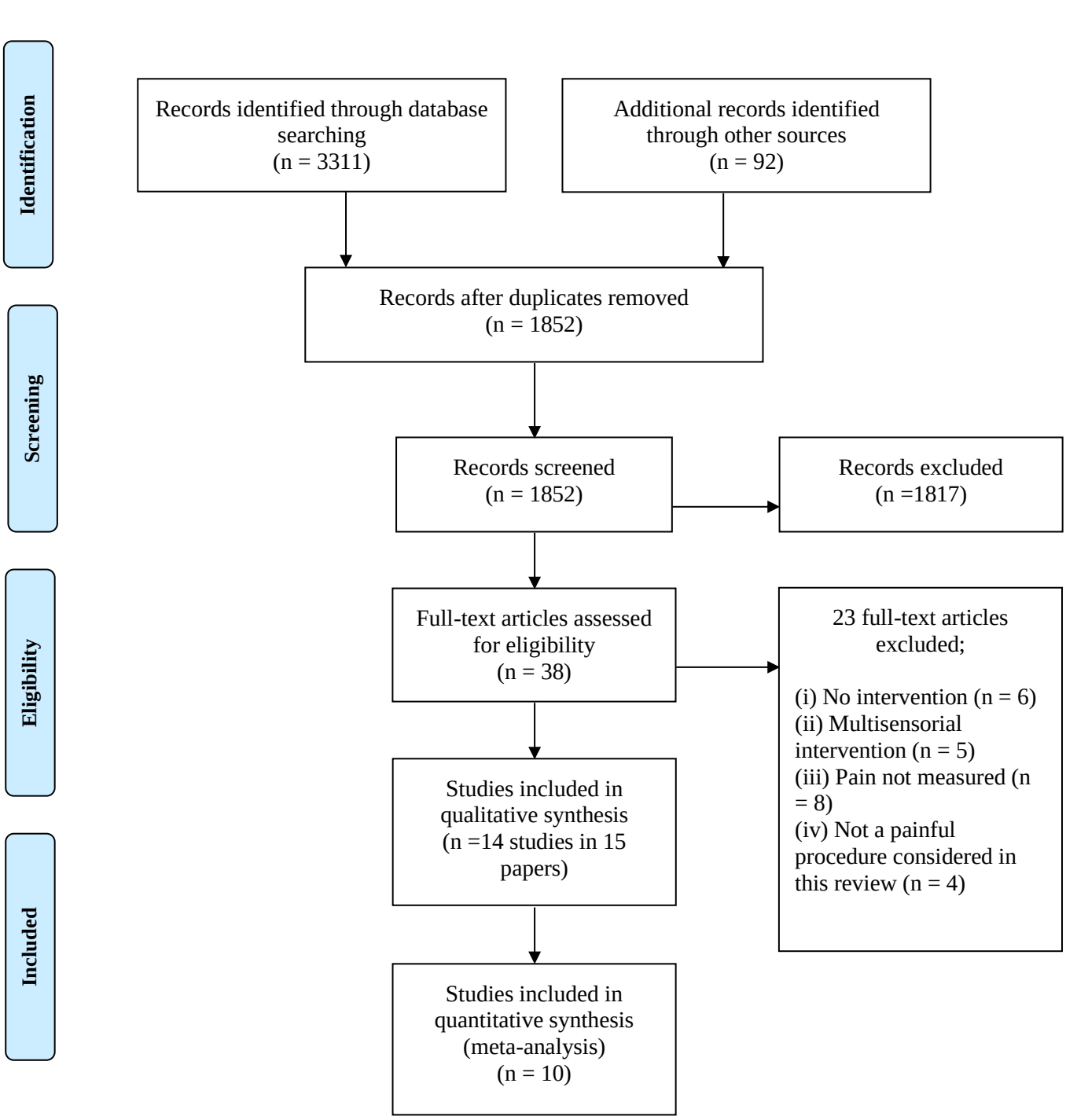


**Figure 1.** PRISMA Flow diagram

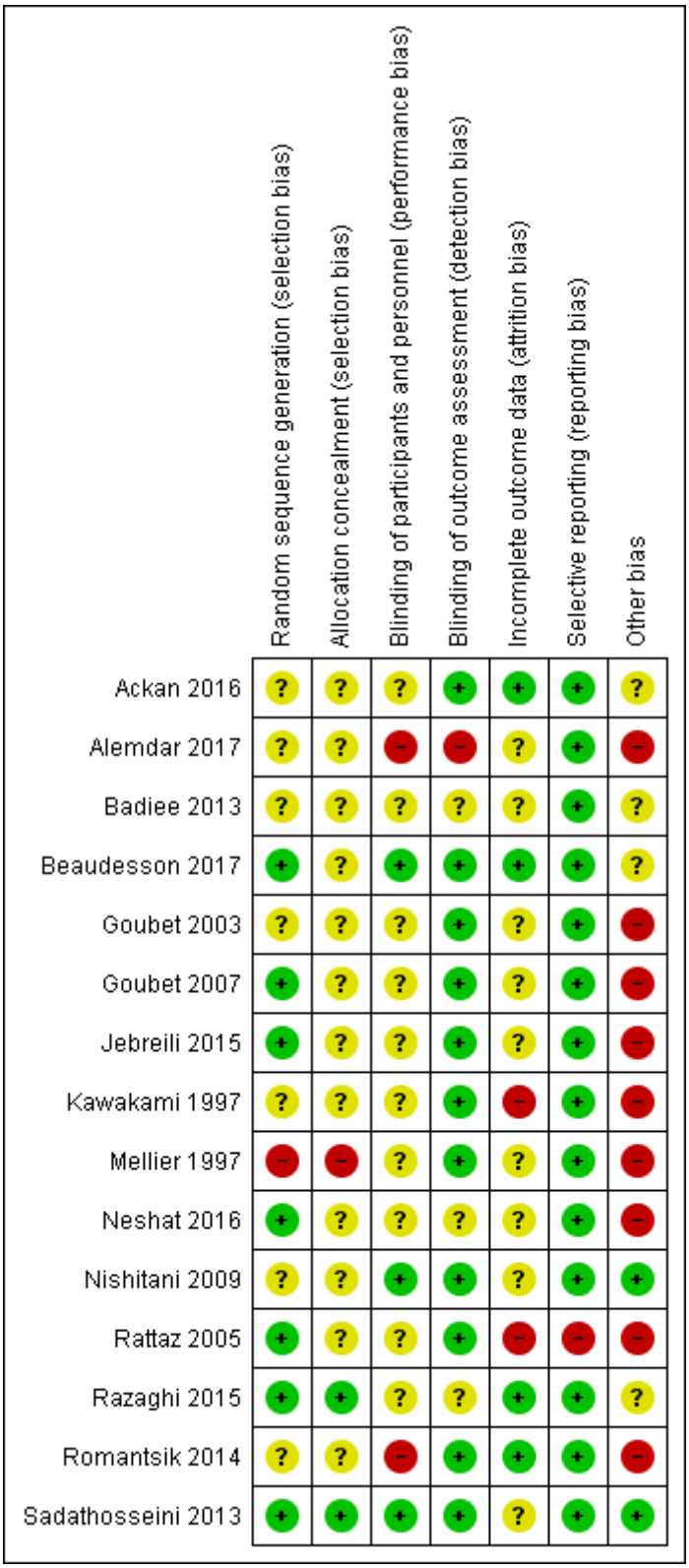


**Figure 2.** Risk of bias summary: review authors' judgements about each risk of bias item for each included study.

**Table 1. Characteristics of included studies**

| Study | Study design | Participants | Painful procedure | Intervention | Outcome | Results |
|---|---|---|---|---|---|---|
| Alemdar et al., 2017 | RCT, four parallel groups | N= 85 preterm<br>Age= 33 weeks (mean) | Heel-prick | G1: n=21, amniotic fluid odor<br>G2: n=22, mother milk odor<br>G3: n=20, mother odor<br>G4: n=22, control group | Pain measured by the PIPP<br>Crying duration | No significant difference between groups. |
| Baudesson de Chaville et al., 2017 | RCT, two parallel groups | N= 33 preterm<br>Age= 30-36+6 weeks | Venipuncture | G1: n=16, mother milk odor<br>G2: n=17, control group | Pain measured by the PIPP and the DAN<br>Crying duration | Mother milk odor is efficient to manage pain. |
| Ackan et al., 2016 | RCT, four parallel groups | N= 102 full-term<br>Age= 38-42 weeks | Heel-prick | G1: n=27, lavender odor<br>G2: n=24, mother milk odor<br>G3: n=26, amniotic fluid odor<br>G4: n=25, control group | Pain measured by the NIPS<br>Variations in heart rate and oxygen saturation | During the painful procedure, G1 experience less pain than G2 and G3. After the painful procedure, G1 and G2 experience less pain than G3. |
| Jebreili et al., 2015<br>Neshat et al., 2015 | RCT, three parallel groups | N= 135 preterm<br>Age= 28-32 weeks | Venipuncture | G1: n=45, mother milk odor<br>G2: n=45, vanilla odor and 12 hours of habituation<br>G3: n=45, control group | Pain measured by the PIPP<br>Variations in heart rate and oxygen saturation | Mother milk odor is efficient to manage pain compared to G2 and G3 (during and after the painful procedure). |
| Razaghi et al., 2015 | RCT, two parallel groups | N= 80 full-term<br>Age= up to 37 weeks (mean at 38 weeks) | Venipuncture | G1: n=40, inhaled lavender<br>G2: n=40, lavender odor and 8 hours of habituation | Pain measured by the DAN<br>Crying duration | Diminution of pain in G1 compared to G2. No difference between groups in crying duration. |
| Romantsik et al., 2014 | RCT, two parallel groups | N= 69 full-term<br>Age= up to 37 weeks (mean not specified) | Heel-prick | G1: n=39, vanilla odor<br>G2: n=30, control group | Pain measured by the NFCS<br>Crying duration | No difference between groups |
| Badiee et al., 2013 | RCT, two parallel groups | N= 50 preterm<br>Age= 32-37 weeks | Heel-prick | G1: n=25, mother milk odor<br>G2: n=25, formula milk odor | Pain measured by the PIPP<br>Crying duration<br>Salivary cortisol level | Mother milk odor is more efficient than formula milk odor on pain score, crying duration and salivary cortisol level. |
| Sadathosseini et al., 2013 | RCT, four parallel groups | N= 135 full-term<br>Age= 37-42 weeks | Arterial puncture | G1: n=45, vanilla odor and 8 hours 45 (mean) of habituation<br>G2: n=45, vanilla odor without habituation<br>G3: n=45, control group | Variations in heart rate and oxygen saturation<br>Crying duration | After the painful procedure, G1 has higher oxygen saturation than G2 and GC. Diminution of crying duration for G1 compared to G2 and GC. |
| Nishitani et al., 2009 | RCT, four parallel groups | N= 48 full-term<br>Age= 38-41 weeks | Heel-prick | G1: n=12, control group<br>G2: n=12, mother milk odor<br>G3: n=15, other mother milk odor<br>G4: n=9, formula milk odor | Crying duration<br>Salivary cortisol level<br>Facial and motor signs in response to pain (not a valid scale) | Diminution of crying duration for G2 compared to G1, G3 and G4. Salivary cortisol level was higher in GC. Facial and motor signs were lower in G2 compared to the other groups. |
| Goubet et al., 2007 | RCT, four parallel groups | N= 44 full-term<br>Age= up to 37 weeks (mean at 39 weeks) | Heel-prick | G1: n=11, vanilla odor and 11 hours (mean) habituation; G2: n=11, vanilla odor on mothers; G3: n=11, vanilla odor without habituation; G4: n=11, control group | Crying duration<br>Facial signs in response to pain (not a valid scale) | Diminution of crying duration for G1 compared to G2, G3 and GC. Diminution of facial signs for G1 and G2 compared to G3 and GC. |

| | | | | | | |
|---|---|---|---|---|---|---|
| Rattaz et al., 2005 | RCT, four parallel groups | N= 44 full-term<br>Age= up to 37 weeks (mean at 39.4 weeks) | Heel-prick | G1: n=11, mother milk odor;<br>G2: n=11, vanilla odor and 16 hours (mean) of habituation;<br>G3: n=11, vanilla odor without habituation; G4: n=11, control group | Crying duration<br>Facial and motor signs in response to pain (not a valid scale) | After the painful procedure, diminution of crying and facial signs for G1 and G2, but not for G3 compared to GC. |
| Goubet et al., 2003 | RCT, six parallel groups | N= 51 preterm<br>Age= 32 weeks (mean) | Venipuncture and heel-prick | *Venipuncture*<br>G1: n=8, vanilla odor and 17 hours 37 (mean) of habituation;<br>G2: n=8, vanilla odor without habituation; G3: n=9, control group<br>*Heel-prick*<br>G4: n=9, vanilla odor and 17 hours 37 (mean) of habituation;<br>G5: n=9, vanilla odor without habituation; G6: n=8, control group | Crying duration<br>Facial signs in response to pain (not a validated scale) | Groups with the familiarization (G1 and G4) have lower crying duration after the painful procedure. |
| Mellier et al., 1997 | Quasi-experimental | N= 69 full-term<br>Age= up to 37 weeks (mean not specified) | Venipuncture | G1: n=21, mother milk odor<br>G2: n=25, formula milk odor<br>G3: n=23, control group | Crying duration<br>Facial and motor signs in response to pain (not a valid scale) | Mother milk odor diminishes 25 to 50% of crying duration after the painful procedure. |
| Kawakami et al., 1997 | RCT, two parallel groups | N= 83 full-term<br>Age= up to 37 weeks (mean not specified) | Heel-prick | G1: n=25, lavender odor<br>G2: n=24, formula milk odor<br>G3: n=34, control group | Cortisol level<br>Facial signs in response to pain (not a valid scale) | No difference between groups. |

**Table 2. Synthesis of results**

| Outcome or subgroup | Studies (n) | Results |
|---|---|---|
| Comparison 1: Familiar odor vs standard care | | |
| **Pain reactivity (all pain scores)** | **5 (n=390)** | **SMD -0.69; 95%CI -0.93 to -0.44; $I^2$ = 20%, *p*<0.00001 (c)** |
| Preterm | 3 (n=234) | SMD -0.64; 95% CI -1.05 to -0.23; $I^2$ = 52%, *p*= 0.002 (**c**) |
| Full-term | 2 (n=156) | SMD -0.73; 95% CI -1.07 to -0.40; $I^2$ = 0%, *p*<0.0001 (**c**) |
| **Pain regulation (all pain scores)** | **4 (n=310)** | **SMD -0.40; 95% CI -0.66 to -0.14; $I^2$ = 13%, *p*= 0.002 (b)** |
| Preterm | 3 (n=234) | SMD -0.42; 95% CI -0.78 to -0.06; $I^2$ = 41%, *p*= 0.02 (**b**) |
| Full-term | 1 (n=76) | SMD -0.32, 95% CI -0.80 to 0.16; $I^2$ = 0%, *p*= 0.19 (**a**) |
| **Crying duration during the procedure** | **7 (n=357)** | **SMD -0.42; 95% CI -0.73 to -0.10; $I^2$ = 47%, *p*= 0.009 (b)** |
| Preterm | 3 (n=131) | SMD -0.20; 95% CI -0.60 to 0.19; $I^2$ = 18%, *p*= 0.31 (**a**) |
| Full-term | 4 (n=226) | SMD -0.59; 95% CI -1.00 to -0.18; $I^2$ = 45%, *p*= 0.005 (**c**) |
| **Crying duration after the procedure** | **5 (n=187)** | **SMD -0.37; 95% CI -0.68 to -0.07; $I^2$ = 0%, *p*= 0.01 (b)** |
| Preterm | 3 (n=131) | SMD -0.41; 95% CI -0.77 to -0.05; $I^2$ = 0%, *p*= 0.02 (**c**) |
| Full-term | 2 (n=56) | SMD -0.34; 95% CI -1.06 to 0.39; $I^2$ = 41%, *p*= 0.36 (**a**) |
| **Heart rate (HR) variation during the procedure** | **3 (n=302)** | **MD -2.29; 95% CI -4.88 to 0.31; $I^2$ = 97%, *p*=0.08 (a)** |
| Preterm | 1 (n=136) | MD -1.95; 95% CI -5.40 to 1.51; *p*= 0.22 (**a**) |
| Full-term | 2 (n=166) | MD -2.76; 95% CI -7.16 to 1.63; $I^2$ = 58%, *p*= 0.22 (**a**) |
| **HR variation after the procedure** | **2 (n=212)** | **MD -3.87; 95% CI -7.36 to -0.38; $I^2$ = 99%, *p*=0.03** |
| Preterm | 1 (n=136) | MD -1.97; 95% CI -6.16 to 2.22; *p*=0.36 (**a**) |
| Full-term | 1 (n=76) | MD -8.39; 95% CI -12.51 to -4.27, *p*<0.0001 |
| **Oxygen saturation (O2) variation during the procedure** | **3 (n=302)** | **MD -0.47; 95% CI -0.86 to -0.08; $I^2$=91%, *p*=0.02** |
| Preterm | 1 (n=136) | MD -0.50; 95% CI -1.05 to 0.05; *p*=0.07 (**a**) |
| Full-term | 2 (n=166) | MD -0.44; 95% CI -1.07 to 0.20; $I^2$=62%, *p*=0.18 (**a**) |
| **O2 variation after the procedure** | **2 (n=212)** | **MD -0.56; 95% CI -0.99 to -0.13; $I^2$=99%, *p*=0.01** |
| Preterm | 1 (n=136) | MD -0.70; 95% CI -1.24 to 0.15; *p*=0.01 (**a**) |
| Full-term | 1 (n=76) | MD -0.34; 95% CI -1.07 to 0.39; p=0.36 (**a**) |
| **Cortisol variation** | **-** | **-** |
| Preterm | 0 (no study) | |
| Full-term | 2 (n=131) | Data not available |
| Comparison 2: Familiar odor vs Artificial odor without habituation | | |
| **Pain reactivity (all pain scores)** | **2 (n=128)** | **SMD -0.28; 95% CI -1.39 to 0.83; $I^2$ = 89%, *p*=0.62 (a)** |
| Preterm | 1 (n=50) | SMD -1.38; 95% CI -2.00 to -0.76; *p*<0.0001 (**d**) |
| Full-term | 1 (n=78) | SMD -0.28; 95% CI -1.39 to 0.83; *p*=0.24 (**a**) |
| **Pain regulation (all pain scores)** | **-** | **-** |
| Preterm | 0 (no study) | |
| Full-term | 1 (n=78) | SMD -0.17; 95% CI -.063 to 0.29; *p*=0.47 (a) |
| **Crying duration during the procedure** | **5 (n=230)** | **SMD -0.49; 95% CI -0.92 to -0.06; $I^2$ = 54%, *p*=0.03 (b)** |
| Preterm | 2 (n=84) | SMD -0.47; 95% CI -1.11 to 0.18; $I^2$ = 53%, *p*=0.16 (**a**) |
| Full-term | 3 (n=146) | SMD -0.46; 95% CI -1.13 to 0.21; $I^2$ = 64%, p=0.18 (**a**) |
| **Crying duration after the procedure** | **3 (n=90)** | **SMD -0.11; 95% CI -0.53 to 0.31; $I^2$=0%, *p*=0.51 (a)** |
| Preterm | 1 (n=34) | SMD -0.05; 95% CI -0.73 to 0.12; *p*=0.88 (**a**) |

| | | |
|---|---|---|
| Full-term | 2 (n=56) | SMD -0.15; 95% CI -0.69 to 0.39; $I^2$ = 0%, p=0.59 (**a**) |
| **HR variation during the procedure-** full-term | 2 (n=168) | MD 0.16; 95% CI -4.54 to 4.87; $I^2$=47%, *p*=0.95 (a) |
| **HR variation after the procedure-** full-term | 1 (n=78) | MD -6.75; 95% CI -12.51 to -0.99; $I^2$=0%, *p*=0.02 (a) |
| **O2 variation during procedure**- full-term | 2 (n=168) | MD 0.10; 95% CI -0.95 to 1.14; $I^2$=86%, *p*=0.86 (a) |
| **O2 variation after the procedure**- full-term | 1 (n=78) | MD 0.38; 95% CI -0.35 to 1.11; $I^2$=56%, *p*=0.30 (a) |
| **Cortisol variation** – full-term | 1 (n=50) | MD -8.70; 95% CI -11.49 to -5.91; *p*<0.00001 |

*Interpretation based of Cohen's d classification: a = no significant difference, b = small effect, c= moderate effect, d= large effect*

**Table 3.** Summary if findings: Familiar odor compared to standard care

| Certainty assessment | | | | | | | Summary of findings | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| № of participants (studies) | Risk of bias | Inconsis-tency | Indirect-ness | Impre-cision | Publica-tion bias | Overall certainty of evidence | Study event rates (%) | | Relative effect (95% CI) | Anticipated absolute effects | |
| | | | | | | | With standard care | With fami-liar odors | | Risk with standard care | Risk difference with familiar odors |
| **Pain reactivity** | | | | | | | | | | | |
| 390 (5 RCTs) | serious [a] | not serious [b] | not serious | serious [c] | none | ⊕⊕◯◯ LOW | 151 | 239 | - | - | SMD **0.68 SD lower** (0.93 lower to 0.44 lower) |
| **Pain regulation** | | | | | | | | | | | |
| 310 (4 RCTs) | serious [a] | not serious [b] | not serious | serious [c] | none | ⊕⊕◯◯ LOW | 111 | 199 | - | - | SMD **0.4 SD lower** (0.66 lower to 0.14 lower) |
| **Crying duration during the procedure** | | | | | | | | | | | |
| 357 (7 RCTs) | serious [a] | serious [d] | not serious | serious [c] | none | ⊕◯◯◯ VERY LOW | 164 | 193 | - | - | SMD **0.42 SD lower** (0.73 lower to 0.1 lower) |
| **Crying duration after the procedure** | | | | | | | | | | | |
| 165 (5 RCTs) | serious [a] | not serious [b] | not serious | serious [c] | none | ⊕⊕◯◯ LOW | 79 | 86 | - | - | SMD **0.41 SD lower** (0.74 lower to 0.09 lower) |
| **Heart rate variation during the procedure** | | | | | | | | | | | |
| 302 (3 RCTs) | serious [a] | very serious [e] | not serious | serious [c] | none | ⊕◯◯◯ VERY LOW | 117 | 185 | - | | MD **2.29 lower** (4.88 lower to 0.3 higher) |
| **Heart rate variation after the procedure** | | | | | | | | | | | |
| 212 (3 RCTs) | serious [a] | very serious [e] | not serious | serious [c] | none | ⊕◯◯◯ VERY LOW | 72 | 140 | - | | MD **3.87 lower** (7.36 lower to 0.38 lower) |
| **Oxygen saturation during the procedure** | | | | | | | | | | | |
| 302 (3 RCTs) | serious [a] | very serious [e] | not serious | serious [c] | none | ⊕◯◯◯ VERY LOW | 117 | 185 | - | | MD **0.47 lower** (0.86 lower to 0.08 lower) |
| **Oxygen saturation after the procedure** | | | | | | | | | | | |
| 212 (2 RCTs) | serious [a] | very serious [e] | not serious | serious [c] | none | ⊕◯◯◯ VERY LOW | 72 | 140 | - | | MD **0.56 lower** (0.99 lower to 0.13 lower) |

**CI:** Confidence interval; **SMD:** Standardised mean difference; **MD:** Mean difference

a. Quality of evidence was downgraded by one level as the majority of risk of bias judgements was rated as "unclear" or "high".
b. Quality of evidence was not downgraded for inconsistency as heterogeneity might not be important (<40%).
c. Quality of evidence was downgraded by one level for imprecision as the total number of participants is less than the threshold for continuous outcomes (<400).
d. Quality of evidence was downgraded by one level for inconsistency due to moderate or substantial heterogeneity among studies (40% to 75%).
e. Quality of evidence was downgraded by two levels for inconsistency due to considerable heterogeneity among studies (75% to 100%).

**Supplemental Table 1**. PubMed search strategy

| | Équation de recherche | Résultats |
|---|---|---|
| 1 | "Odorants"[Mesh] OR "Oils, Volatile"[Mesh] OR "Smell"[Mesh] OR "Perfume"[Mesh] OR "Flower Essences"[Mesh] OR "Aromatherapy"[Mesh] OR "Olfactory Perception"[Mesh] | 37 859 |
| 2 | [Title/Abstract] = (odor* OR smell* OR scent* OR fragranc* OR olfact* OR perfume* OR aroma OR aromatherap* OR "essential oil" OR "essential oils" OR redolence* OR incense*) | 79 183 |
| 3 | "Pain"[Mesh] OR "Pain Management"[Mesh] OR "Pain Perception"[Mesh] OR "Pain Threshold"[Mesh] OR "Pain Measurement"[Mesh] OR "Stress, Physiological"[Mesh] OR "Crying"[Mesh] | 540 711 |
| 4 | [Title/Abstract] = (pain* OR cry OR crying OR cries OR scream* OR suffer* OR tear* OR grimac* OR agitat* OR distress OR stress* OR sooth* OR calm* OR sob* OR weep* OR ache* OR aching OR agony OR agon* OR afflict* OR anguish* OR cramp* OR discomfort OR irritat* OR sore* OR torment OR twinge*) | 1 854 884 |
| 5 | "Infant"[Mesh] OR "Neonatal Nursing"[Mesh] OR "Intensive Care, Neonatal"[Mesh] OR "Intensive Care Units, Neonatal"[Mesh] OR "Neonatology"[Mesh] OR "Neonatologists"[Mesh] | 1 026 963 |
| 6 | [Title/Abstract] = (neonat* OR baby OR babies OR newborn* OR infant* OR (child* AND (premature* OR preterm OR newborn* OR neonat*))) | 658 275 |
| 7 | (#1 OR #2) AND (#3 OR #4) AND (#5 OR #6) | 370 |
| 8 | #7 AND (English[Language] OR French[Language]) | 353 |

**Supplemental Table 2**. Bias in included studies

| | Randomization (random sequence generation) | Allocation bias | Blinding of participants and personnel | Blinding of outcome assessment | Attrition | Selective outcome reporting | Other potential sources of bias |
|---|---|---|---|---|---|---|---|
| **Ackan 2016** | **Unclear risk of bias**<br>Insufficient information about the sequence generation process to permit judgement of Low or High risk. | **Unclear risk of bias**<br>« Newborns were selected for the groups randomly from an opaque envelope."<br>The use of assignment envelopes is described, but it remains unclear whether envelopes were sequentially numbered, opaque and sealed | **Unclear risk of bias**<br>« nurses were not informed which odor sample they would be administering to the infants"<br>It is not clear whether the nurses could tell the difference between the different odors even if they were not informed of the infant's allocation because they used artificial odor (stronger smell). | **Low risk of bias**<br>« Two trained neonatology nurses blinded as to the content of vials<br>evaluated the NIPS of the newborns in all groups from 1<br>minute before the invasive procedure to 1 minute afterward." | **Low risk of bias**<br>No missing outcome data. | **Low risk of bias**<br>The study protocol is not available but it is clear that the published reports include all expected outcomes, including those that were pre-specified | **Unclear risk of bias**<br>It is not clear if the experimental odor can be smelled by the participants in the control group (risk of contamination between groups). |
| **Almedar 2017** | **Unclear risk of bias**<br>"Randomization was performed using a computer program »: Not described in | **Unclear risk of bias**<br>Insufficient information to permit judgement of Low or 'High | **High risk of bias**<br>No blinding. | **High risk of bias**<br>No blinding of outcome assessment, and the outcome measurement is | **Unclear risk of bias**<br>Insufficient reporting of attrition/exclusions to permit judgement (according to the | **Low risk of bias**<br>The study protocol is not available but it is clear that the | **High risk of bias**<br>Had a potential source of bias related to the specific study design used |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | sufficient detail to allow an assessment of its adequacy (i.e. computer random number generator). | risk: allocation concealment not described | | likely to be influenced by lack of blinding | sample size calculation, 123 infants were needed to achieve 0.80 power but only 97 were recruited. 13 lost to follow-up for unknown reasons). | published reports include all expected outcomes, including those that were pre-specified | (no blinding, insufficient information about randomization and allocation). |
| **Badiee 2013** | **Unclear risk of bias** « The random selection of the patients as formula milk and breast milk groups was based on a selected box number from one to 50. Even numbers were allocated to the formula milk group and odd numbers to the breast milk group» Insufficient information to permit judgement of Low or High risk. | **Unclear risk of bias** Insufficient information to permit judgement of Low or High risk. | **Unclear risk of bias** Insufficient information to permit judgment of Low or High risk. | **Unclear risk of bias** Insufficient information to permit judgment of Low or High risk. | **Unclear risk of bias** Insufficient information to permit judgment of Low or High risk. | **Low risk of bias** The study protocol is not available but it is clear that the published reports include all expected outcomes, including those that were pre-specified. | **Unclear risk of bias** Insufficient information to assess whether an important risk of bias exists. |
| **Baudesson de Chanville 2017** | **Low risk of bias** « Computer generated, randomized lists were provided by the Clinical Research Unit (KB) | **Unclear risk of bias** Insufficient information to permit judgement of | **Low risk of bias** « To ensure blinding, only the chief professional nurse was | **Low risk of bias** « The video was then viewed and analyzed by physicians blinded to the treatment group. Analysis was carried out independently by two | **Low risk of bias** No missing data. | **Low risk of bias** The study protocol is not available but it is clear that the | **Unclear risk of bias** Insufficient information to assess whether an important |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | before the beginning of the study with a permuted randomization scheme (block size 4, randomization ratio 1:1)» | Low or High risk. | aware of the neonates' group allocations. A unit professional nurse performed the venipuncture with a 20G needle while following a standardized procedure and was blinded to the allocation group ». We quote low risk of bais because a natural odor were used (not a strong smell). | teams, each composed of a senior and a junior physician. » | | published reports include all expected outcomes, including those that were pre-specified. | risk of bias exists |
| **Goubet 2003** | **Unclear risk of bias**<br>Insufficient information about the sequence generation process to permit judgement of Low or High risk. | **Unclear risk of bias**<br>The method of concealment is not described | **Unclear risk of bias**<br>Insufficient information to permit judgement of Low or High risk. | **Low risk of bias**<br>"Videotapes were coded by two independent observers blind to group assignment (FAM,NFAM,CONT), but not to condition (heelstick or venipuncture) or intervals (baseline, blood collection, and postcollection)." | **Unclear risk of bias**<br>Insufficient information to permit judgement of Low or High risk. | **Low risk of bias**<br>The study protocol is not available but it is clear that the published reports include all expected outcomes, including those that | **High risk of bias**<br>Risk of contamination because the odor of vanilla is strong.<br>The measure of pain was not with a feasible and reliable tool (detection bias). |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | | were pre-specified. | |
| **Goubet 2007** | **Low risk of bias** « Infants were randomly assigned to one of four groups using a random-numbers table.» | **Unclear risk of bias** Insufficient information to permit judgement of 'Low risk' or 'High risk'. | **Unclear risk of bias** Insufficient information to permit judgment of 'Low risk' or 'High risk'; | **Low risk of bias** « The two coders were unaware of group assignment and types of stimuli. » | **Unclear risk of bias** Insufficient information to permit judgment | **Low risk of bias** The study protocol is not available but it is clear that the published reports include all expected outcomes, including those that were pre-specified. | **High risk of bias** Risk of contamination because the odor of vanilla is strong. The measure of pain was not with a feasible and reliable tool (detection bias). |
| **Jebreilli 2015** | **Low risk of bias** « Infants who were eligible for inclusion in the study were assigned randomly into three groups using Rand List software. » | **Unclear risk of bias** Insufficient information to permit judgement of 'Low risk' or 'High risk': the method of concealment is not described | **Unclear risk of bias** Insufficient information to permit judgment of 'Low risk' or 'High risk'; (the odor of vanilla is strong so blinding seems difficult). | **Low risk of bias** "Two observers watched the recorded videos, measured the infants' grimacing time during sampling period, and expressed the results as percentages of measured times to sampling duration time. Observers did not know to what odor the infants were exposed." | **Unclear risk of bias** Insufficient information to permit judgment | **Low risk of bias** The study protocol is not available but it is clear that the published reports include all expected outcomes, including those that were pre-specified. | **High risk of bias** Risk of contamination because the odor of vanilla is strong. |
| **Kawakami 1997** | **Unclear risk of bias** | **Unclear risk of bias** | **Unclear risk of bias** | **Low risk of bias** "The videotaped segments of the heel | **High risk of bias** 'As-treated' analysis done with | **Low risk of bias** | **High risk of bias** |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Insufficient information about the sequence generation process to permit judgement of 'Low risk' or 'High risk'. | Insufficient information to permit judgement of 'Low risk' or 'High risk' : the method of concealment is not described | Insufficient information to permit judgment of 'Low risk' or 'High risk'; | stick procedure were coded for infant's peak facial and vocal expression during 5-s intervals. (…) All data were independently coded by two coders unaware of the experimental hypothesis." | substantial departure of the intervention received from that assigned at randomization. | The study protocol is not available but it is clear that the published reports include all expected outcomes, including those that were pre-specified. | The measure of pain was not with a feasible and reliable tool (detection bias). |
| **Mellier 1997** | **High risk of bias**<br>No randomisation. | **High risk of bias**<br>No randomisation. | **Unclear risk of bias**<br>Insufficient information to permit judgment of 'Low risk' or 'High risk'; | **Low risk of bias**<br>Blinding of outcome assessment ensured, and unlikely that the blinding could have been broken. | **Unclear risk of bias**<br>Insufficient information to permit judgment of Low or High risk. | **Low risk of bias**<br>The study protocol is not available but it is clear that the published reports include all expected outcomes, including those that were pre-specified. | **High risk of bias**<br>Had a potential source of bias related to the specific study design used (quasi-experimental design, no randomization nor allocation). The pain was not measured by a reliable and valid tool (detection bias). |
| **Neshat 2016** | **Low risk of bias**<br>« Infants who were eligible for inclusion in the | **Unclear risk of bias**<br>Insufficient information to | **Unclear risk of bias**<br>Insufficient information to | **Unclear risk of bias**<br>"Heart rate, blood oxygen saturation, and sampling | **Unclear risk of bias**<br>Insufficient information to permit judgment of | **Low risk of bias**<br>The study protocol is | **High risk of bias**<br>Risk of contamination |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | study were randomly assigned into three groups using the Rand List software.» | permit judgement of 'Low risk' or 'High risk': the method of concealment is not described | permit judgment of 'Low risk' or 'High risk'; (the odor of vanilla is strong so blinding seems difficult). | period were registered by two cameras (Olympus, Center Valley, PA, USA). " It is not clear if the coders were aware of the odors or group assignment. | 'Low risk' or 'High risk' | not available but it is clear that the published reports include all expected outcomes, including those that were pre-specified. | because the odor of vanilla is strong. |
| **Nishitani 2009** | **Unclear risk of bias** The authors only write: "Infants were randomly assigned to the following four groups". Insufficient information about the sequence generation process to permit judgement of 'Low risk' or 'High risk'. | **Unclear risk of bias** Insufficient information to permit judgement of 'Low risk' or 'High risk'.: the method of concealment is not described | **Low risk of bias** « the attending physician was not informed what odor would be offered ». (The odors were breast milk and formula milk: similar look and minimal smell ). | **Low risk of bias** « The coder was blind to the group assignment and type of stimulus". | **Unclear risk of bias** Insufficient information to permit judgment of 'Low risk' or 'High risk'. | **Low risk of bias** The study protocol is not available but it is clear that the published reports include all expected outcomes, including those that were pre-specified. | **Low risk of bias** The study appears to be free of other sources of bias. |
| **Rattaz 2005** | **Low risk of bias** « infants were assigned to one of four study groups using random number tables. " | **Unclear risk of bias** Insufficient information to permit judgement of 'Low risk' or 'High risk'.: the method of | **Unclear risk of bias** The odor of vanilla is strong so blinding seems difficult. | **Low risk of bias** « Coders were blind to group assignment and type of stimuli. » | **High risk of bias** 'As-treated' analysis done with substantial departure of the intervention received from that assigned at randomization. | **High risk of bias** The study report fails to include results for a key outcome that would be expected to | **High risk of bias** Risk of contamination because the odor of vanilla is strong. The measure of pain was not |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | concealment is not described | | | | have been reported for such a study (authors change their hypothesis between the background and the results sections). | with a feasible and reliable tool (detection bias). |
| **Razaghi 2015** | **Low risk of bias**<br>« the name of each group was written on a separate piece of paper, then was lottery ". | **Low risk of bias**<br>« Because it was probable for the control group to be exposed to the lavender scent spread from aromatherapy group (diffusion effect), sampling was done only for one group during each week, with groups being randomly selected; in other words, on the first day, the name of each group was written on a | **Unclear risk of bias**<br>Insufficient information to permit judgement. | **Unclear risk of bias**<br>Insufficient information to permit judgement. | **Low risk of bias**<br>No missing outcome data. | **Low risk of bias**<br>The study protocol is not available but it is clear that the published reports include all expected outcomes, including those that were pre-specified. | **Unclear risk of bias**<br>Insufficient information to assess whether an important risk of bias exists. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | separate piece of paper, then was lottery.» | | | | | |
| **Romantsik 2014** | **Unclear risk of bias** Insufficient information about the sequence generation process to permit judgement of 'Low risk' or 'High risk'. | **Unclear risk of bias** Insufficient information to permit judgement of 'Low risk' or 'High risk': the method of concealment is not described or | **High risk of bias** The odor of vanilla is strong so blinding seems difficult." The assisting nurse perceived the odour during the blood sampling procedure at a distance of around 50 cm from the odourised pad." | **Low risk of bias** « Crying duration, infant's state, facial expressions and hand movements were analysed from coded tapes offline, using a video cassette recorder in frame-by-frame mode, as well as in real time, by two independent observers who were blind to the exposure conditions and child gender. » | **Low risk of bias** No missing outcome data. | **Low risk of bias** The study protocol is not available but it is clear that the published reports include all expected outcomes, including those that were pre-specified. | **Hight risk of bias** Insufficient information to assess whether an important risk of bias exists (risk of contamination because the odor of vanilla is strong). |
| **Sadathossei ni 2013** | **Low risk of bias** « sampling was done only for one group during each week, with groups being randomly selected; in other words, on the first day, the name of each group was written on a separate piece of paper, which was enclosed in a envelope, and an independent nurse who worked in the | **Low risk of bias** the first day, the name of each group was written on a separate piece of paper, which was enclosed in a envelope, and an independent nurse who worked in the clinical context and was unaware of purpose of the | **Low risk of bias** « Throughout the duration of the study, the researcher who performed familiarization and olfactory stimulation was the only person who had knowledge of group assignments » | **Low risk of bias** « two trained research assistants who were blinded to the allocation assignment and had no contact with participants » | **Unclear risk of bias** The study did not address this outcome. | **Low risk of bias** The study protocol is not available but it is clear that the published reports include all expected outcomes, including those that were pre-specified. | **Low risk of bias** The study appears to be free of other sources of bias. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | clinical context and was unaware of purpose of the study picked up one envelope randomly as the first group. Then, selection was performed in the same way for the second group, and the remaining envelope was considered to be the third group.» | study picked up one envelope randomly». | | | | | |

**Supplemental Table 3**. Summary of findings: Familiar odor compared to artificial odor without habituation

| Certainty assessment | | | | | | | Summary of findings | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| № of participants (studies) Follow-up | Risk of bias | Inconsistency | Indirectness | Imprecision | Publication bias | Overall certainty of evidence | Study event rates (%) | | Relative effect (95% CI) | Anticipated absolute effects | |
| | | | | | | | With unfamiliar odors | With familiar odors | | Risk with unfamiliar odors | Risk difference with familiar odors |
| **Pain reactivity** | | | | | | | | | | | |
| 128 (2 RCTs) | serious [a] | very serious [b] | not serious | very serious [c] | none | ⨁◯◯◯ VERY LOW | 53 | 75 | - | - | SMD **0.28 SD lower** (1.39 lower to 0.83 higher) |
| **Pain regulation** | | | | | | | | | | | |
| 78 (1 RCT) | serious [a] | not serious [d] | not serious | very serious [c] | none | ⨁◯◯◯ VERY LOW | 28 | 50 | - | - | SMD **0.17 SD lower** (0.63 lower to 0.29 higher) |
| **Crying duration during the procedure** | | | | | | | | | | | |
| 230 (5 RCTs) | serious [a] | serious [e] | not serious | serious [f] | none | ⨁◯◯◯ VERY LOW | 111 | 119 | - | - | SMD **0.49 SD lower** (0.92 lower to 0.06 lower) |

| Certainty assessment | | | | | | | Summary of findings | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Crying duration after the procedure** | | | | | | | | | | | |
| 90 (3 RCTs) | serious [a] | not serious [d] | not serious | very serious [c] | none | ⨁◯◯◯ VERY LOW | 41 | 49 | - | - | SMD **0.11 SD lower** (0.53 lower to 0.31 lower) |
| **Heart rate variation during the procedure** | | | | | | | | | | | |
| 168 (2 RCTs) | not serious [g] | serious [e] | not serious | serious [f] | none | ⨁⨁◯◯ LOW | 73 | 95 | - | | MD **0.15 higher** (4.53 lower to 4.83 higher) |
| **Heart rate variation after the procedure** | | | | | | | | | | | |
| 78 (1 RCT) | serious [a] | not serious [d] | not serious | very serious [c] | none | ⨁◯◯◯ VERY LOW | 28 | 50 | - | | MD **6.77 lower** (12.52 lower to 1.01 lower) |
| **Oxygen saturation variation during the procedure** | | | | | | | | | | | |
| 168 (2 RCTs) | not serious [g] | very serious [b] | not serious | serious [f] | none | ⨁◯◯◯ VERY LOW | 73 | 95 | - | | MD **0.1 higher** (0.95 lower to 1.14 higher) |
| **Oxygen saturation variation after the procedure** | | | | | | | | | | | |

| Certainty assessment | | | | | | | Summary of findings | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 78 (1 RCT) | serious [a] | serious [e] | not serious | very serious [c] | none | ⨁◯◯◯ VERY LOW | 28 | 50 | - | | MD **0.38 higher** (0.35 lower to 1.11 higher) |

**CI:** Confidence interval; **SMD:** Standardised mean difference; **MD:** Mean difference

*Explanations*

a. Quality of evidence was downgraded by one level as the majority of risk of bias judgements was rated as "unclear" or "high".
b. Quality of evidence was downgraded by two levels for inconsistency due to considerable heterogeneity among studies (75% to 100%).
c. Quality of evidence was downgraded by two levels as the total number of participants is less than the threshold for imprecision (<150).
d. Quality of evidence was not downgraded for inconsistency as heterogeneity might not be important (<40%).
e. Quality of evidence was downgraded by one level for inconsistency due to moderate or substantial heterogeneity among studies (40% to 75%).
f. Quality of evidence was downgraded by one level for imprecision as the total number of participants is less than the threshold for continuous outcomes (<400).
g. Quality of evidence was not downgraded as the majority of risk of bias judgements was rated as "low".

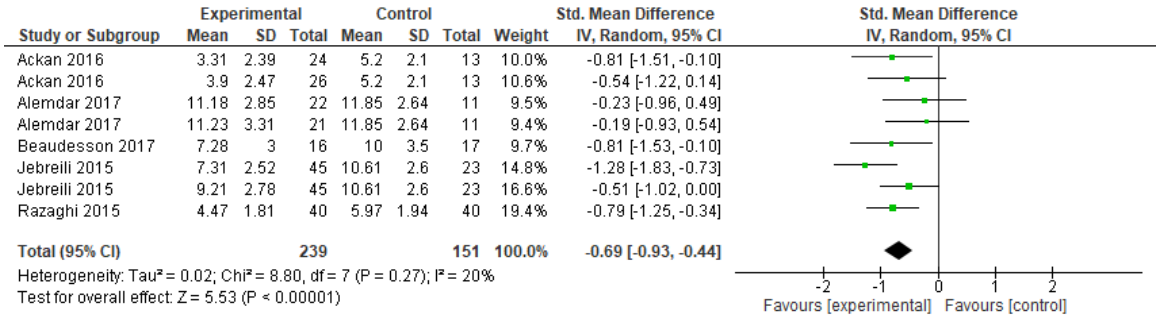


**Figure S1.** Familiar odor vs standard care on pain reactivity

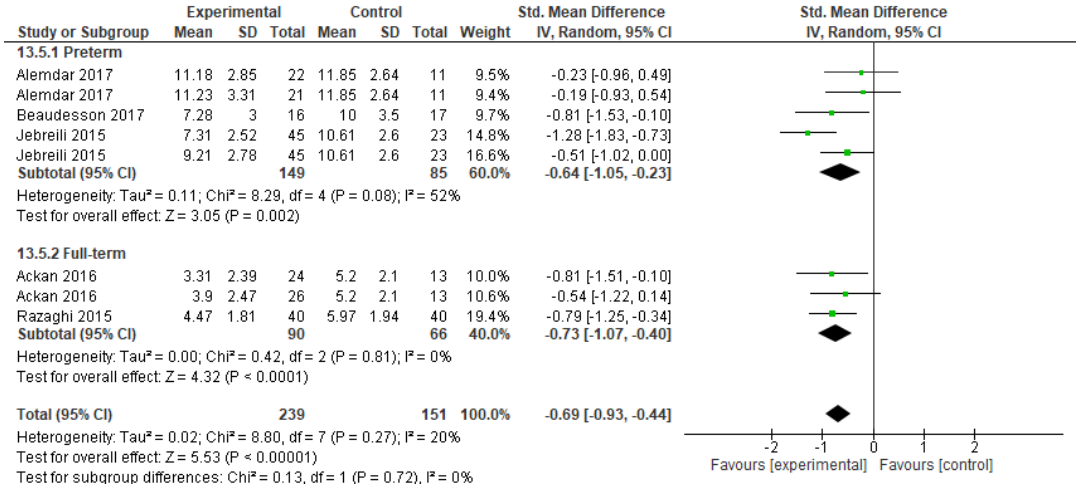


**Figure S2.** Subgroup analysis (population) for familiar odor vs standard care on pain reactivity

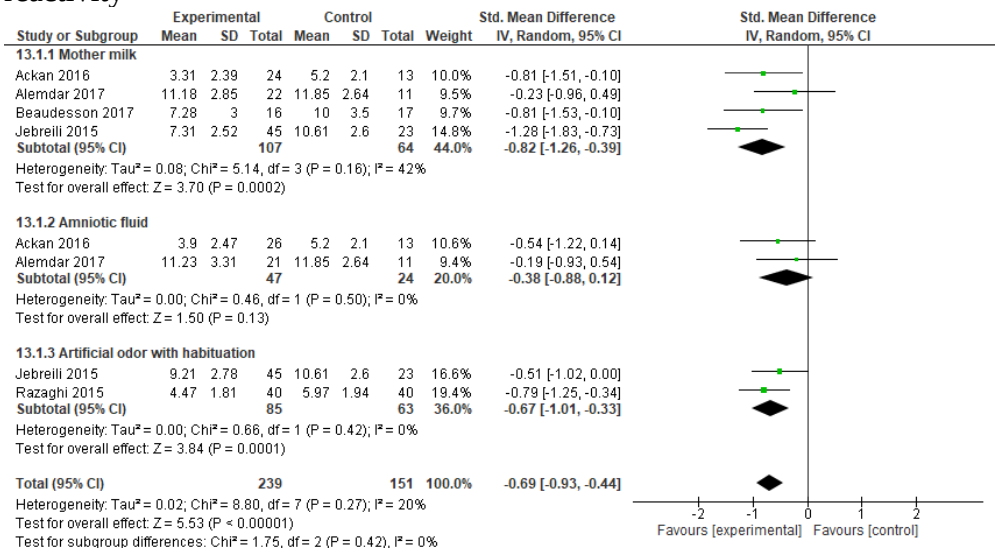


**Figure S3.** Subgroup analysis for familiar odor vs standard care on pain reactivity

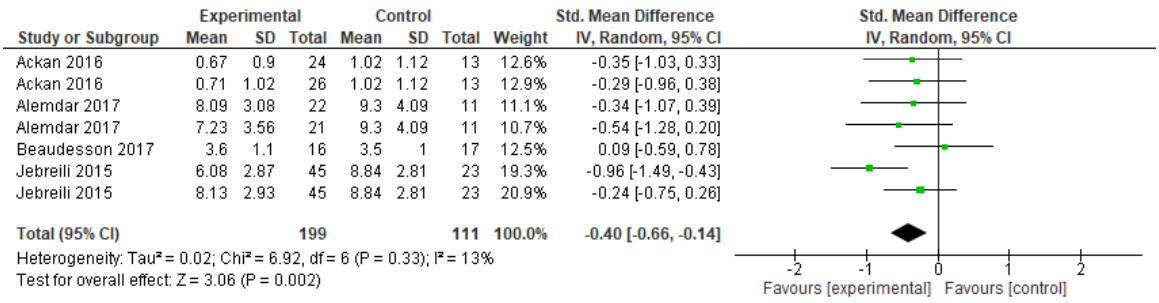


**Figure S4**. Familiar odor vs standard care on pain regulation

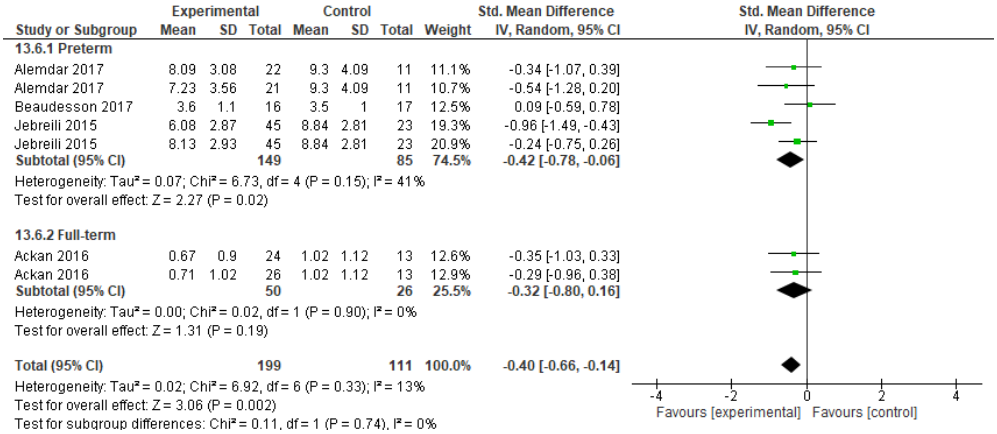


**Figure S5.** Subgroup analysis (population) for familiar odor vs standard care on pain regulation

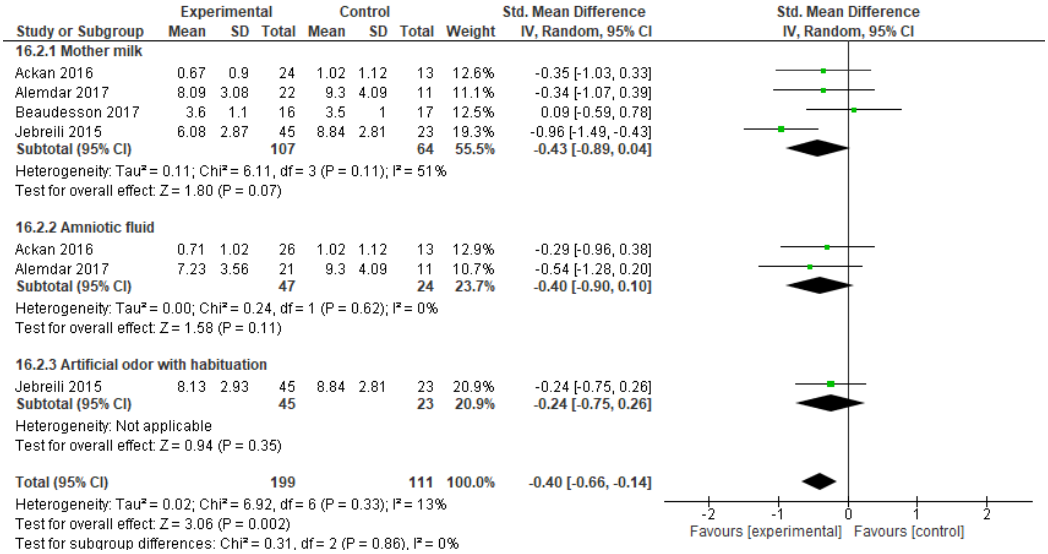


**Figure S6.** Subgroup analysis (type of odors) for familiar odor vs standard care on pain regulation

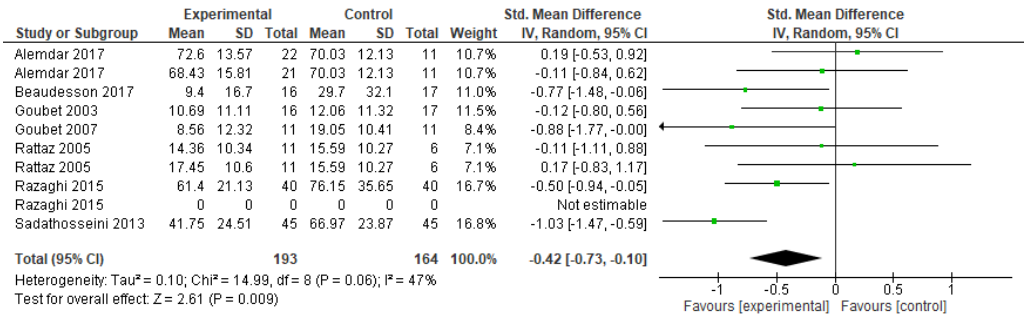


**Figure S7**. Familiar odor vs standard care on crying duration during the procedure

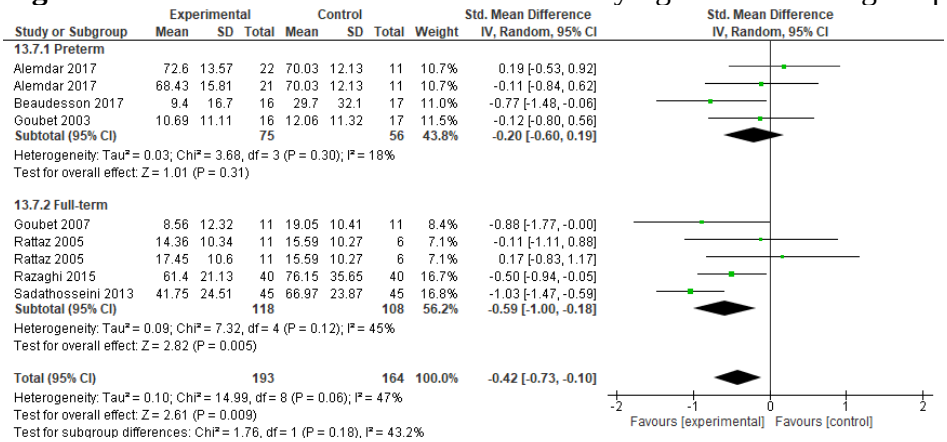


**Figure S8.** Subgroup analysis (population) for familiar odor vs standard care on crying duration during the procedure

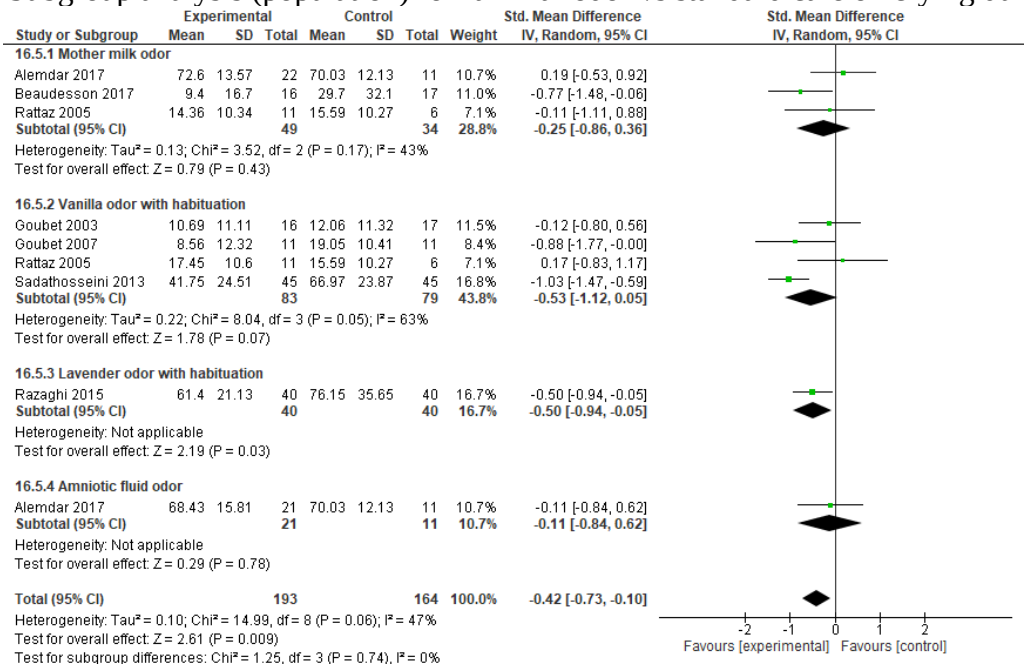


**Figure S9.** Subgroup analysis (type of odors) for familiar odor vs standard care on crying duration during the procedure

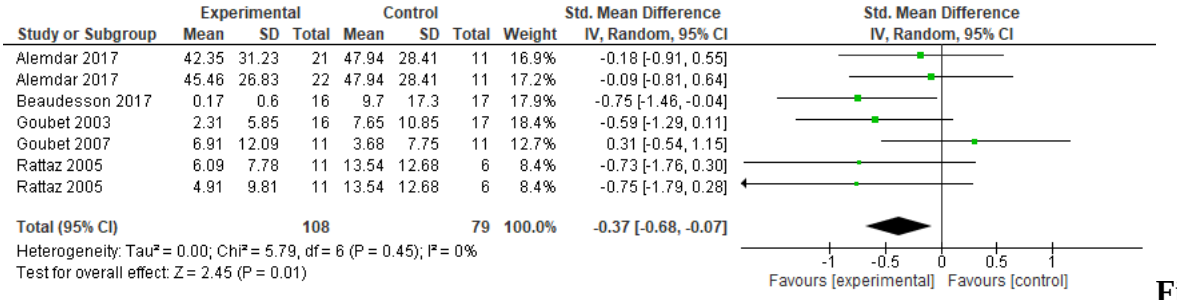


**Figure S10**.

Familiar odor vs standard care on crying duration after the procedure

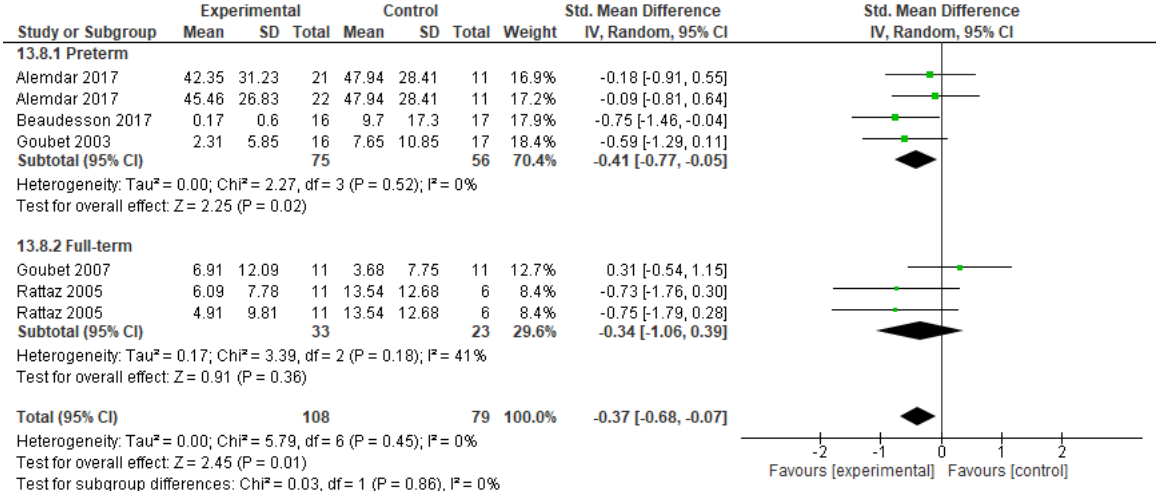


**Figure S11.** Subgroup analysis (population) for familiar odor vs standard care on crying duration after the procedure

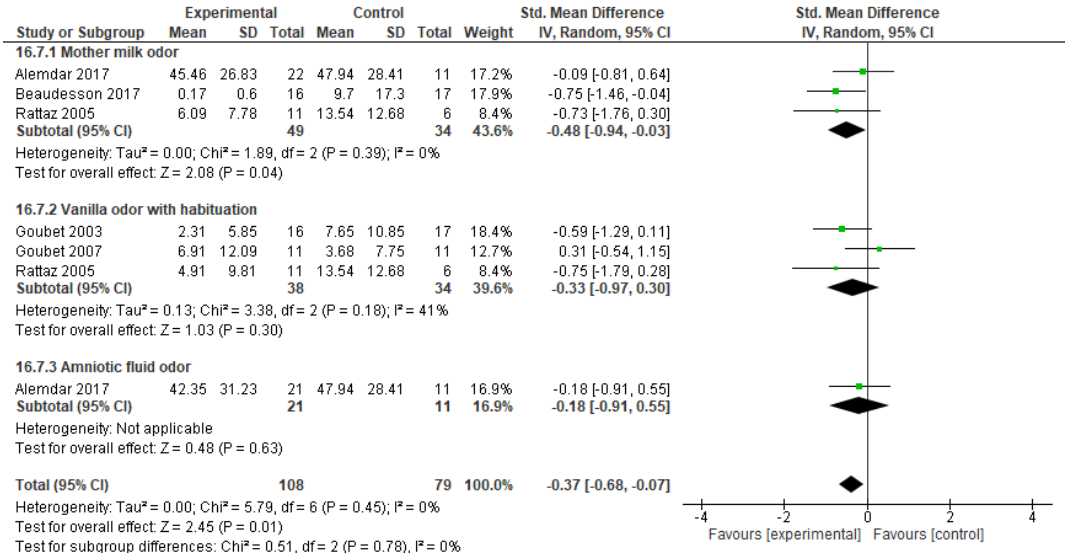


**Figure S12.** Subgroup analysis (type of odors) for familiar odor vs standard care on crying duration after the procedure

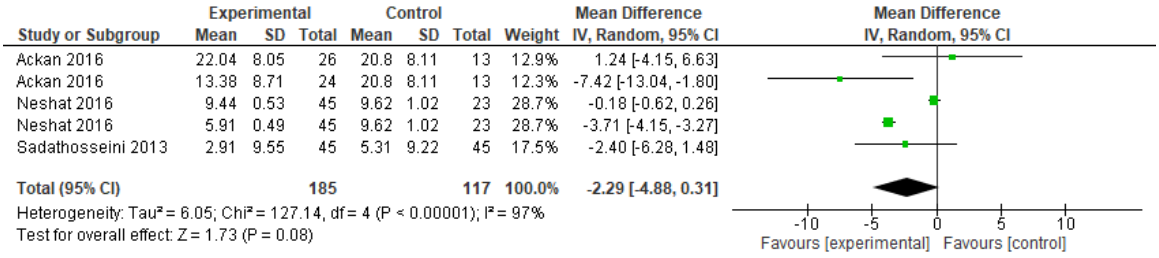


**Figure S13**. Familiar odor vs standard care on heart rate variations during the procedure

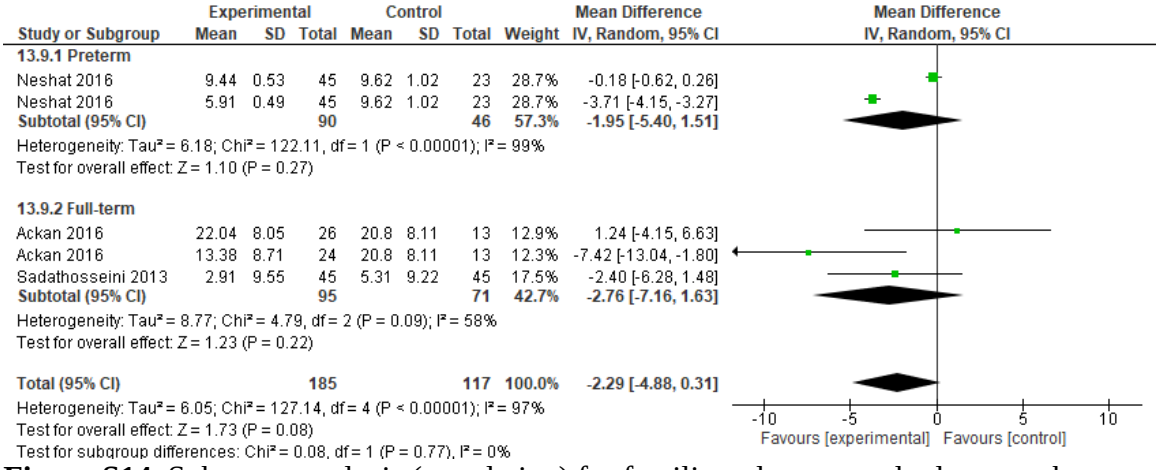


**Figure S14.** Subgroup analysis (population) for familiar odor vs standard care on heart rate variations during the procedure

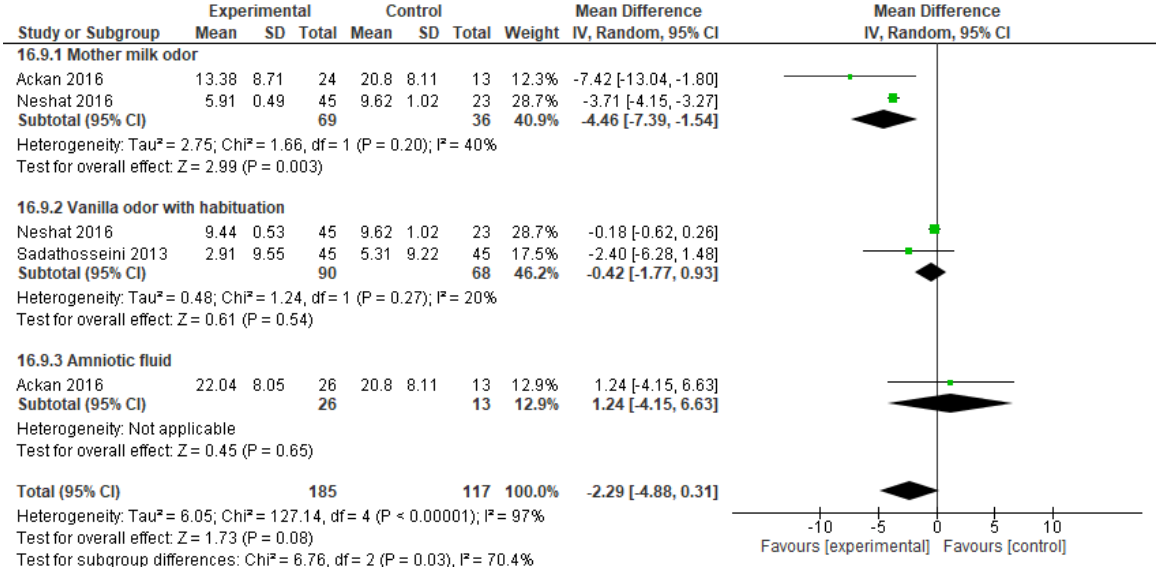


**Figure S15.** Subgroup analysis (type of odors) for familiar odor vs standard care on heart rate variations during the procedure

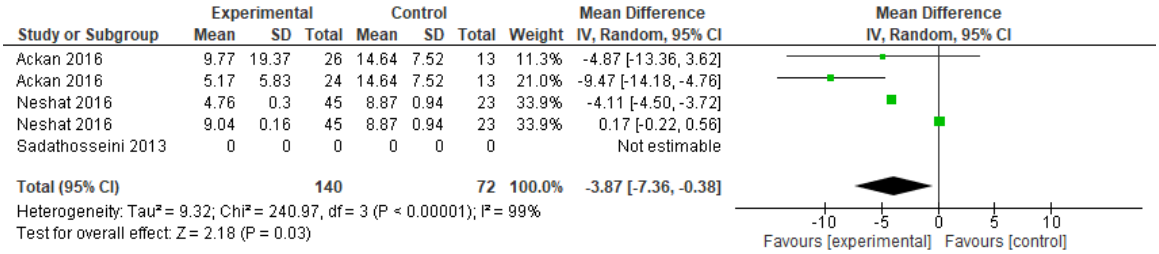


**Figure S16**. Familiar odor vs standard care on heart rate variations after the procedure

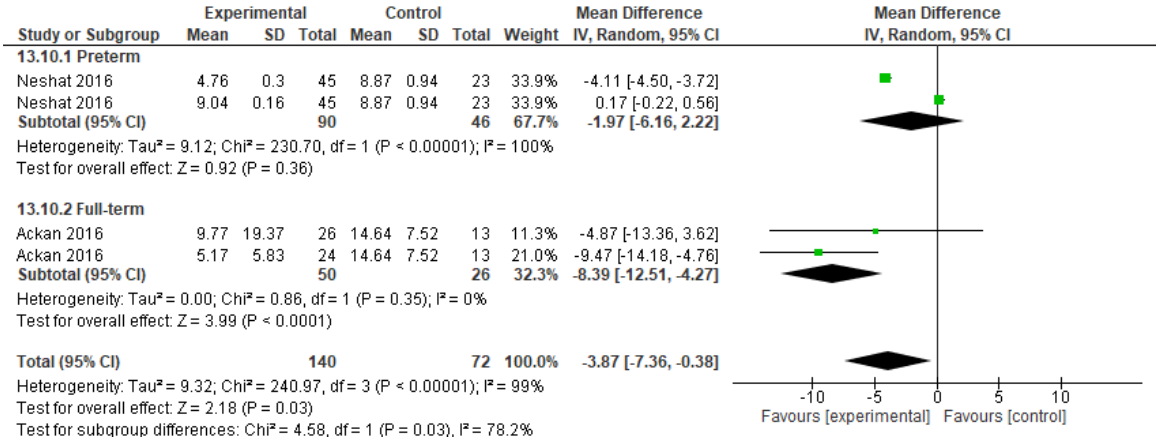


**Figure 17.** Subgroup analysis (population) for familiar odor vs standard care on heart rate variations after the procedure

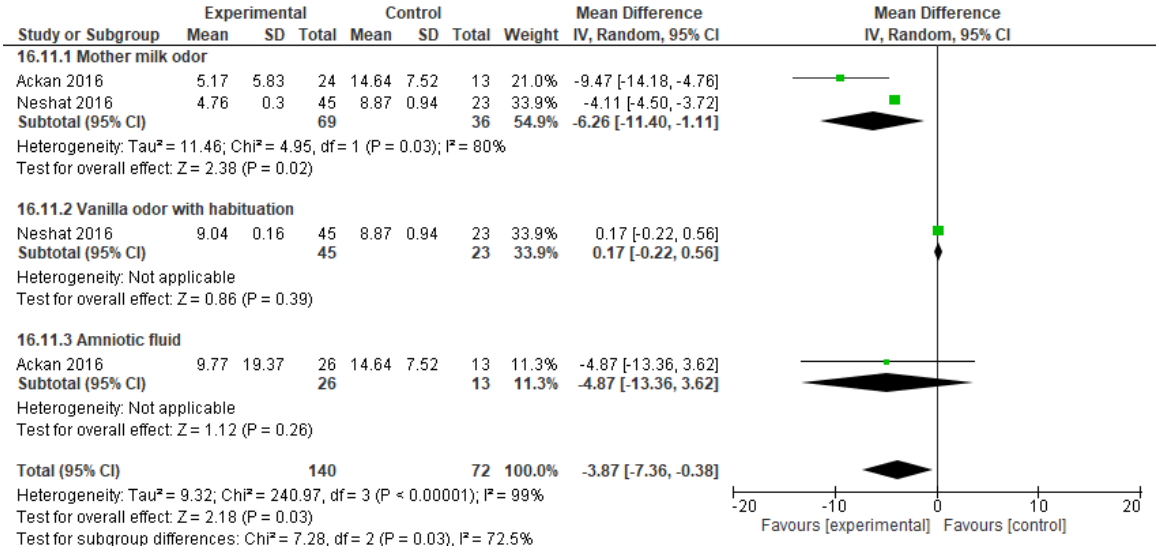


**Figure 18.** Subgroup analysis (type of odors) for familiar odor vs standard care on heart rate variations after the procedure

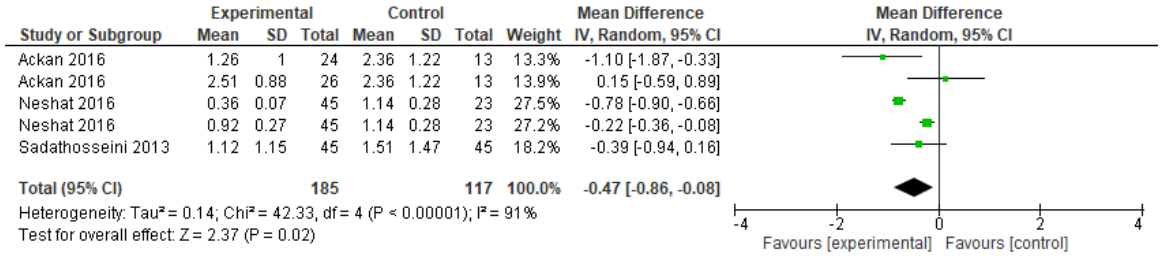


**Figure S19**. Familiar odor vs standard care on oxygen saturation variations during the procedure

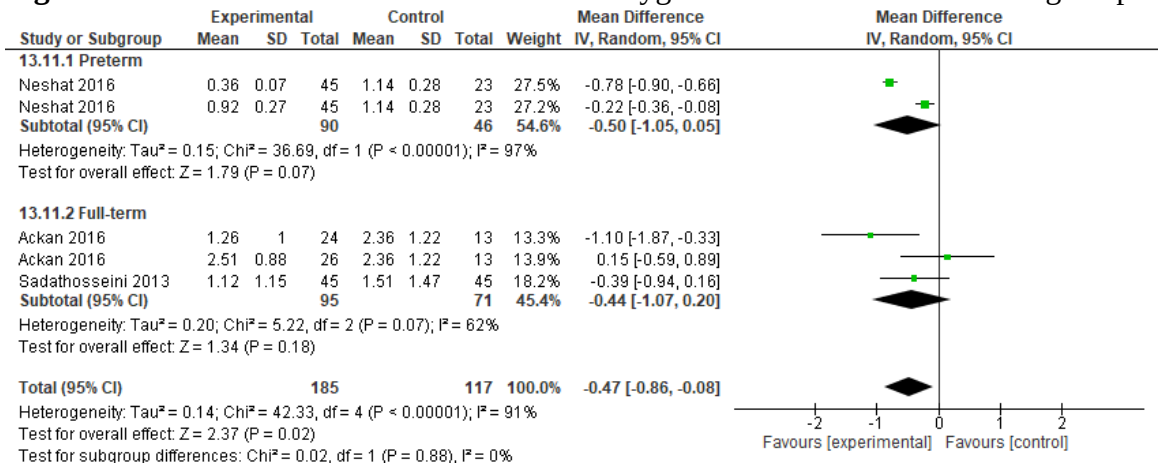


**Figure S20.** Subgroup analysis (population) for familiar odor vs standard care on oxygen saturation variations during the procedure

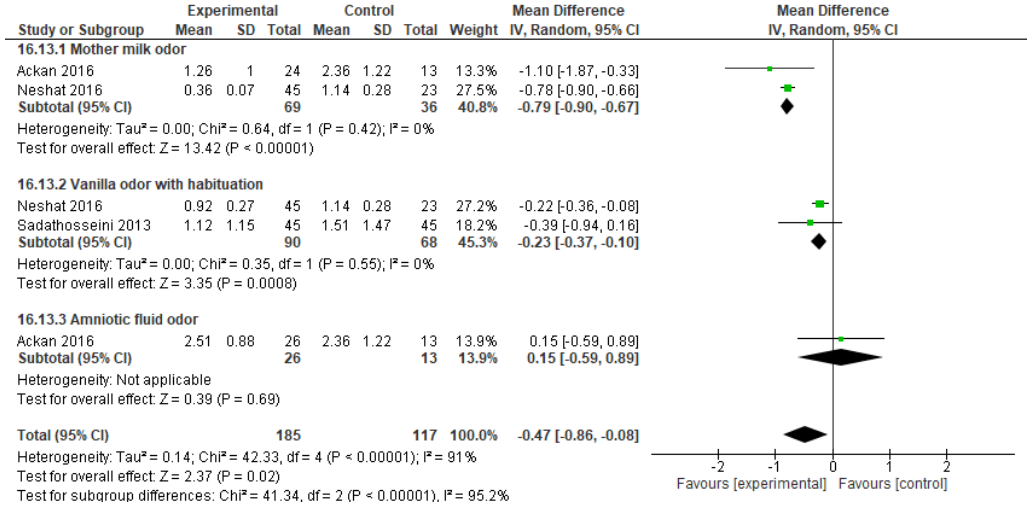


**Figure S21.** Subgroup analysis (type of odors) for familiar odor vs standard care on oxygen saturation variations during the procedure

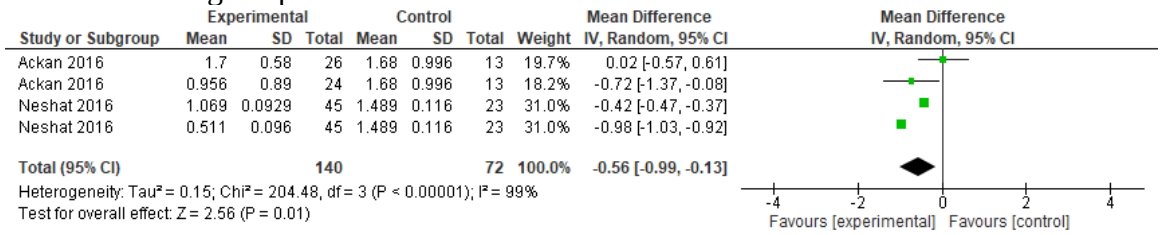

**Figure S22**. Familiar odor vs standard care on oxygen saturation variations after the procedure

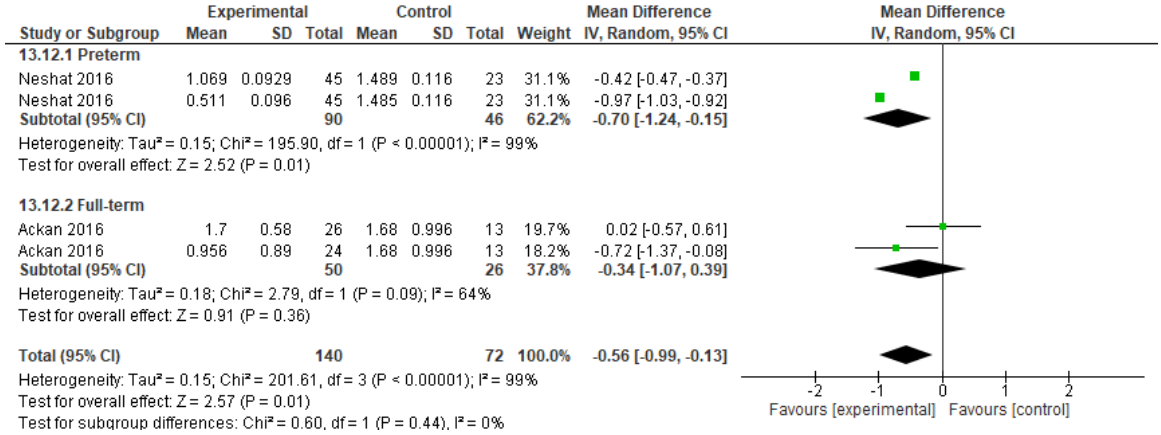


**Figure S23.** Subgroup analysis (population) for familiar odor vs standard care on oxygen saturation variations after the procedure

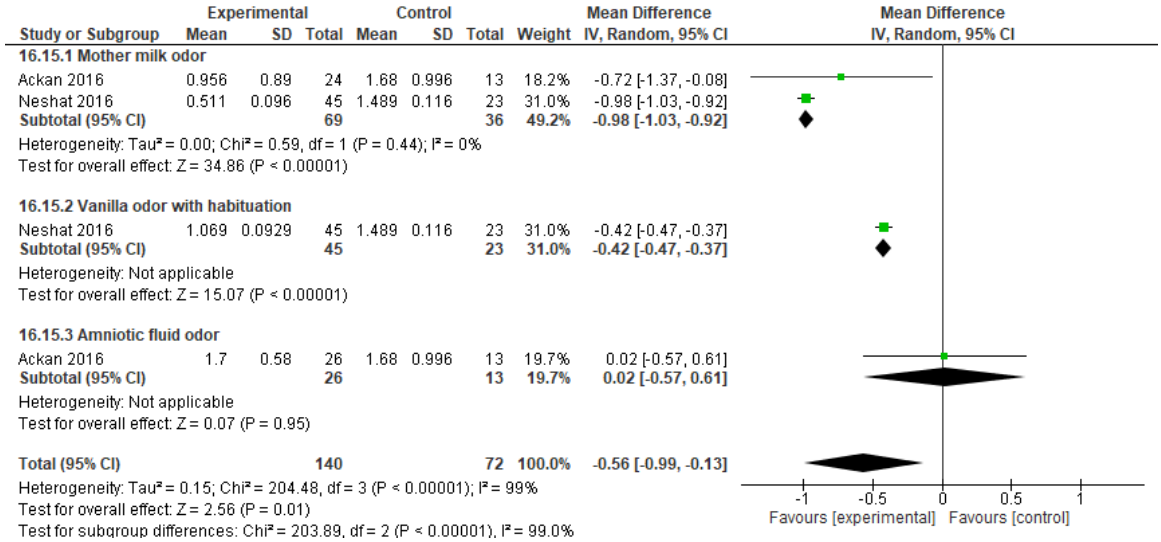


**Figure S24.** Subgroup analysis (type of odors) for familiar odor vs standard care on oxygen saturation variations after the procedure

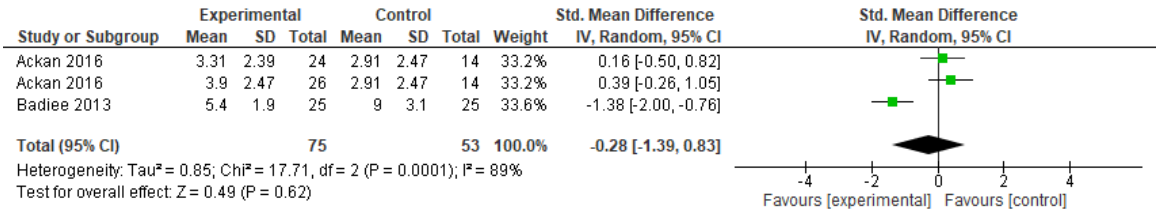


**Figure S25**. Familiar odor vs artificial odor without habituation (unfamiliar odor) on pain reactivity

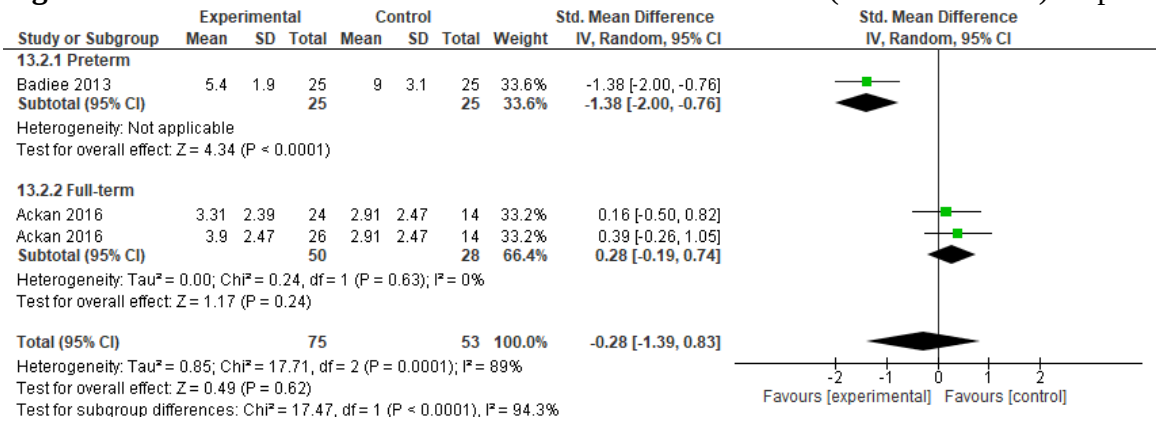


**Figure S26.** Subgroup analysis (populations) for familiar odor vs artificial odor without habituation (unfamiliar odor) on pain reactivity

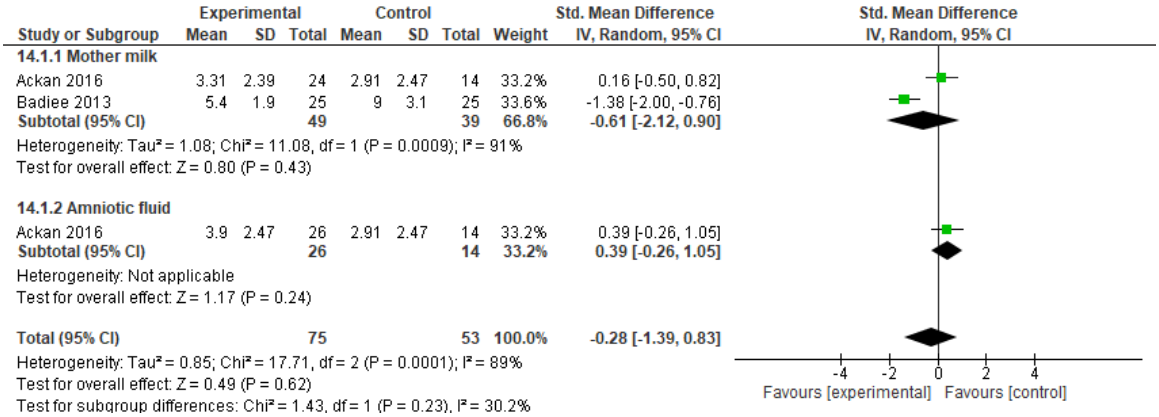


**Figure S27.** Subgroup analysis (type of odors) for familiar odor vs artificial odor without habituation (unfamiliar odor) on pain reactivity

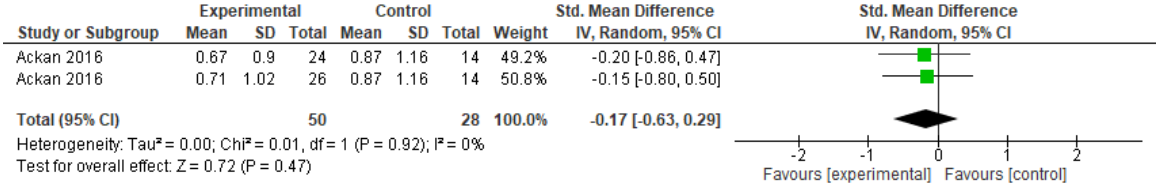


**Figure S28**. Familiar odor vs artificial odor without habituation on pain regulation

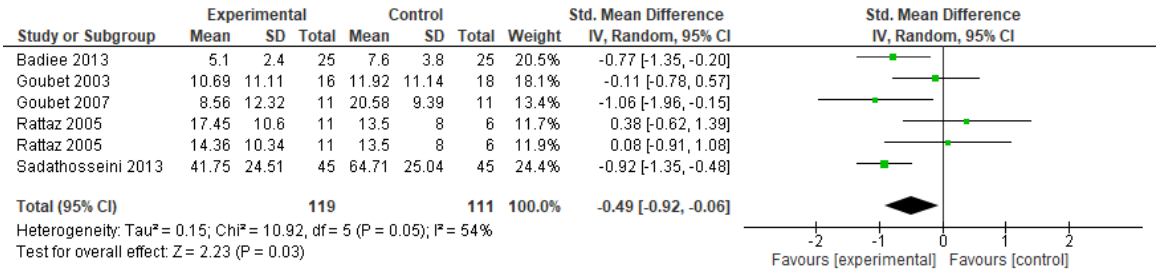


**Figure S29**. Familiar odor vs artificial odor without habituation on crying duration during the procedure

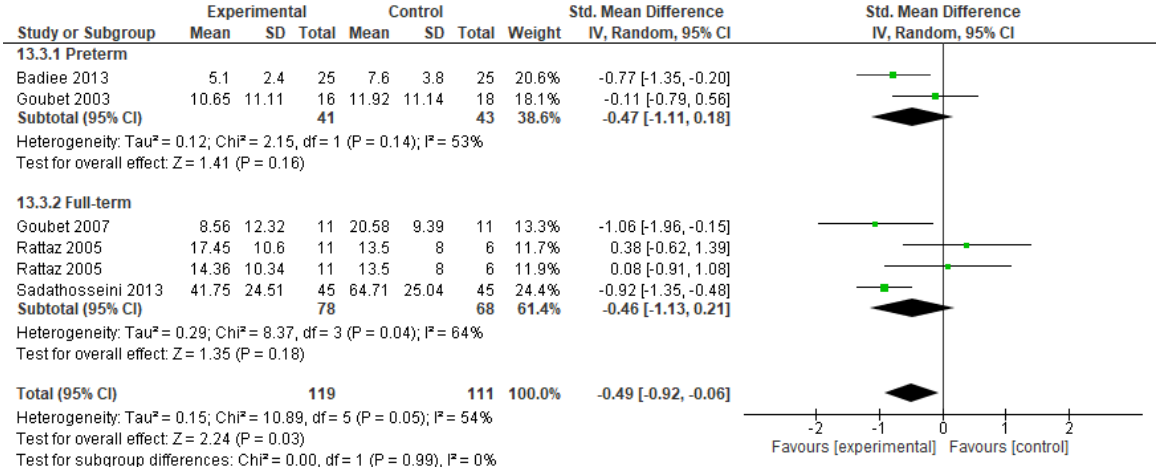


**Figure S30.** Subgroup analysis (populations) for familiar odor vs artificial odor without habituation on crying duration during the procedure

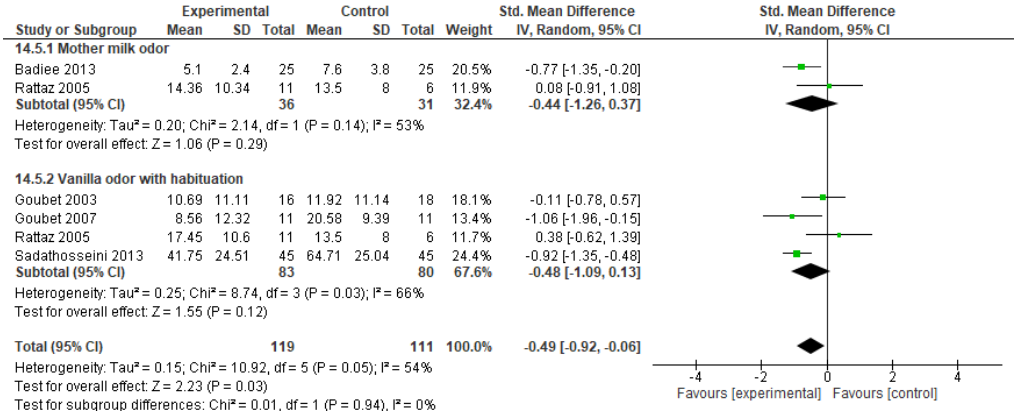


**Figure S31.** Subgroup analysis (type of odors) for familiar odor vs artificial odor without habituation on crying duration during the procedure

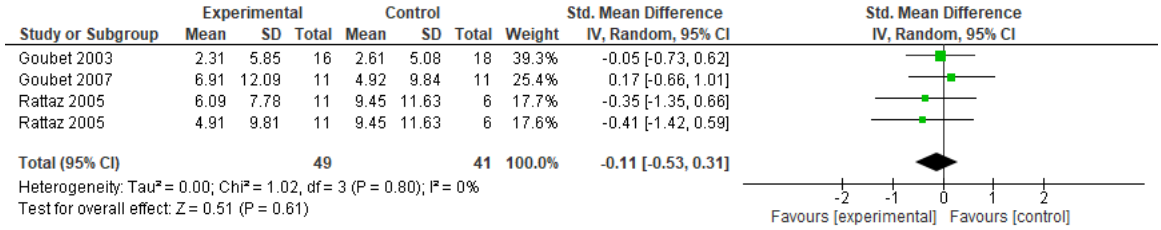


**Figure S32**. Familiar odor vs artificial odor without habituation on crying duration after the procedure

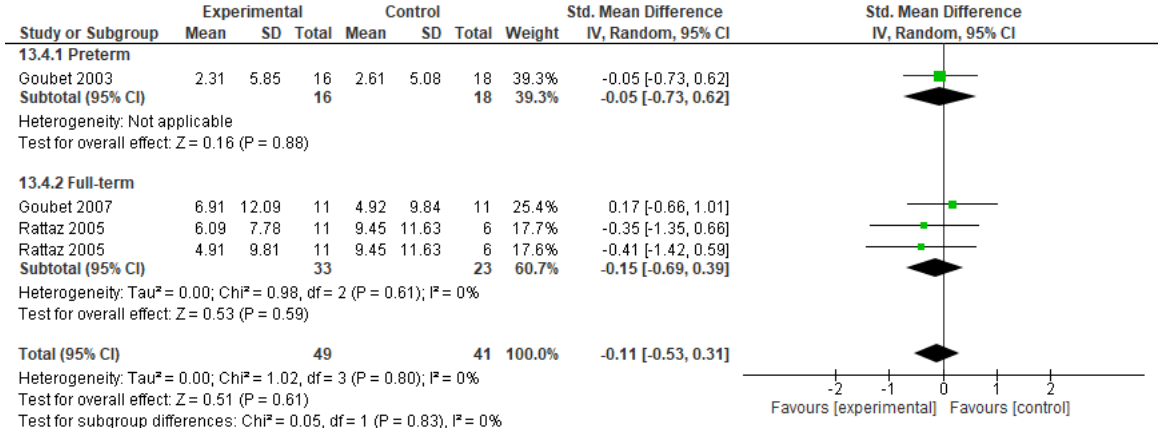


**Figure S33.** Subgroup analysis (populations) for familiar odor vs artificial odor without habituation on crying duration after the procedure

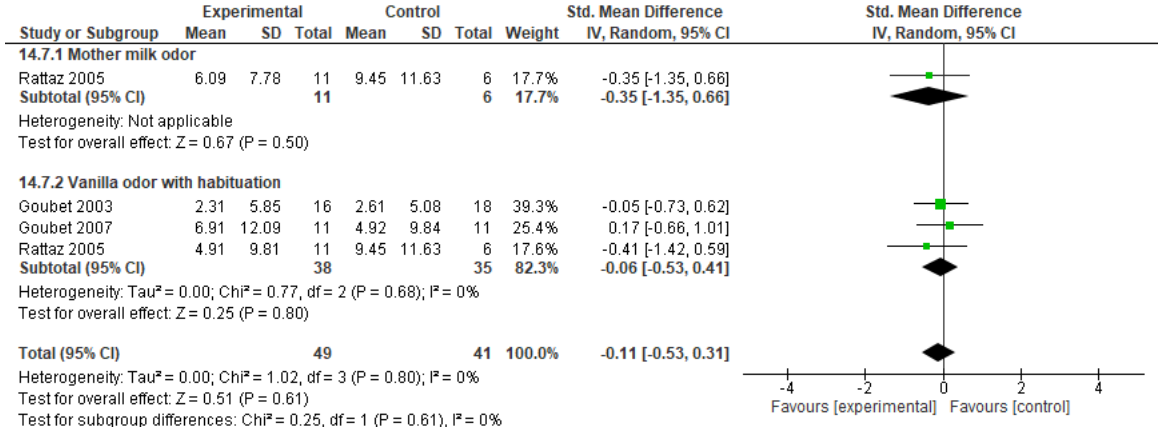


**Figure S34.** Subgroup analysis (type of odors) for familiar odor vs artificial odor without habituation on crying duration after the procedure

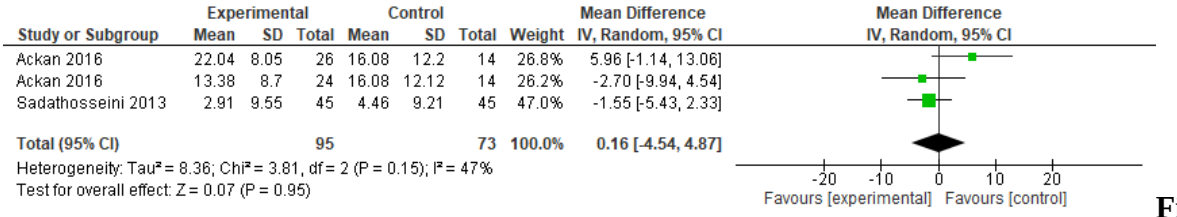


**Figure S352**. Familiar odor vs artificial odor without habituation on heart rate variations during the procedure

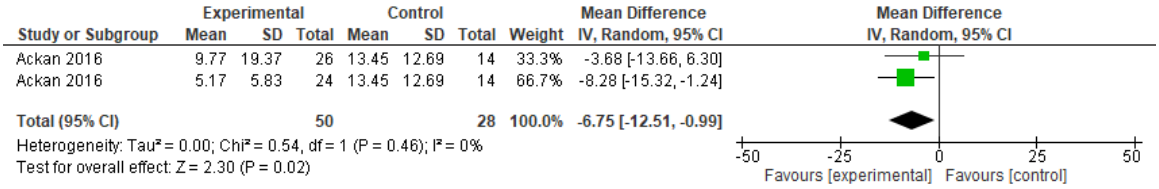


**Figure S36**. Familiar odor vs artificial odor without habituation on heart rate variations after the procedure

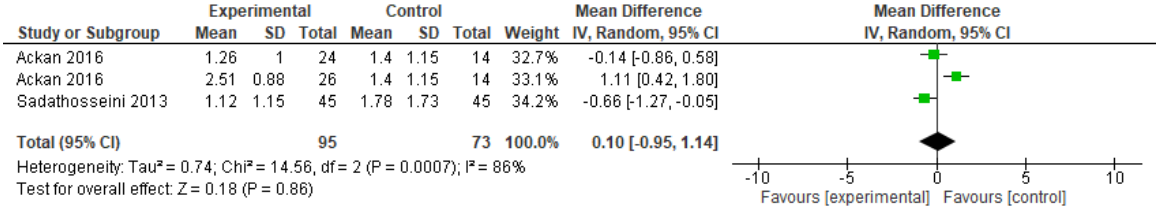


**Figure S37**. Familiar odor vs artificial odor without habituation on oxygen saturation variations during the procedure

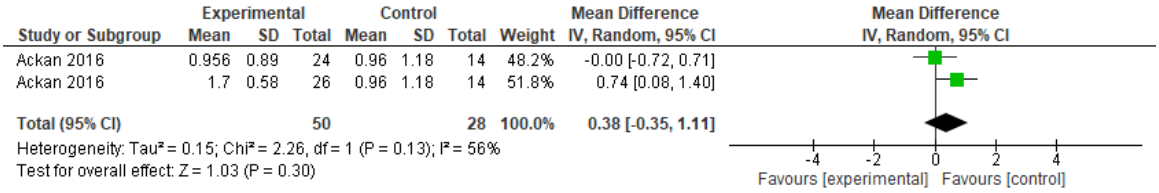


**Figure S38**. Familiar odor vs artificial odor without habituation on oxygen saturation variations after the procedure

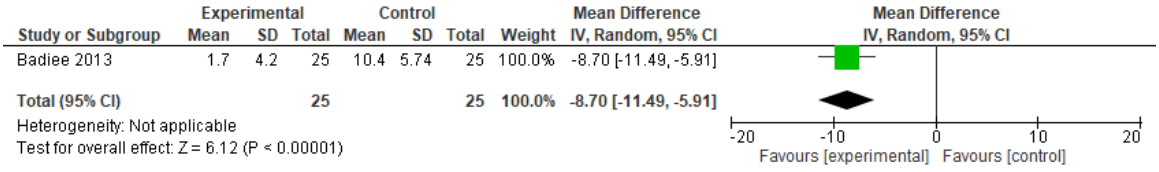


**Figure S39**. Familiar odor vs artificial odor without habituation on cortisol level