Cochrane Australia

# Systematic review of evidence on the clinical effectiveness of aromatherapy

Report prepared by

Cochrane Australia

04 March 2024

Trusted evidence.

Informed decisions.

Better health.

# Contents

In November 2020 Cochrane Australia was contracted by the National Health and Medical Research Council (NHMRC) to design and undertake the systematic review described in this report. This systematic review is one of several independent contracted evidence evaluations being undertaken to update the evidence underpinning the 2015 *Review of the Australian Government Rebate on Natural Therapies for Private Health Insurance* (2015 Review) by the Department of Health (Department). The design and conduct of the review were done in collaboration with the Office of NHMRC (ONHMRC), NHMRC's Natural Therapies Working Committee (NTWC) and the Department of Health and Aged Care's Natural Therapies Review Expert Advisory Panel (NTREAP).

## Copyright



## Attribution



## Use of images



## Contact

To obtain information regarding NHMRC publications or submit a copyright request, contact: communications@nhmrc.gov.au.

## Acknowledgments

Thanks to the members of the Department's NTREAP and NHMRC's NTWC for their advice and comments throughout the creation of this document. PRACI data was provided by Dr Amie Steel at UTS.

Membership and other details of the Panel and Committee can be found at:

https://www.health.gov.au/committees-and-groups/natural-therapies-review-expert-advisory-panel

https://www.nhmrc.gov.au/about-us/leadership-and-governance/committees/natural-therapies-working-committee

# Plain language summary

## What was the aim of the review?

The aim of this review was to examine the effects of aromatherapy in preventing and/or treating injury, disease, medical conditions or preclinical conditions.
Aromatherapy is the therapeutic use of essential oils from plants (flowers, herbs, or trees), via inhalation, massage, or topical use, to treat ill health and promote physical, emotional and spiritual well-being. It is one of the most widely used natural therapies reported by consumers in Western countries.
This review was targeted for the Australian Government Department of Health and Aged Care (formally Department of Health) to assist in their Natural Therapies Review, which was designed to determine whether certain natural therapies, including aromatherapy, have enough evidence of effectiveness to be considered re-eligible for private health insurance rebates. This review was not designed to be a complete review of all studies published for aromatherapy, nor is it intended to inform decisions about whether an individual or practitioner should use aromatherapy.

## Key messages

- There is a large and growing body of evidence examining the effects of aromatherapy on health. Despite this, it is not possible to draw conclusions about the effects of aromatherapy with confidence for any condition or outcome.
- Although we interpret many of the results in the review, the evidence is of low or very low certainty, meaning that the true effect of aromatherapy may be substantially different.
- We are uncertain about the effects of aromatherapy because of serious concerns about the methods used in all of the studies in the review. Another concern is that trialists may have reported findings of large beneficial effects from aromatherapy selectively, and not published findings that showed little or no effect.
- These preventable flaws in how the studies were designed, conducted and reported mean that we cannot tell whether aromatherapy has beneficial effects or little or no effect on health outcomes.

## What was studied in the review?

We examined evidence from randomised trials to study the effect of aromatherapy on

- pain,
- nausea and vomiting,
- sleep quality,
- fatigue,
- emotional functioning and mental health,
- health-related quality of life, and
- physical function.

We examined effects on these outcomes for a wide range of conditions and populations that were agreed through a prioritisation process. For each outcome, we examined the effects of aromatherapy overall (across multiple conditions) and for specific population groups. This approach makes best use of all available evidence to help us decide if there is evidence that aromatherapy works 'in general' or whether any effects might be limited to specific population groups. Assessments of cost-effectiveness, safety and studies of healthy populations were not included in this review.

We applied methods in the Cochrane Handbook for Systematic Reviews of Interventions [2] to search for, collate, appraise, and synthesise evidence from trials. We then applied methods from Grading of Recommendations Assessment, Development and Evaluation (GRADE) working group to interpret the synthesis results in a systematic and transparent way. GRADE is a method used to assess and describe how confident (or certain) we can be that the estimates of the effect (calculated by combining results from multiple studies) reflect the true effects of the intervention. In deciding on our certainty (or confidence) in each result, we consider all relevant information collected in the review. We used four levels to describe our certainty as either:

- High certainty: we are very confident that the true effect lies close to that of the estimate of the effect.

- Moderate certainty: we are moderately confident that the true effect is probably close to the estimate of the effect, but there is a possibility that it is substantially different.
- Low certainty: our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect.
- Very low certainty: we have very little confidence in the estimate and the true effect is likely to be markedly different from the estimated effect. Our confidence in the result is too limited to provide a meaningful interpretation.

Our methods were pre-specified in a publicly available protocol (PROSPERO ID CRD42021268244) that underwent independent review by methods specialists, the Department's panel (NTREAP) and was endorsed by the NHMRC Natural Therapies Working Committee (NTWC) [1]. The review is reported in accordance with the PRISMA 2020 statement [6, 7].

## What were the main results of the review?

We included 323 studies in the review, of which 201 studies contributed results to at least one synthesis of evidence. The largest syntheses included results from over 7000 participants.

The evidence provides low certainty that across multiple conditions and compared to an inactive control (placebo, no intervention, usual care), aromatherapy (delivered by inhalation, massage, or topically) may improve:

- sleep quality (no trials among people living with dementia and behaviour change),
- health-related quality of life, and
- physical function.

For pain, nausea and vomiting, fatigue, emotional functioning and mental health the evidence was very uncertain overall. For these outcomes, the effects varied importantly across studies; some studies showed benefit, others showed little or no effect on the outcome. These inconsistent effects were not explained by differences in the population receiving aromatherapy nor by the way in which aromatherapy was delivered (mode of delivery).

For some population groups the results were somewhat more certain, as follows.

There was low certainty that aromatherapy may improve:

- pain among people with chronic musculoskeletal conditions,
- acute or episodic pain conditions (mainly dysmenorrhea),
- nausea and vomiting during pregnancy,
- mental health among people with symptoms of mental distress,
- physical function among people with chronic musculoskeletal conditions.

There was also low certainty that aromatherapy may have little or no effect on:

- mental health among people living with cancer (no trials among people with non-cancer advanced disease that was not amenable to cure),
- mental health among people living with dementia (mainly agitation),

Fewer studies compared aromatherapy massage to an inactive massage control (comparable to that used to deliver aromatherapy). There was low certainty evidence that health-related quality of life improved with aromatherapy massage, but it was uncertain whether there was benefit or little or no effect on other outcomes. There were no studies that compared aromatherapy massage to an inactive massage control for nausea and vomiting or sleep quality.

## Implications for health policy and research

This review assessed the evidence for certain conditions and groups of people to inform the Australian Government about health policy decisions for private health insurance rebates. The review did not cover all the reasons that people use aromatherapy, or the reasons practitioners prescribe aromatherapy and was not intended to inform individual choices about using aromatherapy.

### Implications for health policy

The evidence is of low or very low certainty for all outcomes and populations considered in this review. This means that our confidence in the effect of aromatherapy on each outcome is limited, and the true effect may be substantially different. Major concerns about inconsistent results (some studies showing benefit, and others little or no effect), and the likelihood that results that show large beneficial effects from aromatherapy may have been selectively published by trialists, should be considered when deciding whether there is any credible evidence to support the use of aromatherapy.

### Implications for future research

Given the extent of concerns about bias in included studies and reporting bias, it is unlikely that systematic reviews will be able to build on the existing evidence base to answer questions about the effects of aromatherapy with any certainty. While further investigation of published and unpublished trials of aromatherapy may help us understand the full extent of flaws in the evidence, it is unlikely to be feasible or possible to conduct these studies. Improving the conduct and, at a minimum, the reporting of trials in this field is essential. The value of conducting more trials on aromatherapy needs to be carefully assessed to avoid research waste.

## How up-to-date is the review?

Searches were conducted from the earliest date included in the databases until 20 August 2021. Studies published after this date are not included in this review.

# Executive summary

## Background

Aromatherapy - the therapeutic use of essential oils from plants (flowers, herbs, or trees) to treat ill health and promote physical, emotional and spiritual well-being - is one of the most widely used natural therapies reported by consumers in Western countries. The Australian Government Department of Health and Aged Care (via the National Health and Medical Research Council) commissioned a suite of independent evidence evaluations to inform the 2019-20 Review of the Australian Government Rebate on Private Health Insurance for Natural Therapies. This report is for one of the evaluations; a systematic review of randomised trials examining the effectiveness of aromatherapy in preventing and/or treating injury, disease, medical conditions or preclinical conditions. In 2015, an overview of systematic reviews conducted for the Australian Government found there was no clear scientific evidence that aromatherapy was effective. The current systematic review considered primary evidence and a wider range of publication dates.

## Objectives

Primary objectives were to answer the following questions:

1. What is the effect of *aromatherapy (delivered by any mode)* compared to an inactive control (placebo, no intervention or usual care) among people with any condition, pre-condition, injury or risk factor on outcomes for which aromatherapy is used (pain, nausea and vomiting, sleep quality, fatigue, emotional functioning and mental health, health-related quality of life, and physical function)?
2. What is the effect of *aromatherapy (delivered by massage)* compared to *massage alone* among people with any condition, pre-condition, injury or risk factor on outcomes for which aromatherapy is used (outcomes as per question 1)?

Secondary objectives related to the following questions:

3. What is the effect of *aromatherapy (delivered by any mode)* compared to an inactive control (placebo, no intervention or usual care) on outcomes for each underlying condition, pre-condition, injury or risk factor (for example, effects on sleep disruption among people living with cancer, people with chronic insomnia, people with chronic pain or people with dementia)?
4. What is the effect of *aromatherapy (delivered by massage)* compared to *massage alone* on outcomes for each underlying condition, pre-condition, injury or risk factor (for example, effects on sleep disruption among people living with cancer, people with chronic insomnia, people with chronic pain or people with dementia)?
5. What evidence exists examining the effects of aromatherapy compared to active comparators?
6. What evidence exists on the effects of aromatherapy compared to inactive controls or other treatments, for conditions that were not prioritised for the review?

We planned to examine the effects of aromatherapy compared to “gold standard” (first line) treatments, however this was not feasible because of the volume of evidence. These studies are listed in an evidence inventory. Other objectives are as stated in the protocol, with editing to include outcomes and conditions in the final framework.

This information will be used by the Australian Government in deciding whether to reinclude aromatherapy as eligible for private health insurance rebates, after aromatherapy was excluded in 2019. This review was not designed to assess all the reasons that people use aromatherapy, or the reasons practitioners prescribe aromatherapy and was not intended to inform individual choices about using aromatherapy.

## Methods

This review was prospectively registered on the international prospective register of systematic reviews (PROSPERO ID CRD42021268244) and the methods pre-specified in a protocol published on the register and in a peer reviewed journal article [1]. The methods were based on the Cochrane Handbook for Systematic Reviews of Interventions [2]. The Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach was used to summarise and assess the certainty of evidence arising from this review [3-5]. The review is reported in accordance with the PRISMA 2020 statement [6, 7] which has been adopted by Cochrane.

## Criteria for including studies in the review

Broad eligibility criteria were defined for including studies in the review, as summarised below.

- ***Types of study designs and comparisons***. We included randomised trials comparing (1) aromatherapy delivered by any mode (inhalation, massage or topically) to inactive controls (placebo, no intervention, usual care) or (2) aromatherapy delivered by massage to an inactive massage control (comparable to that used in the aromatherapy arm). We also included studies comparing aromatherapy delivered by any mode (inhalation, massage or topically) with a co-intervention to the same cointervention.
- ***Types of populations***. Any condition, pre-condition, injury or risk factor (excluding healthy participants without clearly identified risk factors for the condition aromatherapy was used to prevent). Through the prioritisation process, it was agreed to exclude skin conditions, infections, infestations and wounds, and substance withdrawal from the synthesis (studies are included in the evidence inventory for the review).
- ***Types of outcomes***. Any patient-important outcome for which aromatherapy is indicated was eligible for the review. Through the prioritisation process, outcomes determined to be critical or important for the synthesis were pain, nausea and vomiting, sleep, fatigue, emotional functioning and mental health, health-related quality of life and physical function.
- ***Other criteria***. Studies in languages other than English were not eligible for synthesis but were listed in an appendix.

## Search methods

We searched the Cochrane Central Register of Controlled Trials (CENTRAL via the Cochrane Library, Issue 8, 2021), PubMed, AMED (Allied and Complementary Medicine) via Ovid, and Ovid Emcare on 20 August 2021. Searches were not restricted by date, language or format of publication. The public was also invited by the Department to submit references for published research evidence.

## Analytic framework for synthesis and prioritisation process

A staged process, designed to minimise bias in the review, was agreed *a priori* for determining which of the studies eligible for the review would be included in the synthesis (see Summary of methods, Figure 3.1). Through this process, NTWC and NTREAP prioritised outcomes and populations for the synthesis. A framework for the synthesis was finalised prior to commencing data extraction. This framework defined the scope of the evidence synthesis and specified the synthesis questions and associated PICO (populations, interventions, comparators, outcomes) criteria for including studies in each synthesis (see Summary of methods, Figure 3.5.1).

## Data collection and analysis

Screening of citations and full text reports was completed by two authors, independently. Data extraction and risk of bias assessment (ROB 2.0) was piloted by three authors, then completed by a single author and checked by a second.

Comparisons were based on outcome domains (pain, nausea and vomiting, sleep, fatigue, emotional functioning and mental health, health-related quality of life and physical function), both overall and stratified by population groups (e.g. cancer and advanced disease, chronic musculoskeletal pain, dementia). The outcome domains and population groups were defined in the analytic framework for the synthesis. Meta-analysis methods were used to combine results across studies where appropriate. Characteristics of studies eligible for the review but ineligible for the synthesis were tabulated.

GRADE methods were used to assess certainty of evidence and summarise findings. For all results an interpretation was made about whether the observed effect was important (or not) and how certain we were in the finding (high, moderate, low or very low). Certainty accounted for concerns about bias (arising from studies included in and missing from the synthesis), how precisely the effect was estimated, important inconsistency in the results across studies, and how directly the studies in each synthesis addressed the synthesis question defined in the analytic framework.

## Main results

A total of 323 studies were included in the review. Two hundred and thirty-four studies were eligible for the evidence synthesis (following screening of 3547 citations and 664 reports), of which 201 contributed to at least one meta-analysis. Thirty-three of the 234 studies did not contribute to any of the analyses for which they were eligible

because the required data were not available (could not be calculated, imputed, were not reported, or were uninterpretable). Eighty-nine studies were excluded from the synthesis following the prioritisation process. These were primarily studies of skin conditions and skin infections for which essential oils were used topically for their antimicrobial or inflammatory properties, rather than aromatherapy per se. Characteristics of these studies are reported on the evidence inventory. A further 154 studies were listed as awaiting classification (including 81 likely eligible studies in languages other than English, and 33 studies published as abstracts only).

### Effects of aromatherapy

The evidence about the effects of aromatherapy is of **low** or **very low certainty** for all outcomes. This means that our confidence in the estimate of effect for each outcome is limited and the true effect may be substantially different from the estimated effect of the intervention.

Overall, there is uncertainty because the results are inconsistent across studies (some finding benefit, others little or no effect), and because of concerns that beneficial effects may be exaggerated because of methodological limitations of included studies (risk of bias) and selective non-reporting of results that show unfavourable effects (e.g. missing results that show little or no effect).

### Effects of aromatherapy on pain

Overall, the evidence is very uncertain about the effect of aromatherapy (delivered by inhalation, massage, or topically) on pain compared to an inactive control (placebo, no intervention, usual care) (overall analysis, all population groups; 82 studies, 7193 participants; very low certainty evidence).

For the population groups examined in this analysis, the effects are as follows.

| | | |
|---|---|---|
| Aromatherapy may reduce pain | | |
| • from chronic musculoskeletal conditions | low certainty | 7 studies, 347 participants |
| • from acute or episodic pain conditions (mainly dysmenorrhea) | low certainty | 9 studies, 855 participants |

| | | |
|---|---|---|
| The evidence is very uncertain about the effect on pain | | |
| • after surgery (in the acute post-operative period) | very low certainty | 20 studies, 1597 participants |
| • during or after a procedure (peri-procedural period) | very low certainty | 29 studies, 2322 participants |
| • from acute musculoskeletal conditions | very low certainty | 1 study, 60 participants |
| • from headache or migraine | very low certainty | 1 study, 141 participants |
| • from cancer or advanced disease (not amenable to cure) | very low certainty | 2 studies, 338 participants |
| • during labour and birth | very low certainty | 9 studies, 1239 participants |
| • from other chronic conditions | very low certainty | 4 studies, 294 participants |

Overall, the evidence is very uncertain about the effect of aromatherapy massage on pain compared to massage alone (comparable to that used to deliver aromatherapy) (overall analysis, all population groups; 19 studies, 1058 participants; very low certainty evidence).

For the population groups examined in this analysis, the effects were as follows.

| | | |
|---|---|---|
| The evidence is very uncertain about the effect on pain | | |
| • on pain from chronic musculoskeletal conditions | very low certainty | 5 studies, 278 participants |
| • after surgery (in the acute post-operative period) | very low certainty | 3 studies, 110 participants |
| • during or after a procedure (peri-procedural period) | very low certainty | 2 studies, 101 participants |
| • during labour and birth | very low certainty | 1 study, 60 participants |
| • from other chronic conditions | very low certainty | 3 studies, 195 participants |
| • from other acute pain | very low certainty | 5 studies, 314 participants (all in dysmenorrhea) |

| |
|---|
| No studies were included in the analysis of the effects on pain |
| • from headache or migraine (episodic or acute) |
| • from cancer or advanced disease |

### Effects of aromatherapy on nausea and vomiting

Overall, the evidence is very uncertain about the effect of aromatherapy (delivered by inhalation, massage, or topically) on nausea and vomiting compared to an inactive control (placebo, no intervention, usual care) (overall analysis, all population groups; 23 studies, 2032 participants; very low certainty evidence).

For the population groups examined in this analysis, the effects were as follows.

| | | |
|---|---|---|
| Aromatherapy may reduce nausea and vomiting | | |
| • during pregnancy | low certainty | 4 studies, 271 participants |

| | | |
|---|---|---|
| The evidence is very uncertain about the effect on nausea and vomiting among people | | |
| • living with cancer and advanced disease (including those undergoing chemotherapy) | very low certainty | 8 trials, 738 participants |
| • after surgery (in the acute post-operative period) | very low certainty | 10 studies, 982 participants |
| • undergoing procedures | very low certainty | 1 trial, 41 participants |

No studies were included in the synthesis that compared aromatherapy massage to massage alone (comparable to that used to deliver aromatherapy).

### Effects of aromatherapy on sleep quality

Compared to an inactive control (placebo, no intervention, usual care), aromatherapy (delivered by inhalation, massage, or topically) may improve sleep quality (overall analysis, all population groups; 22 studies, 1397 participants; low certainty evidence).

For the population groups examined in this analysis, the effects were as follows.

| | | |
|---|---|---|
| Aromatherapy may improve sleep | | |
| • during hospitalisation (excluding surgery; mainly cardiovascular inpatients) | low certainty | 8 studies, 498 participants |

| | | |
|---|---|---|
| The evidence is very uncertain about the effect on sleep quality | | |
| • among people living with cancer or advanced disease | very low certainty | 3 studies, 163 participants |
| • after surgery (in the acute post-operative period) | very low certainty | 3 studies, 227 participants |
| • among people with chronic insomnia | very low certainty | 3 studies, 131 participants |
| • for people with signs or symptoms of sleep disruption | very low certainty | 5 studies, 378 participants |

| | | |
|---|---|---|
| No studies were included in the analysis | | |
| • among people living with dementia | | |

No studies were included in the synthesis that compared aromatherapy massage to massage alone (comparable to that used to deliver aromatherapy).

### Effects of aromatherapy on fatigue

Overall, the evidence is very uncertain about the effect of aromatherapy (delivered by inhalation, massage, or topically) on fatigue compared to an inactive control (placebo, no intervention, usual care) (overall analysis, all population groups; 18 studies, 1316 participants; very low certainty evidence).

For the population groups examined in this analysis, the effects are as follows.

| | | |
|---|---|---|
| The evidence is very uncertain about the effect on fatigue | | |
| • among people with chronic musculoskeletal conditions | very low certainty | 1 study, 34 participants |
| • among people with cancer and advanced disease | very low certainty | 3 studies, 398 participants |
| • during pregnancy | very low certainty | 1 study, 89 participants |

| | | |
|---|---|---|
| • among people with chronic conditions (mainly those undergoing haemodialysis for kidney disease) | very low certainty | 13 studies, 795 participants |

Overall, the evidence is very uncertain about the effect of aromatherapy massage compared to massage alone on fatigue (overall analysis, all population groups; 4 trials, 252 participants with chronic conditions; very low certainty evidence).

No studies were included among people with cancer and advanced disease or during pregnancy.

### Effects of aromatherapy on emotional functioning and mental health

Overall, the evidence is very uncertain about the effect of aromatherapy (delivered by inhalation, massage, or topically) on emotional functioning and mental health compared to an inactive control (placebo, no intervention, usual care) (overall analysis, all population groups; 86 studies, 7032 participants; very low certainty evidence).

For the population groups examined in this analysis, the effects are as follows.

| | | |
|---|---|---|
| Aromatherapy may improve | | |
| • emotional functioning and mental health among people with symptoms of mental distress (mainly depression symptoms) | low certainty | 5 studies, 440 participants |

| | | |
|---|---|---|
| Aromatherapy may have little or no effect on | | |
| • emotional functioning and mental health among people living with cancer or advanced disease | low certainty | 7 studies, 275 participants |
| • agitation among people living with dementia and behavioural change | low certainty | 7 studies, 521 participants |

| | | |
|---|---|---|
| The evidence is very uncertain about the effect on | | |
| • perioperative anxiety (i.e. in the period immediately before surgery) | very low certainty | 17 studies, 1428 participants |
| • periprocedural anxiety (i.e. before or during a procedure) | very low certainty | 33 studies, 2854 participants |
| • anxiety during hospitalisation for people admitted for cardiovascular conditions | very low certainty | 12 studies, 1030 participants |
| • anxiety during labour and childbirth | very low certainty | 5 studies, 484 participants |

| | | |
|---|---|---|
| No studies were included in the analysis of the effects on emotional functioning and mental health | | |
| • among people living with a diagnosed mental disorder | | |

Overall, the evidence is very uncertain about the effect of aromatherapy massage on emotional functioning and mental health compared to massage alone (comparable to that used to deliver aromatherapy) (overall analysis, all population groups; 11 studies, 664 participants; very low certainty evidence).

For the population groups examined in this analysis, the effects are as follows.

| | | |
|---|---|---|
| Aromatherapy may improve | | |
| • anxiety during hospitalisation | low certainty | 3 studies, 232 participants |

| | | |
|---|---|---|
| The evidence is very uncertain about the effect on | | |
| • perioperative anxiety (i.e. before a surgery) | very low certainty | 2 studies, 130 participants |
| • emotional functioning and mental health among people living with cancer or advanced disease | very low certainty | 2 studies, 134 participants |
| • agitation or other behaviour changes among people living with dementia and behavioural change | very low certainty | 2 studies, 85 participants |
| • emotional functioning and mental health among people with symptoms of mental distress (mainly depression symptoms) | very low certainty | 1 study, 57 participants |

| | | |
|---|---|---|
| No studies were included in the analysis on<br>• periprocedural anxiety (i.e. before or during a procedure)<br>• anxiety during labour and childbirth<br>• emotional functioning and mental health among people living with a diagnosed mental disorder | | |

### Effects of aromatherapy on health-related quality of life (HR-QoL)

Compared to an inactive control (placebo, no intervention, usual care), aromatherapy (delivered by inhalation, massage, or topically) may improve health related quality of life (overall analysis, all population groups; 14 studies, 1048 participants; low certainty evidence).

For the population groups examined in this analysis, the effects are as follows.

| | | |
|---|---|---|
| The evidence is very uncertain about the effect on HR-QoL | | |
| • among people living with cancer or advanced disease | very low certainty | 3 studies, 527 participants |
| • for people living with chronic conditions | very low certainty | 11 studies, 521 participants |

Overall, the evidence is very uncertain about the effect of aromatherapy massage on health-related quality of life compared to massage alone (comparable to that used to deliver aromatherapy) (overall analysis, all population groups; 12 studies, 851 participants; very low certainty evidence).

For the population groups examined in this analysis, the effects are as follows.

| | | |
|---|---|---|
| Aromatherapy may improve HR-QoL | | |
| • for people living with chronic conditions | low certainty | 9 studies, 581 participants |

| | | |
|---|---|---|
| The evidence is very uncertain about the effect on HR-QoL | | |
| • among people living with cancer or advanced disease | very low certainty | 3 studies, 270 participants |

### Effects of aromatherapy on physical function

Compared to an inactive control (placebo, no intervention, usual care), aromatherapy (delivered by inhalation, massage, or topically) may improve physical function (overall analysis, all population groups; 10 studies, 527 participants; low certainty evidence).

For the population groups examined in this analysis, the effects are as follows.

| | | |
|---|---|---|
| Aromatherapy may improve physical function | | |
| • for people with knee osteoarthritis, but the effects are very uncertain for other chronic musculoskeletal conditions | low certainty | 6 studies, 313 participants (265 with knee osteoarthritis) |

| | | |
|---|---|---|
| The evidence is very uncertain about the effect on physical function | | |
| • among people living with cancer and advanced disease | very low certainty | 1 trial, 60 participants |
| • for people with other chronic conditions | very low certainty | 3 studies, 154 participants |

| | | |
|---|---|---|
| No studies were included in the analysis<br>• among people with headache or migraine (chronic or episodic) | | |

Overall, the evidence is very uncertain about the effect of aromatherapy massage on physical function compared to massage alone (comparable to that used to deliver aromatherapy) (overall analysis, all population groups; 7 studies, 434 participants; very low certainty evidence).

For the population groups examined in this analysis, the effects are as follows.

| | | |
|---|---|---|
| The evidence is very uncertain about the effect on function | | |
| • for people with chronic musculoskeletal conditions | very low certainty | 5 studies, 278 participants |
| • for people with other chronic conditions | very low certainty | 2 studies, 156 participants |

No studies were included in the analysis

- people with headache or migraine (chronic or episodic)
- among people living with cancer and advanced disease

## Limitations

### Of the evidence contributing to the review

Limitations of the evidence were considered when interpreting each result by applying the GRADE approach. Overarching concerns that reduce confidence in all findings arise from

- methodological limitations of included trials (for all studies there was either a high risk of bias or some concerns),
- missing results (there was evidence that results may be missing for studies for which results favoured the control), and
- inconsistent results across studies (some showing benefit, others showing little or no effect).

Additional concerns applied to many findings.

In addition to factors addressed in the GRADE assessment, there were problems with the quality of reporting in the included studies. Incomplete and ambiguous reporting affected our ability to understand the study design and confirm design features that could introduce bias. This also precluded inclusion of a large amount of data from the analyses: 41 trials (4415 participants) of which 33 did not contribute to any of the meta-analyses for which the study was eligible. The reasons why data could not be included varied (details reported in section 4.1), but for the majority of studies the problems were such that a summary or other synthesis of the results could be misleading.

### Of the review process

In this review we applied methods recommended in the Cochrane handbook for systematic reviews of interventions and the GRADE approach, as per the detailed protocol that was prospectively registered on PROSPERO after undergoing independent methodological review. The final list of populations and outcomes eligible for the review were determined through a pre-specified prioritisation process, performed by NTWC and NTREAP without knowledge of the included studies or results of those studies. An initial analytic framework for the review was included in the protocol to inform these decisions and propose a structure for the synthesis.

While data extraction for each study was performed by a single reviewer, the selection of outcomes and coding of studies for inclusion in meta-analyses was performed independently by a second experienced review author. All data were checked by a second experienced author, with input from a biostatistician, and all data manipulation and analyses were performed by a biostatistician. These steps minimised the risk of errors or misinterpretation. Risk of bias assessments were performed for each study by a single reviewer following detailed guidance developed for the review and training in the assessment of design features relevant to this review.

While we endeavoured to include all available studies in the analyses (applying all suggested methods from the Cochrane Handbook), many studies reported data that could not be interpreted or from which the required statistics could not be calculated or imputed. The large number of studies in the review meant it was not feasible to contact trialists for additional information, nor was it possible to review the large number of trial registry entries to conduct a thorough assessment of missing results from the synthesis. However, we were able to use graphical methods (funnel plots) to examine whether results may have selectively reported (publication bias).

Finally, we screened and reported citations for studies in languages other than English, however these studies were not included in the synthesis (as per protocol). There is no reason to expect that the results of these studies would differ systematically from those reported in English and, in turn, that exclusion of these studies would bias the

results of the review. Given the amount of data contributing to most analyses, addition of these studies is unlikely to change the review conclusions.

## Conclusions

There is a large and growing body of evidence about the effects of aromatherapy on health. Despite this, it is not possible to draw conclusions about the effects of aromatherapy with confidence for any condition or outcome. Unlike many reviews, this is not due to a lack of evidence from randomised trials, with all but one of the syntheses for the first comparison containing data from over 1000 participants (i.e. the precision of effects estimates is not a serious concern). Instead, the uncertainty reflects significant methodological problems with the evidence base. Although an interpretation is made for many of the results from meta-analyses, the evidence is of low or very low certainty, meaning that the true effect of aromatherapy may be substantially different. Many factors contribute to this uncertainty. Of greatest concern is that results that show large beneficial effects from aromatherapy (beyond what would be seen for many first line therapies) may have been published selectively, while results that show little or no effect are not reported. Together with biases in the conduct of studies (e.g. bias arising from the randomisation process, unblinded outcome assessment; and selection of the reported results), this may be one of the underlying reasons for the inconsistent results observed across studies. In addition, the absence of any studies at low risk of bias meant it was not possible to examine the impact bias in the included studies has on the results using our planned approach (i.e. limiting analyses to studies at low risk of bias to check if the results are robust).

### Implications for health policy

The evidence is of low or very low certainty for all outcomes and populations considered in this review. This means that our confidence in the estimate of effect for each outcome is limited, and the true effect may be substantially different from the estimated effect of the intervention. Major concerns about inconsistent results (some studies showing benefit, and others little or no effect) without a credible explanation, and the likelihood that results that show large beneficial effects from aromatherapy may have been selectively published by trialists, should be considered when deciding whether there is any credible evidence to support the use of aromatherapy.

### Implications for future research

Given the extent of concerns about bias in included studies and bias due to missing results (reporting bias), it is unlikely that systematic reviews will be able to answer questions about the effects of aromatherapy with certainty by building on the very large body of existing evidence. Although a thorough investigation of the integrity of existing research in this field may provide evidence about the extent of reporting bias, our examination of trial registry entries suggests that there may not be sufficient information to conduct these studies using methods proposed for research on research integrity.  Improving the conduct and, at a minimum the reporting, of trials in this field is an imperative. Any future trials must address preventable limitations in the conduct and reporting of trials of aromatherapy (including, but not limited to, bias arising from the randomisation process, the method of outcome assessment; and the reporting of results). Adhering to relevant reporting guidelines such as CONSORT (https://www.equator-network.org/reporting-guidelines/consort/), and addressing errors that rendered results unusable is essential. The value of conducting more trials on aromatherapy would need to be carefully assessed to avoid further research waste.

# 1. Background

In 2015, the Australian Government conducted a *Review of the Australian Government Rebate on Natural Therapies for Private Health Insurance (2015 Review).* Underpinned by systematic reviews of evidence for each natural therapy, one of the findings from the 2015 Review was that there was no clear scientific evidence that aromatherapy was effective. The National Health and Medical Research Council (NHMRC) has been engaged by the Department of Health and Aged Care (Department) to update the evidence underpinning the 2015 Review. This evidence evaluation of aromatherapy is one of a suite of independent contracted systematic reviews that will inform the *Review of the Australian Government Rebate on Private Health Insurance for Natural Therapies 2019-20* (2019-20 Review) [8].

Aromatherapy is one of the most widely used natural therapies reported by consumers in Western countries. A systematic review of 89 surveys (97,222 participants) estimating the prevalence of Complementary Medicine (CM) use by consumers in the United Kingdom (UK) found that aromatherapy was the third most popular CM from among 28 different therapies [9]. In Australia, a cross-sectional survey examining consultation with complementary therapists and use of complementary medicine products found that about half of all respondents (1016/2025 adults) used complementary medicines [10, 11]. Aromatherapy oils were used by 11% of respondents (N=224/2019), and 3.9% of respondents had visited an aromatherapist (N=79/2019) [11]. Based on the average spending on complementary medicines reported in this survey, the study authors estimated the total expenditure on aromatherapy oils in Australia to be $250 million in the previous 12 months (2016-2017) [10].

## 1.1 Description of the intervention

Aromatherapy is the therapeutic use of essential oils from plants (flowers, herbs, or trees) to treat ill health and promote physical, emotional and spiritual well-being [8, 12, 13]. The name 'aromatherapy' suggests that treatments are delivered directly or indirectly through the olfactory system and that 'aroma' is central to therapeutic action. However, there are multiple modes of administration, and these include treatments intended to act through direct contact with the skin and inhalation into the lungs (rather than through an 'aroma' inhaled through the olfactory system). The inclusion of such therapies within the scope of aromatherapy practice has led some professional groups to suggest that a more apt description is "essential oil therapy" [14].

***Active ingredients and choice of essential oils***

Although the scope of aromatherapy practice varies, the use of essential oils is central to all definitions [13-17]. Essential oils are volatile oils extracted using steam distillation or mechanical expression from aromatic plants [13, 18]. While it is possible to extract essential oils using solvents ('absolutes') and to produce synthetic versions of some oils, aromatherapists generally consider that these are not true essential oils and are therefore unsuitable for therapeutic use [13, 18].

Essential oils vary greatly in their molecular composition. This composition determines the aroma of each oil and the pathways by which it is absorbed, distributed and metabolised to produce effects [13, 18]. Aromatherapists tailor treatments to individual needs, selecting essential oils, and their mode of application, based on anticipated therapeutic properties for the targeted condition [8, 13].

***Mode of administration and dose***

Inhalation through passive diffusion in the air (e.g. through mist or heat diffusers, steam vaporisation) and direct inhalation (e.g. individual inhalers, steam inhalation) can be used as the primary mode of administering essential oils. Topical application of diluted essential oils to the skin is also common [13]. The intention of topical application may be to produce local effects at the point of administration (e.g. to alleviate pain in a joint), to mediate effects through inhalation (whether through the lungs or olfactory system), or through skin absorption. Massage is a common co-intervention used with topical application of essential oils. While massage may have a therapeutic effect when used independently of essential oils, it is generally described as an "integral" part of aromatherapy treatment [14]. For topical application, essential oils are diluted in a carrier oil, usually vegetable or nut oil (e.g. sweet almond oil, grapeseed, jojoba oil) [19]. These carrier oils differ from essential oils in that they contain fatty acids, vitamins and minerals, and are believed to aid absorption of the essential oil through the skin [19].

Limiting the dose or concentration of essential oils is considered an important means of avoiding systemic toxicity or adverse effects, such as skin irritation or sensitivity [18, 19]. The typical dose of essential oil used for therapeutic purposes varies depending on indication, and the oil and route of administration, but is generally in the range of a 2.5-5% dilution of essential oils for topical use [18]. Lower concentrations (i.e. higher dilutions) are recommended for some population groups, including people who are pregnant, children, and people with conditions or receiving treatments/medications that may put them at greater risk of adverse effects (e.g. people with skin conditions or damage; people undergoing radiotherapy; people with asthma) [14, 18].

Although other routes of administration are sometimes used, professional associations for aromatherapists in Australia, the United Kingdom, Canada and the United States have position statements recommending against ingestion of essential oils, internal use (on or near mucous membranes), and the use of undiluted essentials oils on the skin [14-16].

***Practitioners of aromatherapy and regulation***

Aromatherapy is practised by natural therapists, including aromatherapists, naturopaths and massage therapists. It is also an increasingly common professional education option for nurses, allied health professionals, and those working in sectors such as palliative care.

Aromatherapy practice is not regulated by the Australian Health Practitioner Regulation National Law, which means there is no requirement for professional registration of practitioners of aromatherapy [20, 21]. The International Aromatherapy and Aromatic Medicine Association (IAAMA) offers membership to aromatherapy practitioners in Australia who have completed accredited training through the National Quality Training Framework [22]. The IAAMA, and other associations for natural therapists in Australia, also set standards for practice and ethical conduct, and have requirements for continuing professional education [22, 23]. Some professional associations have safety guidelines and position statements aimed at preventing the use of contraindicated oils, unsafe therapies and treatments that are not widely accepted by the profession (for examples, see [14-17]).

In the 2016-17 cross-sectional survey examining use of complementary medicine products, only a minority of those who reported therapeutic use of aromatherapy oil consulted a complementary medicine practitioner (12.5%) for a prescription, whereas self-prescription was common (43%) [10]. Indeed, part of the appeal of aromatherapy may be the accessibility of essential oils, which do not require a prescription. The Australian Government provides a safeguard for consumers by regulating essential oils intended for therapeutic use through the Therapeutic Goods Administration (TGA). However, most essential oils are designated as lower risk medicines, which means they are assessed by the TGA for quality and safety, but not effectiveness [24].

## 1.2 How aromatherapy might work

The research literature and guidance on aromatherapy describes multiple theories of how aromatherapy works (for examples, see [13, 14]). This is perhaps unsurprising given that the exact mechanism by which aromatherapy brings about effects is likely to differ according to the molecular composition of the essential oil and the mode of administration. Similarly, the mechanism of action may vary across outcomes. For example, the mechanism(s) through which aromatherapy might relieve pain may be different from the mechanism for relieving nausea and vomiting [25]. If massage is used as a co-intervention, then the interaction between massage, the essential oil, and the carrier oil may also influence the mechanism [13, 19]. Research on these mechanisms comes predominantly from mainstream neurophysiological research on olfaction and pharmacological research. Some is specific to essential oils, but very little originates from literature on aromatherapy [13]. This research is comparatively recent, and evidence about the mechanisms of action for specific oils and modes of delivery is limited [13, 26]

The prevailing description of how aromatherapy works – and one that aligns intuitively with the practice of aromatherapy – is that aromatherapy acts through the olfactory system. Volatile molecules in the aromatherapy oil (the odorant) interact with receptors in the nose, generating an electrical signal to the brain that triggers the perception of smell [13, 26, 27]. This perception includes responses initiated in the limbic system, which is involved in controlling memory and emotion, and through which odours are thought to produce effects on mood, alertness, mental stress, arousal and perceived health [13]. Biochemical or physiological pathways are likely to mediate the effects of essential oils applied to the skin, where either local or systemic effects may be possible depending on whether the active component diffuses through the skin [26]. Some of these effects are suggested to arise from antibacterial, anti-inflammatory and analgesic properties of essential oils [13, 28, 29].

Aromatherapy professional associations also describe a pathway involving an 'energetic' or spiritual response. Such mechanisms are described as a 'vibrational interaction' between the active component of the essential oil and 'the energy flows within the body' [14].

## 1.3 Description of conditions for which aromatherapy is used

Although texts on aromatherapy describe a breadth of clinical indications, aromatherapy is often used to treat symptoms of a condition and the side effects of treatment rather than the underlying condition. Examples include the use of aromatherapy to alleviate pain, symptoms of anxiety (that occur as a reaction to stress), low mood, sleep disturbance, behavioural disturbance, vomiting and nausea, and fatigue [13, 30-33]. These indications align with the most commonly treated conditions reported in a 2015 survey completed by 36 practising aromatherapists in Australia [21, 34]. Stress was the condition most frequently reported as 'often treated' (by 79% of aromatherapists). Musculoskeletal conditions associated with chronic pain were also frequently reported as often treated, especially neck (64% of aromatherapists), arthritis (54%), sciatica (42%), and knee pain (42%). Other conditions that were reported as 'often treated' were headache and migraine (66%), mental health conditions (40%), insomnia (47%), sports injury (27%), cancer (24%) and palliative care (21%).

There is particular interest in using aromatherapy in circumstances where mainstream interventions may not provide satisfactory relief of symptoms, for example for people with unremitting chronic pain, cancer or advanced disease (not amenable to cure) [13, 31, 35, 36]. Among people with cancer and advanced disease, aromatherapy is used as a form of supportive care to enhance physical and emotional well-being, in addition to alleviating specific symptoms [13, 31, 35, 36]. In other cases, aromatherapy is used as an alternative or adjunctive therapy by those seeking to avoid pharmacological or invasive treatment. For example, aromatherapy has been used to ameliorate behavioural and sleep disturbances among people with dementia [30], to relieve pain during labour [37] and to treat postoperative nausea and vomiting [38]. These treatments may be delivered in a range of healthcare settings (primary, acute and subacute care), with varying levels of integration with conventional providers [39].

Because aromatherapy is often sought or prescribed for relief of symptoms, those receiving aromatherapy for the same indication may have very different underlying conditions (e.g. cancer, arthritis, chronic insomnia) or be undergoing different treatments (e.g. surgery, chemotherapy, minor procedures). Examining the effects of aromatherapy on outcomes for a particular condition may be of interest in some circumstances, but for many commonly treated symptoms or side effects, there is no clear clinical rationale for why the effects of aromatherapy would differ importantly by condition. Where this is the case, a broad synthesis across multiple conditions addresses whether there is a consistent effect for the outcome of interest (benefit, little or no effect, harm), in addition to enabling exploration of whether the effect of aromatherapy differs by condition (e.g. smaller or larger effects).

## 1.4 Why it is important to do this review

This systematic review will inform the Australian Government's Natural Therapies Review 2019-20, which is evaluating evidence of the clinical effectiveness of 16 therapies (including aromatherapy). The conclusion from the evidence evaluation conducted on aromatherapy for the *2015 Review* was that "there was no clear evidence demonstrating efficacy of aromatherapy" [40]. The evidence evaluation used overview methods, synthesising results from 20 systematic reviews published up to May 2013. Of the primary studies included in these systematic reviews (N=45), all but one were published prior to 2012. Since the completion of the original evidence evaluation, there has been substantial growth in published research on aromatherapy. A bibliometric analysis of scientific articles on aromatherapy found a steady increase in the number of primary studies and reviews from 1995 to 2014 [41]. Of the 549 research articles published in this period, a third (N=190) were published between 2012 and 2014. This finding indicated that there could be evidence available (either published in the last five years or not incorporated in systematic reviews at the time the overview was conducted) that would change the conclusions about the effects of aromatherapy [8]. In contrast to the 2015 aromatherapy evidence evaluation, this review examined evidence from eligible primary studies published from database inception until the date of the last search for this systematic review.

# 2. Objectives

The overall objective of this systematic review was to examine the evidence for the clinical effectiveness of aromatherapy in preventing and/or treating injury, disease, medical conditions or preclinical conditions [8]. The review focused on outcomes (and underlying conditions) for which aromatherapy is commonly sought or prescribed in Australia, and which are relevant to the 2019-20 Review of the Private Health Insurance rebate.

The questions for the review follow (framed as primary and secondary objectives). Decisions about the final synthesis questions and criteria for including studies in each synthesis were made through a staged prioritisation process (describe in section 3.4). The prioritisation process aimed to align the questions addressed with priorities for the 2019-20 Review, ensure a consistent approach across the evidence evaluations of natural therapies (where appropriate), and make best use of available evidence.

### Primary objectives to address the following questions were

1. What is the effect of *aromatherapy (delivered by any mode)* compared to an inactive control (placebo, no intervention or usual care) among people with any condition, pre-condition, injury or risk factor on outcomes for which aromatherapy is used (pain, nausea and vomiting, sleep quality, fatigue, emotional functioning and mental health, health-related quality of life, and physical function)?
2. What is the effect of *aromatherapy (delivered by massage)* compared to *massage alone* among people with any condition, pre-condition, injury or risk factor on outcomes for which aromatherapy is used (outcomes as per objective 1)?

### Secondary objectives related to the following questions

3. What is the effect of *aromatherapy (delivered by any mode)* compared to an inactive control (placebo, no intervention or usual care) on outcomes for each underlying condition, pre-condition, injury or risk factor (for example, effects on sleep disruption among people living with cancer, people with chronic insomnia, people with chronic pain or people with dementia)?
4. What is the effect of *aromatherapy (delivered by massage)* compared to *massage alone* on outcomes for each underlying condition, pre-condition, injury or risk factor (for example, effects on sleep disruption among people living with cancer, people with chronic insomnia, people with chronic pain or people with dementia)?
5. What evidence exists examining the effects of aromatherapy compared to active comparators?
6. What evidence exists on the effects of aromatherapy compared to inactive controls or other treatments, for conditions that were not prioritised for the review?

The final synthesis questions and criteria for including studies in each synthesis are presented in Figure 3.5.1.

Outcomes listed in the objectives above were agreed through the prioritisation process.

We planned to examine the effects of aromatherapy compared to “gold standard” (first line) treatments, however the volume of evidence meant it was not feasible to do so. These studies are listed on the evidence inventory (objective 4). Other objectives were as stated in the protocol, with editing to include outcomes and conditions in the final framework.

# 3. Summary of methods

This review followed methods pre-specified in the protocol endorsed by the NHMRC, NTWC and NTREAP. The protocol was prospectively registered on the International prospective register of systematic reviews (PROSPERO ID CRD42021268244) and published in the open access journal Systematic Reviews [1]. The methods were based on the Cochrane Handbook for Systematic Reviews of Interventions [2]. The Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach was used to summarise and assess the certainty of evidence arising from this review [4, 5]. The review is reported in accordance with the PRISMA 2020 statement [6, 7].

A staged approach was taken to developing the questions and criteria for including studies in the synthesis (Figure 3.1). A summary of each stage is described in the methods that follow (see Appendices A and B for a complete description of methods; Appendix I for Abbreviations used in the report). The framework for the synthesis was finalised prior to commencing data extraction (Figure 3.1, panel 4). It defines the scope of the evidence synthesis and specifies the synthesis questions and associated PICO (population, intervention, comparator, outcome) criteria for including studies in each synthesis. Studies that met the eligibility criteria for the review but not the evidence synthesis are reported on the evidence inventory (Appendix E3).

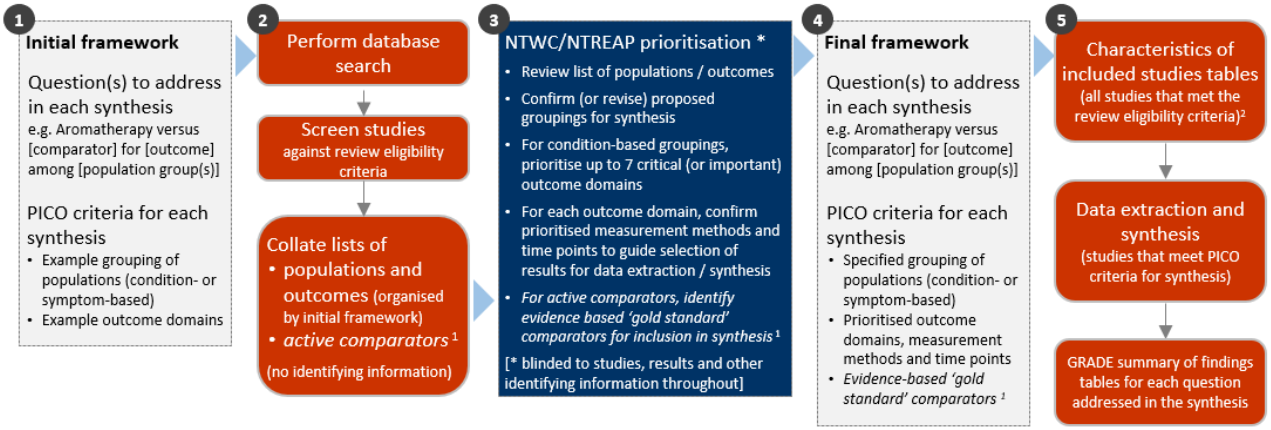


**Fig 3.1** | Staged approach for developing the questions and analytic framework for this review. [1]Active comparators were not considered in the prioritisation process because there were few studies with active comparators. These studies (or study arms) are included on the evidence inventory. [2] Separate tables are presented for studies included for the evidence synthesis (Appendix E1 and E2) and those in the evidence inventory (Appendix E3)

## 3.1 Criteria for considering studies for this review

### 3.1.1 Types of studies

We included randomised controlled trials (RCTs) (including individually and cluster randomised, and crossover trials) and controlled trials where there was an attempt to have some kind of 'randomisation' to groups (e.g. sequence generation based on alternation, dates (of birth or attendance at a clinic) and patient record numbers) [42].

We excluded: non-randomised studies of interventions (NRSIs); studies described as 'randomised trials' or 'controlled clinical trials' without some kind of randomisation (e.g. participants allocated to groups based on clinician choice); and studies for which available reports had not been peer reviewed (grey literature, including theses).

***Date and language restrictions.*** There were no restrictions on publication date. Potentially eligible studies published in languages other than English were eligible for the review but not the synthesis.

### 3.1.2 Types of participants

Studies involving participants with any disease, medical condition, injury, or preclinical condition were eligible for the review. This included healthy participants with clearly identified risk factors for a condition (evident from study eligibility criteria or baseline data) that aromatherapy was administered to prevent. There were no restrictions on age. Healthy populations seeking health improvement were excluded.

As per the provision in the protocol, NTWC/NTREAP reviewed and accepted a proposal to exclude some conditions from the synthesis to ensure the synthesis was manageable (see 3.4). Decisions were guided by whether conditions were identified in the PRACI survey as frequently treated by practitioners in Australia and whether findings could be applicable to other indications for aromatherapy.

### 3.1.3 Types of interventions

Aromatherapy was defined as "Administration of aromatherapy oils by inhalation, diluted topical use and massage" [8]. Aromatherapy treatments were eligible irrespective of the type of essential oil, carrier or dispersant, mode of delivery or route of administration, whether self-administered or provided by a practitioner, the training or qualifications of the practitioner, and the dose and duration of treatment. Interventions were excluded if essential oils were ingested, internally administered, applied topically without dilution, or known to be unsafe for therapeutic use in humans.

**Comparisons**

1. Aromatherapy (delivered by any mode, including massage) *versus* any inactive comparator (placebo/sham, no intervention, wait list control, usual care, or a co-intervention that was given to both groups).

2. Aromatherapy delivered by massage *versus* massage alone (this comparison was included to isolate the effects of aromatherapy)

We excluded head-to-head comparisons of aromatherapy (e.g. comparison of two oils, two modes of delivery or different dilutions of the same oil). Active comparators were eligible for the review but not the synthesis (any pharmacological or non-pharmacological intervention, except natural therapies in other evidence evaluations).

### 3.1.4 Types of outcomes

Any patient-important outcome that aligned with the reasons why aromatherapy is sought by patients and prescribed by practitioners was eligible. Studies were included in the review irrespective of the outcome(s) measured, but the synthesis was limited to outcomes considered to be critical or important for each population group. Experience of care (e.g. satisfaction), safety, quality, and economic outcomes were excluded.

The outcome domains determined to be critical or important for the synthesis through the prioritisation process were:

- Pain
- Nausea and vomiting
- Sleep quality
- Fatigue
- Emotional functioning and mental health
- Health-related quality of life
- Physical function.

From each study, we selected only one outcome per outcome domain for data extraction (results), risk of bias assessment and inclusion in the synthesis. In selecting outcomes for synthesis, we considered the outcome measure (any measure was eligible but a pre-specified hierarchy was applied to select the most relevant measure if multiple were available), timing of outcome measurement (first measure after end of aromatherapy treatment period) and suitability of data for meta-analysis.

## 3.2 Search methods for identification of studies

We searched the Cochrane Central Register of Controlled Trials (CENTRAL via the Cochrane Library, Issue 8, 2021), PubMed, AMED (Allied and Complementary Medicine) via Ovid, and Ovid Emcare on 20 August 2021.

## 3.3 Selection of studies

We piloted guidance for title and abstract screening on a sample of 50 records to ensure the review eligibility criteria were applied consistently. All records were screened independently by two reviewers at both the title and abstract screening and full-text review stages. Disagreements at either stage of screening were resolved by consensus among members of the review team. We documented the flow of studies through the review in a PRISMA diagram (Figure 4.1.1).

Studies that did not meet the review eligibility criteria were excluded and the reason for exclusion was recorded at full-text screening. For studies that originated from the call for evidence, NTREAP, or the Committee, we recorded and reported exclusion decisions irrespective of whether the study was excluded during title and abstract screening or full text review. Studies in languages other than English were included on the list of studies awaiting classification categorised according to whether they were likely to be eligible or eligibility could not be confirmed.

## 3.4 Prioritisation of populations and outcomes for the synthesis

Decisions about the final synthesis questions and criteria for including studies in each synthesis were made through the prioritisation process in Figure 3.1. The process was designed to minimise bias in the selection of results for inclusion in the synthesis while ensuring coverage of relevant populations and outcomes. In brief:

We screened studies against the review eligibility criteria and collated information about the populations and outcomes addressed in the eligible studies.

The information was de-identified and presented in aggregate form so that it was not possible to identify the studies (no bibliographic information, titles etc). No information was provided about the number of studies, number of participants, methodological quality of studies or results.

NTWC/NTREAP prioritised outcome domains, population groups and population-specific outcomes for the synthesis (Figure 3.1).

***Prioritisation and selection of population-specific outcomes***. To prioritise outcomes for each population we:

- Compiled a list of population-specific outcomes from included studies and example outcome measures.
- Categorised outcomes by the outcome domains and population groups in the initial framework for the review (Appendix A1). Outcomes in other domains were also listed.
- Asked NTWC to indicate whether each of the listed outcome domains (or population-specific outcomes) was critical, important or of limited importance for understanding the effects of aromatherapy on each population group. Only critical and important outcomes were considered in the synthesis.

## 3.5 Final framework: synthesis questions and criteria for including studies in each synthesis

The final analytic framework for the review and the evidence synthesis as agreed through the prioritisation process is presented in Figure 3.5.1. Panel A shows the final list of populations and outcome domains eligible for the evidence synthesis. There is a meta-analysis for each outcome domain with population groups within as listed.

***Population groups included in the synthesis***. Some refinements were made to the populations listed in the initial framework (Figure 3.5.1, Panel A). We separated acute conditions and indications from chronic and longer-term conditions, to provide greater clarity about which outcomes were relevant. For example, for people with osteoarthritis undergoing knee replacement surgery, the population was categorised as 'surgery' rather than 'chronic' if treatment was focused on outcomes in the acute perioperative period rather than longer-term outcomes.

***Population groups excluded from the synthesis.*** Given the number of studies included in the review, agreement was reached through the prioritisation process to exclude studies of aromatherapy for the treatment of skin conditions (22 studies), skin infections, infestations or wounds (20 studies), and substance withdrawal (2 studies) from the synthesis (Panel B). Outcomes specific to these population groups were also excluded from the synthesis (Panel B, blue boxes).

***Prioritised outcomes***. The outcome domains specified in the initial framework were endorsed, and the outcomes relevant to each population groups were agreed with some refinement to the presentation in the initial framework.

- An additional outcome domain was added (physical function).
- Fatigue, health-related quality of life and physical function were considered for chronic and longer-term conditions only (where aromatherapy treatment was over weeks or longer (not days) and outcomes measured in a timeframe likely to detect meaningful improvement in the outcome).
- The outcomes listed in Panel B, white box, were not prioritised for any population.

***Outcome measures.*** A hierarchy of outcome measures was agreed for each population and the timeframe for outcome measurement prioritised (for outcome selection when a study reported multiple measures for an outcome domain).

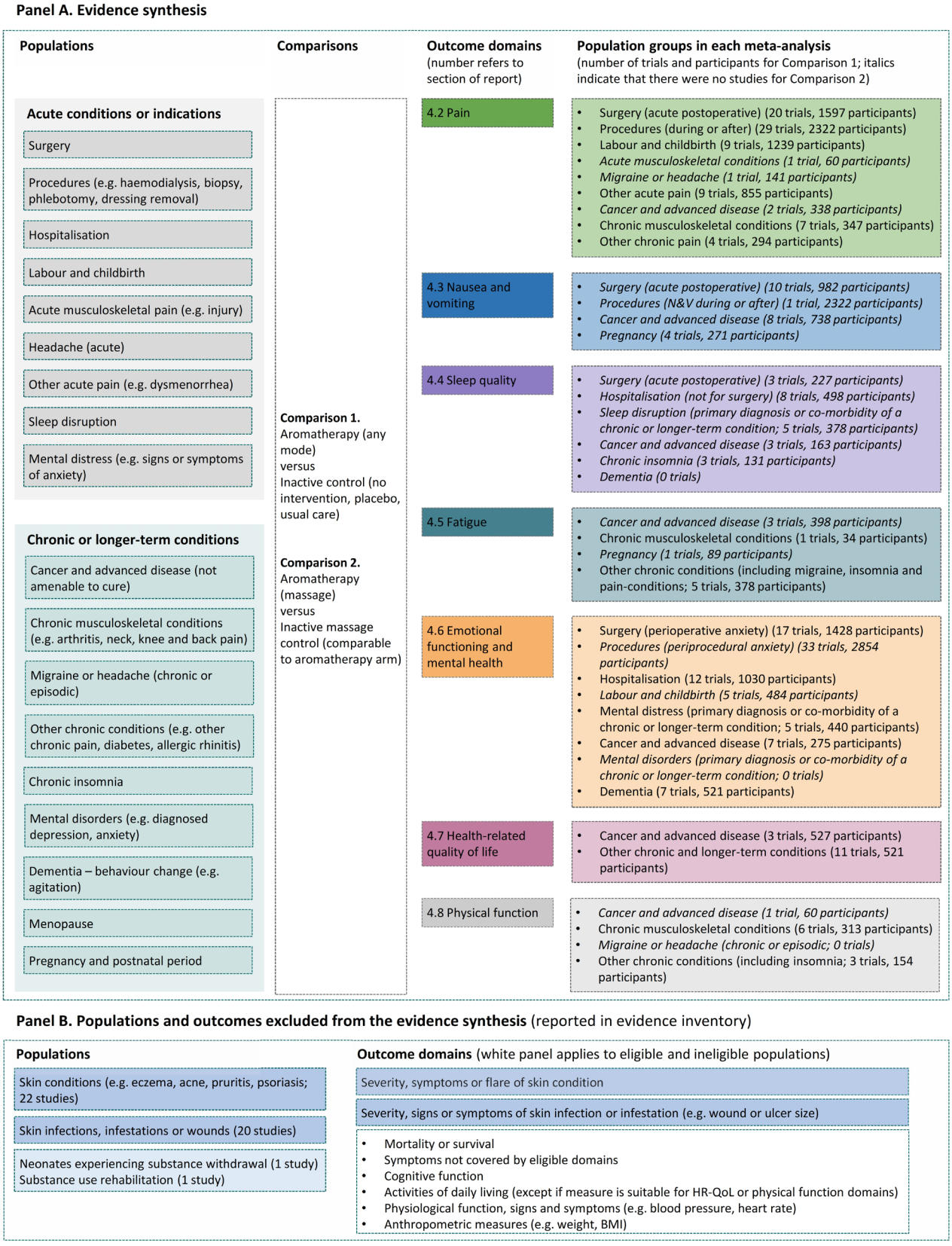


**Fig 3.5.1** | Final analytic framework for the review as agreed through the prioritisation process (Appendix A5). Panel A, column 1 lists population groups eligible for the synthesis, column 2 the outcome domains that form the basis of meta-analyses, and column 3 the populations included for each analysis. Panel B, blue boxes show populations and associated outcomes excluded from the synthesis to limit the size of the review. The white box shows outcomes that were not prioritised as critical or important for any eligible population group. Studies that only reported one or more of these ineligible outcomes are in the evidence inventory.

## 3.6 Data extraction and management

### 3.6.1 Data extraction

Study data were collected and managed using REDCap electronic data capture tools [43, 44]. A two-step data extraction process was implemented wherein a senior author (SB, MM) coded the study PICO to allocate studies for analysis according to the analytic framework and selected the outcome (result) for inclusion in each synthesis using pre-specified decision rules. For each included study, one review author (KB, IF, PN, AS, ST) then extracted study characteristics and quantitative data. A second author (MM) independently verified the data. Steps taken to ensure the completeness, accuracy and consistency of data included pretesting the form and providing coding guidance, training, and feedback for data extractors. Quantitative data were reviewed by a biostatistician when queries arose.

### 3.6.2 Assessment of risk of bias in individual studies

We assessed the risk of bias in included studies using the revised Cochrane 'Risk of Bias' tools (RoB 2) for randomised trials [42, 45, 46]. After piloting of the tools by senior authors (SB, MM, SM, AS), we developed review-specific guidance for assessors (KB, IF, PN, AS). One review author (KB, IF, PN, AS) then applied the tool to the selected results from each study following the RoB 2 guidance [42], and a second author (SB) checked assessments. Supporting information and justifications for judgements for each domain (low, some concerns, high risk of bias) was recorded. We derived an overall summary of the risk of bias from each assessment, following the algorithm in the RoB 2 guidance as implemented in the excel assessment tool [42].

### 3.6.3 Measures and interpretation of treatment effect

We anticipated that many of the outcomes would be continuous (e.g. pain, anxiety), and that varying measurement instruments would be used to measure the same underlying construct across the studies. For this reason, we quantified the effects of aromatherapy using the standardised mean difference (SMD).

Our interpretation was based on whether there was an important effect or not [3, 47], with an SMD of 0.2 standard units set as the threshold for an important difference. If the SMD fell within the pre-specified range of -0.2 to 0.2 (i.e. within both thresholds), the effect of aromatherapy was considered to be no different from control. An SMD above 0.2 or below -0.2 was interpreted as an important effect. We opted to use the most intuitive interpretation of effect estimates for each outcome, so positive values indicate benefit for some outcomes (an increase in sleep quality, function, health-related quality of life, and physical function) and harm for other outcomes (an increase in pain, nausea and vomiting, anxiety or agitation, and fatigue).

## 3.7 Data synthesis

### 3.7.1 Meta-analysis

Separate comparisons were set up based on outcome domains agreed in the final framework (see Figure 3.5.1). These comparisons were stratified by the population groups in the final framework. This approach to structuring the meta-analysis yielded an overall estimate of the effect of aromatherapy for the outcome (review objectives 1 and 2), as well as estimates within each population group (review objective 3 and 4). Forest plots were used to visually depict the intervention effect estimates and their confidence intervals. Forest plots are stratified by condition and risk of bias (within population group).

### 3.7.2 Summary of findings tables and assessment of certainty of the body of evidence

For each result, one author (SB) used the GRADE approach to assess our certainty in whether there is an important effect (or not). In accordance with GRADE guidance [4, 47, 48], an overall GRADE of high, moderate, low or very low certainty is reported for each result based on whether there are serious, very serious or no concerns in relation to each of the following domains [3].

- **Risk of bias**. whether the studies contributing each synthesis have methodological limitations that might lead to over (or under) estimation of the effect
- **Imprecision**. whether the confidence interval for the synthesised result crosses one or both of the thresholds for an important effect (an SMD of 0.2 or -0.2) meaning that the result is compatible with different

interpretations (e.g. the upper bound of the interval lies above 0.2 indicating 'an important effect' whereas the lower bound lies between -0.2 and 0.2 indicating 'little or no effect')
- **Inconsistency**. whether there are important, unexplained inconsistency in results across studies
- **Indirectness**. whether there are important differences between the characteristics of studies included in each synthesis and the question we were seeking to address, such that the effects observed may not apply to our question (i.e. the applicability of the evidence).
- **Publication bias**. whether results missing from each analysis may bias the effect estimate they are selectively not reported because the results (or studies) showed unfavourable effects

A summary of findings is tabulated for each meta-analysis. These summary of findings tables include:

- estimates of the effects of aromatherapy reported as standardised mean differences
- the overall GRADE (rating of certainty) and an explanation of the reason(s) for rating down (or borderline decisions) [49].
- the study design(s), number of studies and number of participants contributing data
- a plain language statement interpreting the evidence for each comparison and outcome, following GRADE guidance for writing informative statements (see 3.7.3 interpretation of findings) [50].

### 3.7.3 Interpretation of findings (evidence statements)

When interpreting results, we followed GRADE guidance for writing informative statements [50]. All interpretations are based on where the point estimate lies in relation to the pre-specified thresholds for an important effect (an important effect or not) and the direction of effect (beneficial or harmful). The certainty of evidence is communicated by qualifying the interpretation of effect (e.g. 'probably' improves for moderate certainty). For low certainty evidence the interpretation is qualified with the word 'may'. For example, 'Aromatherapy may improve sleep quality' indicates that the point estimate lies above the threshold for important benefit (an SMD >0.2) and that the evidence is of low certainty. For very low certainty evidence, we do not provide an interpretation of the result except to state 'The evidence is very uncertain about the effect of aromatherapy on outcome'. This is one of two options that GRADE provides for interpreting findings based on very low certainty of evidence. The decision not to interpret very low certainty results was made independently by the NTWC to ensure a consistent and clear interpretation of findings across Natural Therapy Review reports (see Appendix G).

# 4. Results

## 4.1 Results of the search

The flow of studies through the review is summarised in Figure 4.1.1, the PRISMA flowchart.

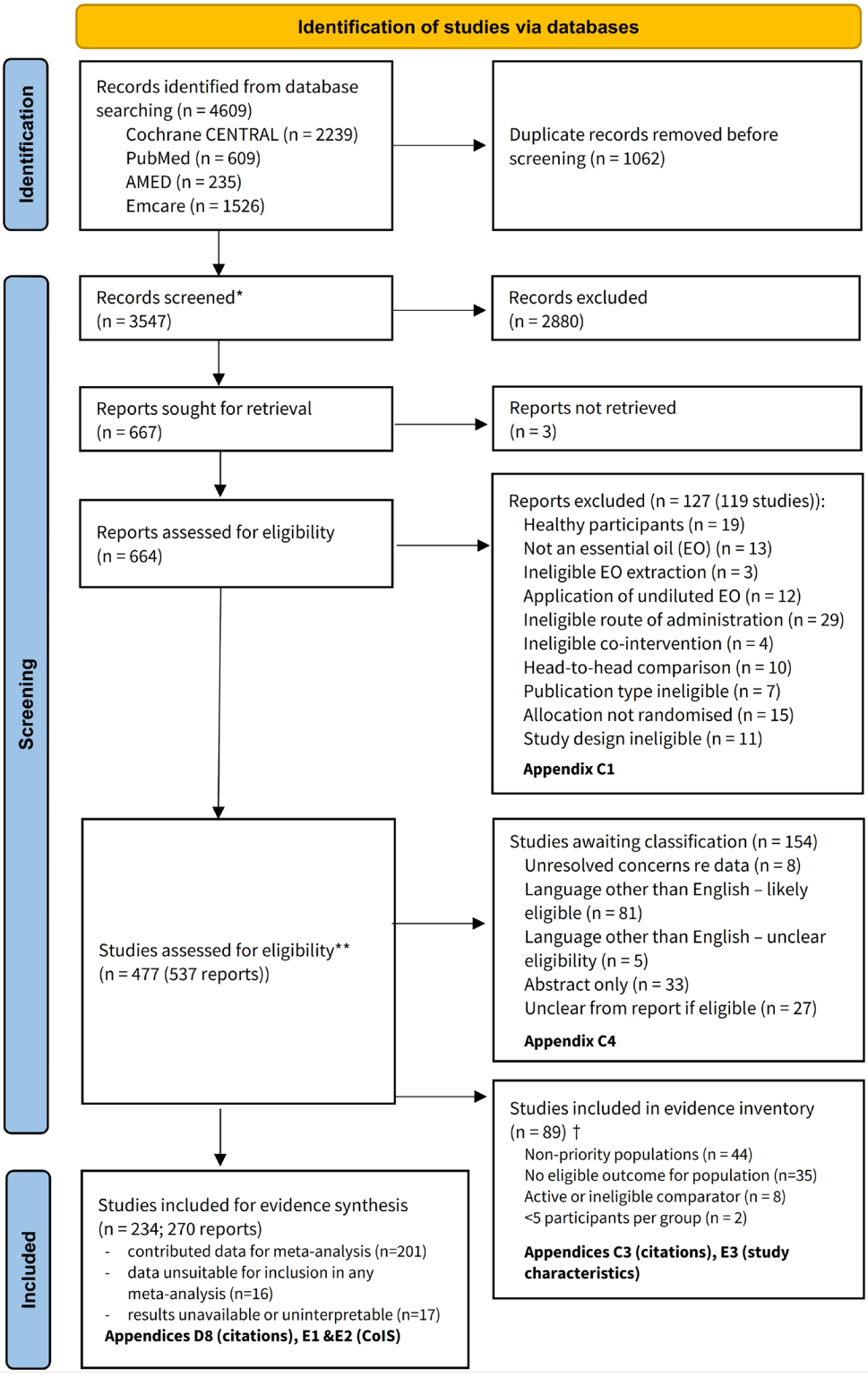


**Fig. 4.1.1** | PRISMA diagram showing the flow of studies through the review. * In addition to records from the search, 134 public submissions were received and screened, of which 26 were unique records. All 26 were systematic reviews and therefore excluded (see Appendix C2). ** Studies are the unit of interest in the review. For each study there may be multiple reports. † Exclusion of these studies from synthesis was agreed through the prioritisation process (Fig 3.5.1; Methods appendix A5, A6). CoIS: characteristics of included studies.

## Included studies

### Studies included for the evidence synthesis

Studies were eligible for one or both of two main comparisons

Comparison 1 (C1). Aromatherapy (any mode) versus inactive control (usual care, placebo, no intervention)

Comparison 2 (C2). Aromatherapy (massage) versus massage

For each of these comparisons, studies could contribute to the synthesis for one or more of seven outcome domains.

1. Pain
2. Nausea and vomiting
3. Sleep
4. Fatigue
5. Emotional functioning and mental health
6. Health-related quality of life (HR-QoL)
7. Physical functioning

Of the 234 studies that were eligible for the evidence synthesis (i.e., studies that addressed at least one comparison and outcome of interest):

- 201 contributed data to at least one meta-analysis; of these, 193 reported data required for inclusion in all of the meta-analyses for which the study was eligible and 8 had required data for a subset only,
- 33 studies did not report data suitable for inclusion in any of the meta-analyses for which the study was eligible.

A breakdown of the number of studies and participants included and missing from each analysis is shown in Table 4.1.1.

Three sets of studies are presented in the table as follows.

**Set (a) is studies included in each meta-analysis**. For these, the data required for meta-analysis was reported, able to be calculated or, if necessary, imputed (see Appendix B for assumptions made when calculating the required statistics and Appendix D for results of sensitivity analyses testing the impact of these decisions).

**Sets (b) and (c) are studies missing from each meta-analysis.**

- Set (b). Data from these studies was *unsuitable for meta-analysis* because the required statistics were unavailable and could not be calculated or imputed (details in Appendix D). While these results were interpretable, the extent and nature of incomplete reporting raised concerns about the validity of results. For this reason, the results have not been summarised.
- Set (c). Results from these studies were *unavailable or could not be interpreted*, and as such the study is missing from a meta-analysis from which it is eligible. The reasons are as follows.
    - No result was reported for a measured outcome.
    - The required statistics were incompletely or ambiguously reported.
    - The results could not be interpreted in relation to the synthesis PICO question (e.g. an effect estimate from a model that combined eligible and ineligible intervention/comparator groups; no information in the paper or other sources from which to interpret an outcome measure or direction of effect; an effect estimate adjusted for multiple 'baseline' measures of the outcome collected post-randomisation).
    - Major or multiple errors in the data, with substantial impact on the results (e.g. reporting % that were not possible for the number of participants).

**Table 4.1.1**. Summary of the number of studies and participants included and missing from each analysis

| Outcome domain | | Studies in MA Set (a) | | Studies with results missing from the meta-analysis | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Set (b). data unsuitable | | Set (c). results unavailable or uninterpretable | | Total missing | |
| | | No. studies | Participants | No. studies | Participants | No. studies | Participants | No. studies | Participants |
| Pain | C1 | 82 | 7193 | 11 | 1011 | 7 | 1297 | 18 | 2308 |
| | C2 | 19 | 1058 | 1 | 287 | 1 | 80 | 2 | 367 |
| N&V | C1 | 23 | 2032 | 0 | 0 | 4 | 428 | 4 | 428 |
| | C2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Sleep | C1 | 22 | 1397 | 3 | 303 | 1 | 70 | 4 | 373 |
| | C2 | 0 | 0 | 1 | 150 | 0 | 0 | 1 | 150 |
| Fatigue | C1 | 18 | 1316 | 2 | 359 | 1 | 90 | 3 | 449 |
| | C2 | 4 | 252 | 0 | 0 | 0 | 0 | 0 | 0 |
| EFMH | C1 | 86 | 7032 | 4 | 313 | 9 | 1005 | 13 | 1318 |
| | C2 | 11 | 664 | 2 | 92 | 3 | 168 | 5 | 260 |
| HR-QoL | C1 | 14 | 1048 | 2 | 164 | 2 | 189 | 4 | 353 |
| | C2 | 12 | 851 | 1 | 118 | 0 | 0 | 1 | 118 |
| Physical function | C1 | 10 | 527 | 2 | 238 | 0 | 0 | 2 | 238 |
| | C2 | 7 | 434 | 1 | 118 | 0 | 0 | 1 | 118 |
| *TOTAL (unique in each set) | | 201 | | 16 not in any MA (20 total) | 1944 | 17 not in any MA (21 total) | 2471 | 41 | 4415 |

* Studies may contribute multiple outcomes or comparisons, hence values in yellow and red columns do not sum. Totals in bottom row are the sum of unique studies/participants missing from all analyses.
**Abbreviations**: MA – meta-analysis; C1 - Comparison 1. AT (any mode) vs inactive control (placebo, no intervention, usual care); C2 - Comparison 2. AT (massage) vs inactive massage control; N&V – nausea and vomiting, EFMH – emotional functioning and mental health, HR-QoL – health-related quality of life

**Studies included in the evidence inventory**

Of the 323 studies included in this review, 89 were included in the evidence inventory but not the evidence synthesis. Reasons for excluding these studies from the synthesis are summarised in Figure 4.1.1, study characteristics are reported per study in Appendix E3 and references are in Appendix C3.

In brief, the majority of studies (79 of 89) were excluded from the synthesis after priority populations and outcomes for the review were agreed through the independent prioritisation process (Appendix A5 for method and A6 for results).

***Population*** exclusions implemented to manage the scope of the synthesis were as follows.

- Skin infections, infestations and wounds (20 studies)
- Skin conditions (e.g. eczema, acne, pruritis, psoriasis) (22 studies; two additional studies where the population was categorised as having a chronic condition but the outcomes were ineligible for the population)
- Neonatal substance withdrawal (1 study)
- Substance use rehabilitation (1 study)

***Outcomes.*** Studies that did not measure an outcome eligible for the synthesis (i.e. rated as critical or important in the prioritisation process) are as follows.

Twenty-two studies only measured an ineligible outcome (i.e. not included in the final analytic framework).

- Physiological function, signs and symptoms (11 studies)
- Other symptoms (e.g. duration of labour, enuresis, 11 studies)

Thirteen studies measured an outcome listed in the final analytic framework for the review, but the outcome was ineligible for the population (some measured more than one ineligible outcome).

- Emotional functioning/mental health in populations without evidence of mental distress, mental disorder or at risk of situational anxiety such as prior to surgery or a procedure (6 studies)
- Sleep quality in populations without evidence of sleep disturbance (6 studies)
- HR-QoL measured less than one month from commencement of aromatherapy (5 studies). Exceptions were made to this rule if a core outcome set or similar indicated that shorter term follow up was acceptable (Appendix A1.1.4).

The remaining ten studies excluded from the synthesis are as follows.

- Eight studies had a comparator that was active (e.g. music therapy, topical diclofenac) or another natural therapy (reflexology, 1 study).
- Two studies had 5 participants per group or less. While these studies were described as randomised, they were deemed to have too few participants for randomisation to be successful.

### Excluded studies

After full-text screening, 119 studies (127 reports) were excluded from the review. Reasons for exclusion are summarised in Figure 4.1.1 and reported per study in Appendix C1. Of the 127 excluded reports, the majority were excluded because the intervention or comparison was ineligible. This included:

- undiluted topical application of essential oils (12 studies),
- ineligible routes of administration (e.g., ingested or internal; 29 studies),
- synthetic products, other aromatic products that were not essential oils, and oils extracted using solvents (26 studies), and
- comparisons of two forms of aromatherapy (10 studies).

### Studies awaiting classification

Following screening and linking of multiple study reports, 154 studies were categorised as awaiting classification. The reasons why a decision could not be made on the eligibility of these reports are summarised in Figure 4.1.1 and reported per study in Appendix C4. For 119 of these studies, a decision had to be made on the basis of title and abstract alone. These included studies reported in conference abstracts (33 studies that could not be matched to a full text report) and studies for which the full text report was in a language other than English (86 studies).

#### Studies in languages other than English

Of the 86 studies in languages other than English, 81 were judged likely to be eligible based on the title and abstract (listed in Appendix C4). Because study design and characteristics tend to be incompletely reported in abstracts (especially the outcomes measured), the proportion of these studies eligible for the review and the evidence synthesis is unknown. For these reasons, a full analysis of the impact of these studies on each of the meta-analyses was not possible, however the likely implications of non-inclusion of these studies in the synthesis is as follows.

- **Implications of study in languages other than English**. There is no reason to believe that, on average, the results from studies in languages other than English would differ systematically from studies included in our analysis. Given this, and the volume of evidence contributing to each analysis, non-inclusion of these studies is unlikely to change the results or conclusions for each outcome.

### Ongoing and unpublished studies

Our search of trial registry entries from CENTRAL identified 1120 records, of which 91 were linked to included studies. Of the remaining records, 500 were screened to determine the likely proportion of records requiring further screening. A high proportion were identified as potentially eligible (i.e. based on the limited information contained in the CENTRAL record). Given the volume of records and low likelihood that the records would provide additional information for assessing the impact of missing results, we decided that it was not worthwhile to screen the full registry records. While we are unable to quantify the number of ongoing and unpublished studies, the number is likely to be substantial given

the number of records and proliferation of studies in recent years. For unpublished studies listed in registry records, it is difficult to distinguish between studies that are yet to be completed (truly ongoing) and studies that remain unpublished because the findings were considered by the trialists or others to be unfavourable (harm or little or no benefit). The two have different implications for the results and conclusions, as follows.

- **Implications of ongoing studies**. As with studies in languages other than English, there is no reason to believe that, on average, the results of ongoing studies would differ from those of studies included in our analysis. Given this, and the large amount of data contributing to each analysis, non-inclusion of these studies is unlikely to change the results or conclusions for each outcome.
- **Implications of non-reporting of completed studies**. Non-reporting of completed studies is of concern because of potential that these missing results bias the estimates of effect. We consider the potential for bias due to missing results from the synthesis in relation to our synthesis of results for each outcome. Because of the large amount of data contributing to each analysis, we were able to use sensitivity analyses and funnels plots to determine whether missing results were likely to bias the estimates of effect (detailed in Appendix D and considered in GRADE judgements of publication bias, as reported in Summary of Findings tables).

## 4.2 Pain

Overall, 90 studies that examined the effect of aromatherapy on pain were included for meta-analysis, and 11 of these contributed to both Comparisons 1 and 2. A further 19 studies (2388 participants) were eligible for one or both of the pain meta-analyses, but could not be included (see below).

***Comparison 1***.

- 82 studies (7193 participants) contribute to the comparison of aromatherapy delivered by any mode compared to an inactive control (Figure 4.2.1).
- An additional 18 trials (2308 participants) were eligible for this comparison, but either did not report results that could be included in the meta-analysis, or the results were unavailable or uninterpretable.

***Comparison 2***.

- 19 studies (1058 participants) contribute to the comparison of aromatherapy delivered by massage to an inactive massage control (comparable to massage in the aromatherapy arm) (Figure 4.2.2).
- Two additional trials (367 participants) were eligible for this comparison, one reported results that were unsuitable for analysis and the other had unavailable/uninterpretable results.

### Characteristics of included studies

#### Types of populations

Included studies examined the effect of aromatherapy on pain

- after specific types of surgery (e.g. caesarean, Coronary Artery Bypass Graft (CABG), open heart, abdominal) in the acute postoperative period
- during or after a procedure among adults (e.g. haemodialysis, burn dressing changes, gynaecological procedures) or children (e.g. phlebotomy, vaccination, dental procedures)
- from chronic musculoskeletal conditions (mainly knee osteoarthritis)
- from acute musculoskeletal conditions (one trial among people with fracture)
- from headache and migraine (one trial among people with migraine)
- from cancer or advanced disease (one study among people undergoing chemotherapy, a second unspecified)
- during labour and childbirth
- from other chronic conditions (mainly neuropathy)
- from other acute or episodic conditions (mainly dysmenorrhoea)

The specific condition addressed in each trial is reported in the forest plot (column 3, Figure 4.2.1 and Figure 4.2.2) with full details for each study including eligibility criteria, participant characteristics, and ICD 11 codes in Appendix E1.

#### Types of interventions

***Comparison 1 (any mode of aromatherapy delivery).*** Of the 82 trials included in Comparison 1, aromatherapy was delivered by inhalation in 60 trials, by massage in 16 trials and topically in 6 trials. Thirteen of the 82 trials examined the effects of two or more aromatherapy treatments that we combined prior to inclusion in the meta-analysis. These were different essential oils (7 trials), a different dose, duration or timing of the same essential oil (4 trials), or aromatherapy with a co-intervention (2 trials).

Lavender was the most commonly evaluated essential oil, either alone (46/82 trials) or in a blend (4 trials), followed by rose (11 trials, one in a blend with lavender), eucalyptus oil (4 trials), chamomile (3 trials), ginger (2 trials), and orange (2 trials). Many other essential oils were evaluated in a single trial (e.g. peppermint, tea-tree, rosemary, nutmeg, sage). Four trials evaluated a blend of essential oils, and one trial gave participants a choice of several oils.

The treatment period varied in length, but this generally reflected the treatment goal (i.e. for an acute or chronic indication).

- For people undergoing surgery or procedures, aromatherapy was administered soon after surgery or during/immediately after the procedure, with one or multiple doses delivered within 24 hours in most trials (42/49 trials). The single study of acute musculoskeletal pain (fracture) had a similarly short aromatherapy treatment period.

- For people with chronic musculoskeletal conditions and other chronic pain, the aromatherapy intervention period was of longer duration, ranging from two to six weeks (daily or weekly treatments depending on mode).
- For migraine, the single trial had a 3-month aromatherapy treatment period (weekly treatments).
- For people living with cancer, the trial among those receiving aromatherapy for pain while undergoing chemotherapy received one week of aromatherapy, whereas the trial aimed at relieving cancer symptoms involved a four-week period of aromatherapy.
- For labour and birth, all but one trial delivered aromatherapy during the first stage of labour.
- For other acute pain, including episodic pain such as dysmenorrhea, the aromatherapy intervention period was variable, ranging from short term (less than one day) to treatment over months (multiple menstrual cycles).

***Comparison 2 (aromatherapy delivered by massage)***. The use of essential oil blends was relatively common in trials that compared aromatherapy massage to massage alone (6/19 trials). Lavender, alone or in a blend, was the most commonly evaluated essential oil (11/19 trials).

The treatment period was similar in studies that delivered aromatherapy by massage and by other modes.

- For chronic musculoskeletal conditions and other chronic pain, weeks to months.
- For most acute indications, the treatment period was short: within 24 hours of surgery for perioperative pain, during labour, and the days immediately prior to or at onset of menstruation.
- For procedural pain, the two trials were among dialysis patients, who received aromatherapy massage at dialysis sessions over a 4-week period.

**Types of outcomes**

The outcome measure from which data were included for meta-analysis is reported for each trial in the forest plots (column 2, Figure 4.2.1 and Figure 4.2.2). Full details for each study are in Appendix E1, including the timing of outcome measurement in relation to intervention and details of which outcome was selected when multiple were available.

All studies measured pain intensity on a scale, almost all using either a visual analogue scale (VAS) or a numeric rating scale (NRS). Exceptions included for chronic musculoskeletal pain (4 trials used the WOMAC-pain scale), cancer (where cancer-specific measures such as the pain scale from EORTC-QLQ-C30 was used), and studies of procedural pain among children (where child-specific pain measures were used, including the Neonatal Infant Pain Scale (NIPS), the COMFORT scale, and the Oucher scale).

Most results were reported as a score on the original scale (e.g. pain intensity on a VAS). Three trials presented results as ordinal data (e.g. in categories such as mild, moderate or severe pain); no trials reported dichotomous outcomes (e.g. the proportion of patients who met a predefined threshold for reduction in pain, such as a 30% reduction in pain intensity). Where possible, we selected a result reported on the original scale. When necessary, an effect estimate was calculated from ordinal data (odds ratio) and transformed to a standardised mean difference (Appendix B).

### Effects of aromatherapy on pain

**Comparison 1: Aromatherapy (any mode) versus inactive control (usual care, placebo, no intervention)**

The effect of aromatherapy on pain is uncertain overall (all population groups). Aromatherapy may improve pain for chronic musculoskeletal pain and acute or episodic pain conditions (mainly dysmenorrhea) but the effect is uncertain for each other population group (surgery, procedures, acute musculoskeletal pain, headache or migraine, cancer and advanced disease, labour and childbirth, other chronic pain).

Factors that reduced our certainty in the combined estimates of effect differed somewhat for each population group, as explained in the GRADE summary of findings table (Table 4.2.1, explanations). In combination, these factors raise concern that any observed benefit could be overestimated (or harm underestimated). Major concerns are as follows.

- ***Publication bias***. There is evidence that there could be studies (or results) missing from the analysis that show effects favouring the control (i.e. selective non-reporting based on the direction and statistical significance of results).
- ***Risk of bias in included trials***. All trials in the analysis have methodological limitations (high risk of bias or some concerns). The absence of trials at low risk of bias meant that it was not possible to examine the impact

of these methodological limitations on the estimate of the intervention effect using the approach specified in our protocol (limiting analyses to studies at low risk of bias).

- ***Inconsistent results that lead to different conclusions about the effects of aromatherapy***. There is evidence that the size of the intervention effect differs across studies beyond what would be expected by chance. These differences were not explained by population group (Figure 4.2.1; as explained in Appendix D, Section D.1) or the mode by which aromatherapy was delivered (Appendix D, Section D.1 and Figure D.1.1). This reduced our confidence in the combined estimate because some studies found an important reduction in pain (greater than the threshold for important benefit, an SMD of - 0.2 or lower) while others found little or no difference between aromatherapy and control, with no credible evidence to explain whether this reflects true differences in the effects of aromatherapy or methodological problems in some studies.

Concerns relating to each finding were considered in the GRADE assessment when interpreting the result. The findings are as follows.

**Results for which an interpretation was made** (low certainty evidence)

- ***Chronic musculoskeletal pain***. Aromatherapy (any mode) may reduce chronic musculoskeletal pain (SMD 1.00 lower, 95% CI 1.56 lower to 0.43 lower; $I^2$ = 75 %; 7 studies, 347 participants; low certainty, Figure 4.2.1).
- ***Other acute pain***. Aromatherapy (any mode) may reduce otheracute pain (mainly dysmenorrhea, renal colic) (SMD 1.01 lower, 95% CI 1.53 lower to 0.48 lower; $I^2$ = 89%; 9 studies, 855 participants; low certainty, Figure 4.2.1).

**Results considered too uncertain to interpret**

- ***Overall (all population groups).*** The evidence is very uncertain about whether aromatherapy reduces pain (82 studies, 7193 participants; very low certainty, Figure 4.2.1).
- ***Surgery***. The evidence is very uncertain about the effect of aromatherapy (any mode) on pain after surgery (in acute postoperative period) (20 studies, 1597 participants; very low certainty, Figure 4.2.1).
- ***Procedures***. The evidence is very uncertain about the effect of aromatherapy (any mode) on pain during or after a procedure (acute procedural period) (29 studies, 2322 participants; very low certainty, Figure 4.2.1).
- ***Acute musculoskeletal pain***. The evidence is very uncertain about the effect of aromatherapy (any mode) on acute musculoskeletal pain (1 trial, 60 participants; very low certainty).
- ***Headache and migraine (chronic or episodic).*** The evidence is very uncertain about the effect of aromatherapy (any mode) on headache or migraine pain (1 study, 141 participants with migraine; very low certainty).
- ***Cancer and advanced disease***. The evidence is very uncertain about the effect of aromatherapy (any mode) on cancer pain (2 studies, 338 participants; very low certainty, Figure 4.2.1).
- ***Labour and childbirth***. The evidence is very uncertain about the effect of aromatherapy (any mode) on pain during labour and childbirth (9 studies, 1239 participants; very low certainty, Figure 4.2.1).
- ***Other chronic pain***. The evidence is very uncertain about the effect of aromatherapy (any mode) on other chronic pain (4 studies, 294 participants).

**Table 4.2.1**. Summary of findings for Comparison 1. the effect of aromatherapy (any mode) versus inactive control (usual care, no intervention, placebo) on pain.

| Outcomes | Anticipated absolute effects* (95% CI) | | Relative effect (95% CI) | № of participants (studies) | Certainty of the evidence (GRADE) | Comments |
|---|---|---|---|---|---|---|
| | **Risk with inactive control (usual care, placebo, no intervention)** | **Risk with aromatherapy (any mode)** | | | | |
| Pain: All population groups[a] | - | SMD **1.29 SD lower** (1.62 lower to 0.96 lower) | - | 7193 (82 RCTs) | ⨁◯◯◯ Very low[b,c,d] | The evidence is very uncertain about the effect of aromatherapy (any mode) on pain. |

| Outcomes | Anticipated absolute effects* (95% CI) | | Relative effect (95% CI) | № of participants (studies) | Certainty of the evidence (GRADE) | Comments |
|---|---|---|---|---|---|---|
| | **Risk with inactive control (usual care, placebo, no intervention)** | **Risk with aromatherapy (any mode)** | | | | |
| Pain after surgery (acute postoperative period)[a] | - | SMD **1.17 SD lower** (1.75 lower to 0.58 lower) | - | 1597 (20 RCTs) | ⨁◯◯◯ Very low[d,e,f,g] | The evidence is very uncertain about the effect of aromatherapy (any mode) on pain after surgery (in acute postoperative period). |
| Pain during or after a procedure (acute procedural period)[a] | - | SMD **0.97 SD lower** (1.3 lower to 0.65 lower) | - | 2322 (29 RCTs) | ⨁◯◯◯ Very low[d,e,h] | The evidence is very uncertain about the effect of aromatherapy (any mode) on pain during or after a procedure (acute procedural period). |
| Pain: chronic musculoskeletal conditions (knee OA, knee pain, carpal tunnel syndrome, rheumatoid arthritis) | - | SMD **1 SD lower** (1.56 lower to 0.43 lower) | - | 347 (7 RCTs) | ⨁⨁◯◯ Low[d,e,i] | Aromatherapy (any mode) may reduce chronic musculoskeletal pain. |
| Pain: acute musculoskeletal conditions (fracture, emergency department care) | - | SMD **1.26 SD lower** (1.18 lower to 0.71 lower) | - | 60 (1 RCT) | ⨁◯◯◯ Very low[d,j,k,l] | The evidence is very uncertain about the effect of aromatherapy (any mode) on acute musculoskeletal pain. |
| Pain from headache or migraine (chronic or episodic) | - | SMD **2.76 SD lower** (4.31 lower to 1.2 lower) | - | 141 (1 RCT) | ⨁◯◯◯ Very low[d,j,k,m,n] | The evidence is very uncertain about the effect of aromatherapy (any mode) on headache or migraine pain. |
| Pain: cancer & advanced disease[a] | - | SMD **0.14 SD lower** (0.43 lower to 0.14 higher) | - | 338 (2 RCTs) | ⨁◯◯◯ Very low[d,o,p,q] | The evidence is very uncertain about the effect of aromatherapy (any mode) on cancer pain. |
| Pain during labour and childbirth | - | SMD **2.32 SD lower** (4.01 lower to 0.64 lower) | - | 1239 (9 RCTs) | ⨁◯◯◯ Very low[d,e,r] | The evidence is very uncertain about the effect of aromatherapy (any mode) on pain during labour and childbirth. |
| Pain: other chronic (mainly neuropathic pain) | - | SMD **3.72 SD lower** (11.07 lower to 3.63 higher) | - | 294 (4 RCTs) | ⨁◯◯◯ Very low[d,e,s,t] | The evidence is very uncertain about the effect of aromatherapy (any mode) on other chronic pain. |
| Pain: other acute (mainly dysmenorrhea, renal colic) | - | SMD **1.01 SD lower** (1.53 lower to 0.48 lower) | - | 855 (9 RCTs) | ⨁⨁◯◯ Low[d,e,u] | Aromatherapy (any mode) may reduce other acute pain (mainly dysmenorrhea, renal colic). |

***The risk in the intervention group** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI). **CI:** confidence interval; **SMD:** standardised mean difference
The threshold for an important difference was an SMD of 0.2 (used for interpreting point estimates and confidence intervals). For pain, the resulting interpretation is: < -0.2 is beneficial, -0.2 to 0.2 is trivial or unimportant ("little or no difference" between treatments), > 0.2 is harmful

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

## Explanations

a. Measures varied. VAS, VRS, NRS and some condition- or population-specific measures.

b. Serious risk of bias (-1). All 82 studies in the analysis are at high risk of bias or some concerns, such that the observed benefit may be overestimated.

c. Serious inconsistency (-1). Confidence intervals do not overlap for many studies, heterogeneity statistics indicate inconsistent results, and the prediction interval suggests that the true effect in a new study could range from important benefit to important harm. This might suggest very serious inconsistency, however the point estimate in the majority of studies indicates important benefit or trivial effects, not important harm. For this reason, we have downgraded for serious inconsistency.

d. Publication bias strong suspected (-1). Evidence from sensitivity analysis and contour enhanced funnel plot that there could be missing studies which show effects favouring the control, and nonsignificant effects in general (see Appendix D). Applies overall and to population groups with a high proportion of small studies showing large, statistically significant effects favouring aromatherapy. Publication bias is not suspected for population groups for which the combined estimate indicates a trivial (i.e. unimportant) effect.

e. Serious risk of bias (-1). All studies in the analysis are at high risk of bias or some concerns, such that the observed benefit may be overestimated.

f. Serious inconsistency (-1). Confidence intervals do not overlap for many studies, heterogeneity statistics indicate inconsistent results, and the prediction interval indicates that the effect in a new study could range from important benefit to important harm. However, the point estimate for most studies (16/20; majority of weight in analysis) indicates important benefit (SMD <-0.2) or trivial effect (SMD -0.2 to 0.2; 3/20) not harm. For this reason, we have downgraded for serious not very serious inconsistency.

g. No serious imprecision. Both the upper and lower limits of the 95% confidence interval (SMD -1.75 to -0.58) are compatible with an important reduction in pain (SMD < -0.2).

h. Serious inconsistency (-1). Confidence intervals do not overlap for many studies, heterogeneity statistics indicate inconsistent results, and the prediction interval indicates that the effect in a new study could range from important benefit to important harm. However, the point estimate for most studies (24/29; majority of weight in analysis) indicates important benefit (SMD <-0.2) or trivial effect (SMD -0.2 to 0.2; 5/29) not harm. For this reason, we have downgraded for serious not very serious inconsistency.

i. No serious inconsistency. Heterogeneity statistics indicate inconsistent results; however, the confidence intervals overlap for all but one study, and all point estimates indicate important benefit (SMD of -0.2 or lower). For this reason, we have not downgraded for inconsistency.

j. Very serious risk of bias (-2). All studies in analysis are at high risk of bias.

k. Inconsistency not assessed: single study

l. Serious indirectness (-1): Evidence from one small study among people receiving acute care for fracture. Uncertain whether results apply to other populations with acute pain.

m. Serious indirectness (-1). Evidence from one small study among people receiving care for migraine pain. Uncertain whether results apply to other populations with headache or migraine.

n. Serious imprecision (-1). Both the upper and lower limits of the 95% confidence interval are compatible with an important reduction in pain (SMD < -0.2) indicating no important imprecision; however, the result is downgraded due to concerns that the model may not yield a valid estimate for this data. The data were dichotomised (ordinal data: mild pain=event; moderate/severe=no event) to enable inclusion in the analysis, and the control group experienced zero events (all had moderate/severe pain).

o. Serious risk of bias (-1). Both studies in analysis are at high risk of bias; however, there is little or no difference between treatments so downgraded for serious not very serious risk of bias.

p. Serious indirectness (-1): Evidence from two small studies among people with cancer. Uncertain whether results are generalisable to other people with cancer

q. Serious imprecision (-1). The 95% confidence interval crosses the threshold for small but important benefit (SMD of -0.2), which means the result is compatible with important benefit (SMD 0.43 lower) and little or no difference (SMD 0.14 higher).

r. Serious inconsistency (-1). Heterogeneity statistics indicate inconsistent results and the confidence intervals do not overlap for many studies. The point estimate for most studies indicates important benefit (SMD of -0.2 or lower; 8 of 9 studies, majority of weight in analysis) or a trivial effect (SMD -0.2 to 0.2; 1 of 9 studies), not harm. This could suggest unimportant inconsistency; however, the estimates from four studies are implausibly large and, for this reason, we have downgraded for inconsistency.

s. Serious inconsistency (-1). Confidence intervals do not overlap and effect estimates vary widely.

t. Extremely serious imprecision (-3). The 95% confidence interval crosses the threshold for both small but important benefit (SMD -0.2) and small but important harm (SMD 0.2), and is too wide for the result to be interpretable (SMD -11.07 to 3.63).

u. No serious imprecision. Both the upper and lower limits of the 95% confidence interval (SMD -1.53 to -0.48) are compatible with an important reduction in pain (SMD < -0.2).

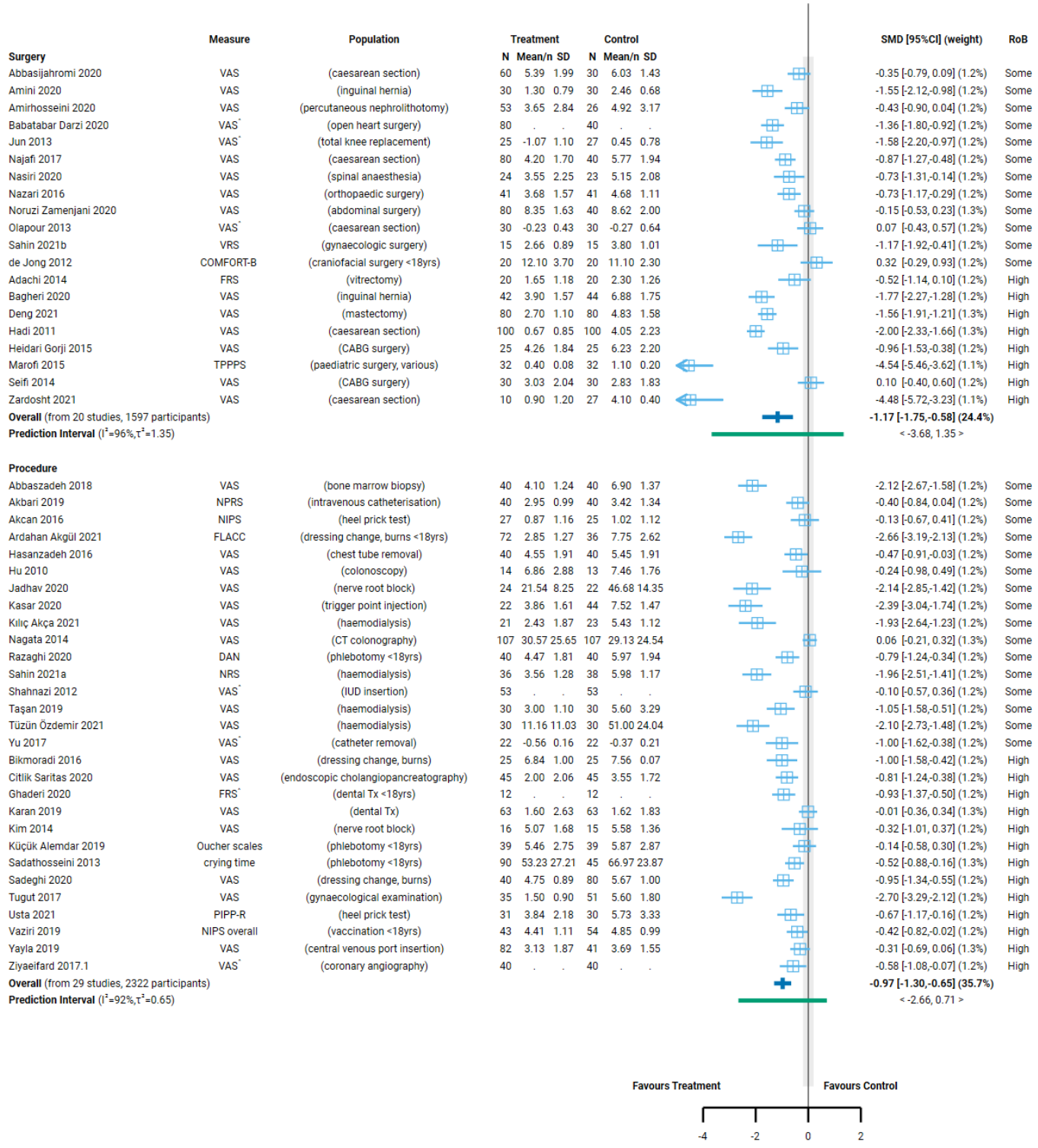


**Fig 4.2.1** | Forest plot for comparison 1. The effect of aromatherapy (any mode) versus inactive control (usual care, no intervention, placebo) on pain. See next page for continuation of plot and figure caption.

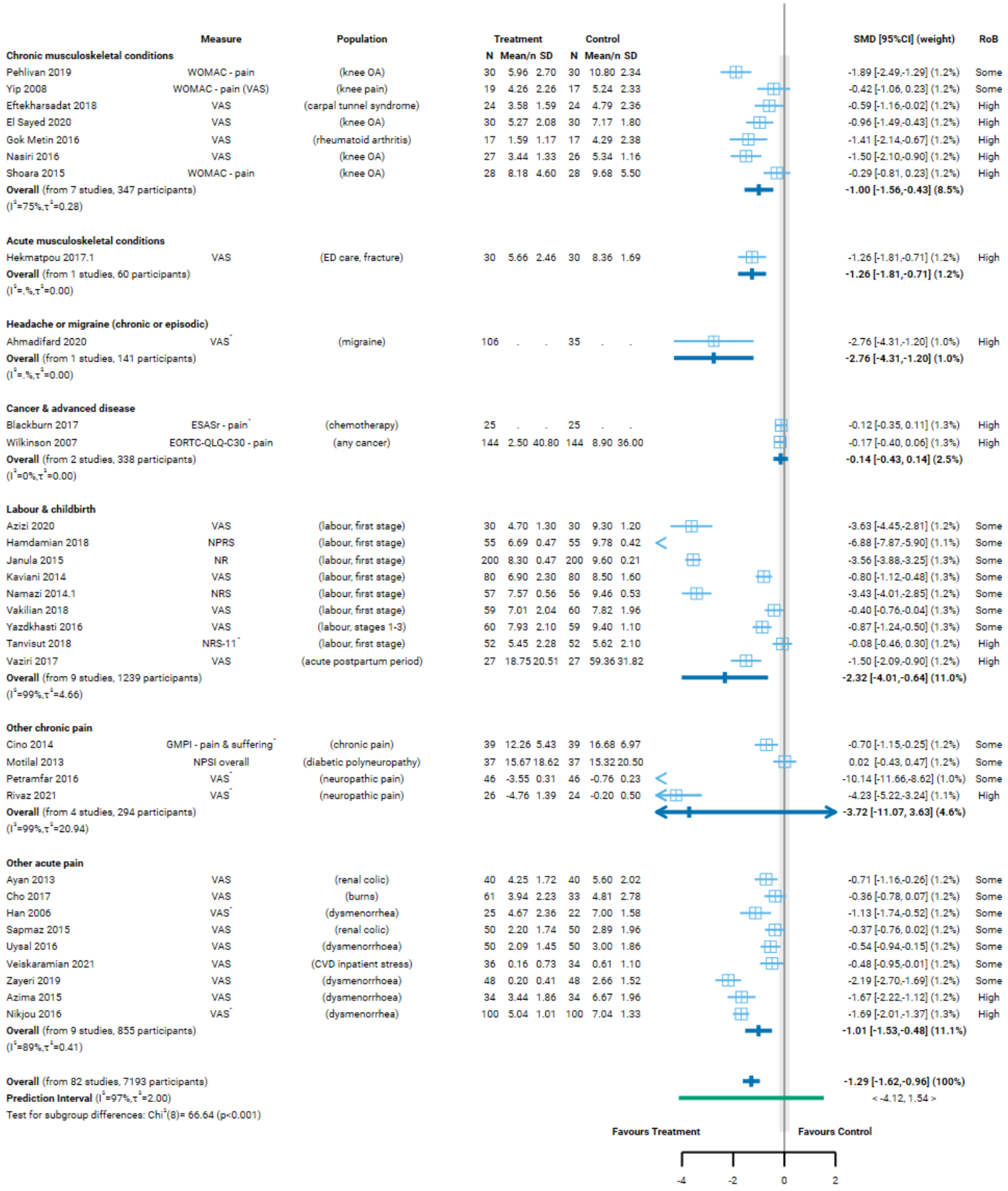


**Fig 4.2.1** | Forest plot for comparison 1. The effect of aromatherapy (any mode) versus inactive control (usual care, no intervention, placebo) on pain. SMD=standardised mean difference. Blue lines show 95% confidence intervals (CI) and green lines show prediction intervals (PI). The shaded grey area indicates the pre-specified range where the effect of aromatherapy is considered to be no different from control (SMD -0.2 to 0.2 standard units). ^ indicates studies for which data transformation or imputation was required to include the result in the meta-analysis. This may include crossover trials and studies that reported results as a dichotomous or ordinal outcome (identifiable because no mean or SD is reported for the study in the forest plot). * Denotes studies for which the direction of effect was changed to match the overall plot (e.g. positive numbers are beneficial).

**Comparison 2. Aromatherapy (massage) versus massage**

The effect of aromatherapy massage on pain is uncertain overall (all population groups) and for each population group (surgery, procedures, chronic musculoskeletal pain, labour and childbirth, other chronic pain, other acute pain).

Factors that reduced our certainty in the combined estimates of effect differed somewhat for each population group, as explained in the GRADE summary of findings table (Table 4.2.2, explanations). In combination, these factors raise concern that any observed benefit could be overestimated (or harm underestimated). Major concerns are as follows.

- ***Publication bias***. There is evidence that there could be studies (or results) missing from the analysis that show effects favouring the control (i.e. selective non-reporting based on the direction and statistical significance of results).
- ***Risk of bias in included trials***. All trials in the analysis have methodological limitations (high risk of bias or some concerns). The absence of trials at low risk of bias meant that it was not possible to examine the impact of methodological limitations on the estimate of the intervention effect using the approach specified in our protocol (limiting analyses to studies at low risk of bias).
- ***Inconsistent results that lead to different conclusions about the effects of aromatherapy***. There is evidence that the size of the intervention effect differs across studies beyond what would be expected by chance. These differences were not explained by population group (Figure 4.2.2; as explained in Appendix D, Section 4.1). This reduced our confidence in the combined estimate because some studies found an important reduction in pain (greater than the threshold for important benefit, an SMD of – 0.2 or lower) while others found little or no difference between aromatherapy and control, with no credible evidence to explain whether this reflects true differences in the effects of aromatherapy or methodological problems in some studies.

Concerns relating to each finding were considered in the GRADE assessment when interpreting the result. The findings are as follows.

**Results for which an interpretation was made** (low certainty evidence)

- ***None***

**Results considered too uncertain to interpret**

- ***Overall (all population groups).*** The evidence is very uncertain about the effect of aromatherapy (massage) on pain compared to massage alone (19 studies, 1058 participants; very low certainty, Figure 4.2.2).
- ***Chronic musculoskeletal pain***. The evidence is very uncertain about the effect of aromatherapy (massage) on chronic musculoskeletal pain compared to massage alone (5 studies, 278 participants; very low certainty, Figure 4.2.2).
- ***Surgery***. The evidence is very uncertain about the effect of aromatherapy (massage) on pain after surgery (acute postoperative period) compared to massage alone (3 studies, 110 participants; very low certainty, Figure 4.2.2).
- ***Procedures***. The evidence is very uncertain about the effect of aromatherapy (massage) on pain during or after a procedure (acute procedural period) compared to massage alone (2 studies, 101 participants; very low certainty).
- ***Labour and childbirth***. The evidence is very uncertain about the effect of aromatherapy (massage) on pain during labour and childbirth compared to massage alone (1 study, 60 participants; very low certainty).
- ***Other chronic pain***. The evidence is very uncertain about the effect of aromatherapy (massage) on other chronic pain compared to massage alone (3 studies, 195 participants; very low certainty).
- ***Other acute pain***. The evidence is very uncertain about the effect of aromatherapy (massage) on other acute pain compared to massage alone (5 studies, 314 participants; very low certainty, Figure 4.2.2).

Eligible populations for this analysis for which no trials were identified were people living with cancer and advanced disease, headache and migraine, and acute musculoskeletal pain.

**Table 4.2.2**. Summary of findings for Comparison 2. the effect of aromatherapy (massage) versus inactive massage control on pain.

| Outcomes | Anticipated absolute effects* (95% CI) | | Relative effect (95% CI) | № of participants (studies) | Certainty of the evidence (GRADE) | Comments |
|---|---|---|---|---|---|---|
| | **Risk with inactive massage control** | **Risk with aromatherapy (massage)** | | | | |
| Pain: All population groups[a] | - | SMD **0.72 SD lower** (1.19 lower to 0.25 lower) | - | 1058 (19 RCTs) | ⨁◯◯◯ Very low[b,c,d,e] | The evidence is very uncertain about the effect of aromatherapy (massage) on pain compared to massage alone |
| Pain: chronic musculoskeletal conditions (knee OA, knee pain, neck pain) | - | SMD **0.52 SD lower** (1.30 lower to 0.27 higher) | - | 278 (5 RCTs) | ⨁◯◯◯ Very low[f,g,h] | The evidence is very uncertain about the effect of aromatherapy (massage) on chronic musculoskeletal pain compared to massage alone. |
| Pain after surgery (acute postoperative period)[a] | - | SMD **0.12 SD higher** (0.77 lower to 1.01 higher) | - | 110 (3 RCTs) | ⨁◯◯◯ Very low[f,i] | The evidence is very uncertain about the effect of aromatherapy (massage) on pain after surgery (acute postoperative period) compared to massage alone. |
| Pain during or after a procedure (haemodialysis; periprocedural period)[a] | - | SMD **0.47 SD lower** (5.21 lower to 4.27 higher) | - | 101 (2 RCTs) | ⨁◯◯◯ Very low[f,j,k] | The evidence is very uncertain about the effect of aromatherapy (massage) on pain during or after a procedure (acute procedural period) compared to massage alone. |
| Pain during labour and childbirth | - | SMD **2.67 SD lower** (3.36 lower to 1.98 lower) | - | 60 (1 RCT) | ⨁◯◯◯ Very low[f,l,m,n] | The evidence is very uncertain about the effect of aromatherapy (massage) on pain during labour and childbirth compared to massage alone. |
| Pain: other chronic (neuropathic pain, prostatitis) | - | SMD **1.22 SD lower** (6.29 lower to 3.84 higher) | - | 195 (3 RCTs) | ⨁◯◯◯ Very low[f,o,p] | The evidence is very uncertain about the effect of aromatherapy (massage) on other chronic pain compared to massage alone. |
| Pain: other acute (dysmenorrhea) | - | SMD **0.90 SD lower** (1.41 lower to 0.39 lower) | - | 314 (5 RCTs) | ⨁◯◯◯ Very low[q,e] | The evidence is very uncertain about the effect of aromatherapy (massage) on other acute pain compared to massage alone. |

***The risk in the intervention group** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI). **CI:** confidence interval; **SMD:** standardised mean difference.

The threshold for an important difference was an SMD of 0.2 (used for interpreting point estimates and confidence intervals). For pain, the resulting interpretation is: < -0.2 is beneficial, -0.2 to 0.2 is trivial or unimportant ("little or no difference" between treatments), > 0.2 is harmful

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

## Explanations

a. Measures varied. VAS, VRS, NRS and some condition- or population-specific measures.

b. Serious risk of bias (-1). All 19 studies in the analysis are at high risk of bias or some concerns, such that the observed benefit may be overestimated.

c. Serious inconsistency (-1). Confidence intervals do not overlap for many studies, heterogeneity statistics indicate inconsistency, and the prediction interval indicates that the effect in a new study could range from important benefit to important harm. However, the point estimate for most studies (17/19; majority of weight in analysis) indicates important benefit (SMD <-0.2) or trivial effect (SMD -0.2 to 0.2) not harm. For this reason, we have downgraded for serious not very serious inconsistency.

d. No serious imprecision. Both the upper and lower limits of the 95% confidence interval (SMD -1.19 to -0.25) are compatible with an important reduction in pain (SMD < -0.2).

e. Publication bias strong suspected (-1). Evidence from contour enhanced funnel plot that there could be missing studies which show effects favouring the control, especially nonsignificant effects (see Appendix D). Applies overall and to population groups with a high proportion of small studies showing large, statistically significant effects favouring aromatherapy (combined effect estimate is moderate to large). Publication bias is not suspected for population groups for which the combined effect estimate is trivial (i.e. an unimportant effect).

f. Serious risk of bias (-1). All studies in the analysis are at high risk of bias or some concerns, such that the observed benefit may be overestimated.

g. Serious inconsistency (-1). Heterogeneity statistics indicate inconsistency. However, the confidence intervals overlap for most studies and the point estimate for all studies indicates important benefit (SMD <-0.2; 3/5 studies) or trivial effect (SMD -0.2 to 0.2; 2/5 studies) not harm. For this reason, we have downgraded for serious not very serious inconsistency.

h. Serious imprecision (-1). The 95% confidence interval crosses the threshold for both a small but important reduction in pain (SMD -0.2) and a small but important increase in pain (SMD 0.2), so the result is compatible with important benefit (SMD 1.30 lower) and important harm (SMD 0.27 higher). However, this is likely influenced by inconsistent results, so we rated imprecision as serious not very serious.

i. Very serious imprecision (-2). The 95% confidence interval crosses the threshold for both a small but important reduction in pain (SMD -0.2) and a small but important increase in pain (SMD 0.2)] so the result is compatible with important benefit (SMD -0.77 lower) and important harm (SMD 1.01 higher).

j. Serious indirectness (-1): Evidence from two small studies among people receiving haemodialysis. Uncertain whether results apply to prevention or relief of procedural pain more generally.

k. Extremely serious imprecision (-3). The 95% confidence interval crosses the threshold for both small but important benefit (SMD -0.2) and small but important harm (SMD 0.2), and is too wide for the result to be interpretable (SMD -5.21 to 4.27).

l. Inconsistency not assessed: single study

m. Serious indirectness (-1): Evidence from one small study among people receiving pain relief during labour and childbirth. Uncertain whether results apply to other populations during labour and childbirth.

n. Publication bias strongly suspected (-1). Evidence from sensitivity analysis and contour enhanced funnel plot that there could be missing studies which show effects favouring the control, and nonsignificant effects in general (see Appendix D). Most concerning for labour and childbirth (single small study, large effect).

o. Serious inconsistency (-1). Heterogeneity statistics indicate inconsistency. However, the confidence intervals overlap for most studies and the point estimate for all studies indicates important benefit (SMD <-0.2; 1/3 studies) or trivial effect (SMD -0.2 to 0.2; 2/3 studies) not harm. For this reason, we have downgraded for serious not very serious inconsistency.

p. Extremely serious imprecision (-3). The 95% confidence interval crosses the threshold for both small but important benefit (SMD -0.2) and small but important harm (SMD 0.2), and is too wide for the result to be interpretable (SMD -6.29 to 3.84).

q. Very serious risk of bias (-2). All studies in the analysis are at high risk of bias, such that the observed benefit may be overestimated.

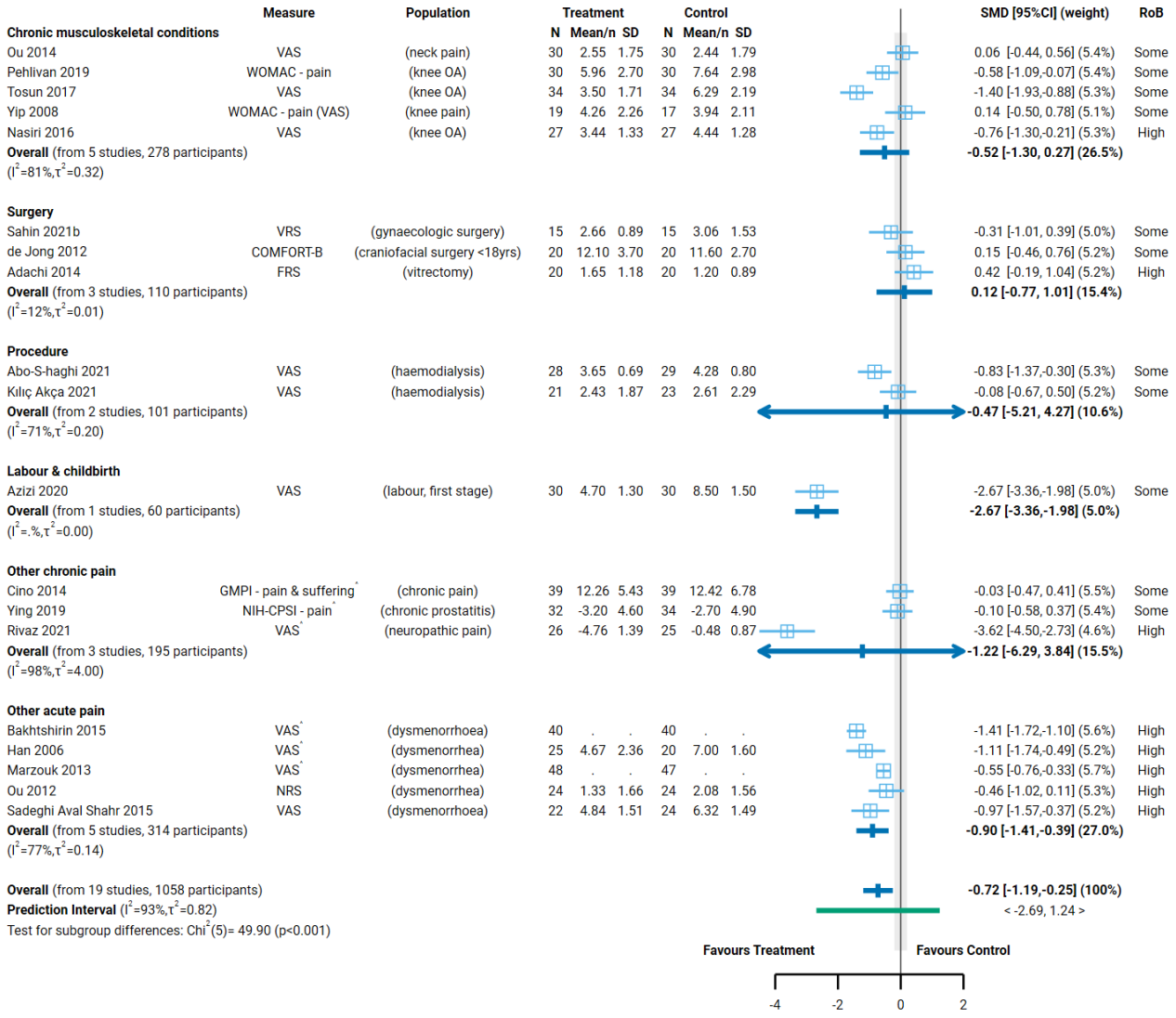


**Fig 4.2.2** | Forest plot for comparison 2. the effect of aromatherapy (massage) versus inactive massage control on pain. SMD=standardised mean difference. Blue lines show 95% confidence intervals (CI) and green lines show prediction intervals (PI). The shaded grey area indicates the pre-specified range where the effect of aromatherapy is considered to be no different from control (SMD -0.2 to 0.2 standard units). ^ indicates studies for which data transformation or imputation was required to include the result in the meta-analysis. This may include crossover trials and studies that reported results as a dichotomous or ordinal outcome (identifiable because no mean or SD is reported for the study in the forest plot). * Denotes studies for which the direction of effect was changed to match the overall plot (e.g. positive numbers are beneficial).

## 4.3 Nausea and vomiting

Twenty-three studies (2032 participants) examined the effect of aromatherapy (any mode) compared to an inactive control on nausea, vomiting or both (Comparison 1, Figure 4.3.1). An additional four trials (428 participants) were eligible for this comparison, but either did not report results despite having measured the outcome (1 study, 120 people undergoing a dental procedure), or the results were uninterpretable (2 studies, 271 people undergoing surgery; 1 study, 37 children undergoing stem cell transplantation).

No studies examined the effect of aromatherapy (massage) compared to massage alone on nausea and vomiting (Comparison 2).

### Characteristics of included studies

#### Types of populations

Included studies examined the effect of aromatherapy on nausea, vomiting or both among

- people living with cancer or advanced disease (to relieve treatment related side-effects, primarily chemotherapy, in all but one trial),
- post-operatively after any type of surgery or a specific surgery (e.g. caesarean, nephrectomy),
- post-procedurally (stem cell transplantation), and
- during pregnancy (Figure 4.3.1, column 3).

The specific condition addressed in each of the trials is reported in the forest plot (column 3, Figure 4.3.1) with full details for each study including eligibility criteria, participant characteristics, and ICD 11 codes in Appendix E1.

#### Types of interventions

Aromatherapy was delivered by inhalation in all but two trials (both among people living with cancer or advanced disease). Peppermint was the most commonly used essential oil (alone or as part of a blend), followed by ginger and lavender.

- For people living with cancer or advanced disease, the aromatherapy intervention period was variable, ranging from short term (less than one day) to treatment over months (multiple chemotherapy cycles).
- For people undergoing surgery or procedures, aromatherapy was generally administered soon after surgery, with one or multiple doses delivered within 24 hours.
- For pregnancy, the intervention period ranged from four to seven days.

#### Types of outcomes

The outcome measure from which data were included for meta-analysis is reported for each trial in the forest plot (column 2, Figure 4.3.1). Full details for each study are in Appendix E1, including the timing of outcome measurement in relation to intervention and details of which outcome was selected when multiple outcomes were available.

Studies varied in what was measured (nausea, vomiting or both), how the outcome was measured (on a scale, an event, a count per person) and how the data were handled and reported (Figure 4.3.1, column 2). Variations were as follows.

In some studies, outcomes were measured on a scale and reported as:

- a score on the scale (e.g., severity score on VAS),
- a dichotomous outcome for which a threshold was used to categorise participants as having the event or not (e.g., any nausea or no nausea), or
- an ordinal outcome for which multiple thresholds were used to create ordinal categories (e.g., none, mild, moderate or severe).

The scales used varied (Figure 4.3.1, column 2), as did the thresholds for dichotomising results or creating ordinal categories.

In other studies, outcomes were measured as:

- an event (e.g. 'any vomiting' reported as the number of participants with at least one episode of vomiting), or
- a count (e.g. number of episodes per participant).

Where possible, we selected a result reported on the original scale. When necessary, an effect estimate was calculated from dichotomous and ordinal data (odds ratio) and transformed to a standardised mean difference (Appendix B).

### Effects of aromatherapy on nausea and vomiting

**Comparison 1: Aromatherapy (any mode) versus inactive control (usual care, placebo, no intervention)**

The effect of aromatherapy on nausea and vomiting is uncertain overall (all population groups). Aromatherapy may improve nausea and vomiting for pregnancy, but is uncertain and for each other population group (cancer and advanced disease, surgery, procedures).

Factors that reduced our certainty in the combined estimates of effect differed somewhat for each population group, as explained in the GRADE summary of findings table (Table 4.3.1, explanations). In combination, these factors raise concern that any observed benefit could be overestimated (or harm underestimated). Major concerns are as follows.

- ***Publication bias***. There is evidence that there could be studies (or results) missing from the analysis that show effects favouring the control (i.e. selective non-reporting based on the direction and statistical significance of results).
- ***Risk of bias in included trials***. All trials in the analysis have methodological limitations (high risk of bias or some concerns). The absence of trials at low risk of bias meant that it was not possible to examine the impact of these methodological limitations on the estimate of the intervention effect using the approach specified in our protocol (limiting analyses to studies at low risk of bias).
- ***Inconsistent results that lead to different conclusions about the effects of aromatherapy***. There is evidence that the size of the intervention effect differs across studies beyond what would be expected by chance. These differences were not explained by population group (Figure 4.3.1; as explained in Appendix D, Section D.2) or the mode by which aromatherapy was delivered (Appendix D, Section D.2 and Figure D.2.1). This reduced our confidence in the combined estimate because some studies found an important reduction in nausea and vomiting (greater than the threshold for important benefit, an SMD of - 0.2 or lower) while others found little or no difference between aromatherapy and control, with no credible evidence to explain whether this reflects true differences in the effects of aromatherapy or methodological problems in some studies.

Concerns relating to each finding were considered in the GRADE assessment when interpreting the result. The findings are as follows.

**Results for which an interpretation was made** (low certainty)

- ***Pregnancy***. There is low certainty evidence that aromatherapy (any mode) may reduce nausea and vomiting during pregnancy (SMD 0.52 lower, 95% CI 1.08 lower to 0.04 higher; $I^2$ = 49%; 4 studies, 271 participants; low certainty, Figure 4.3.1).

**Results considered too uncertain to interpret**

- ***Overall (all population groups).*** The evidence is very uncertain about the effect of aromatherapy (any mode) on nausea and vomiting (23 studies, 2032 participants; very low certainty, Figure 4.3.1).
- ***Cancer and advanced disease***. The evidence is very uncertain about the effect of aromatherapy (any mode) on nausea and vomiting among people undergoing chemotherapy for cancer (8 trials, 738 participants; very low certainty).
- ***Surgery***. The evidence is very uncertain about the effect of aromatherapy (any mode) on post-operative nausea and vomiting (10 studies, 982 participants; very low certainty, Figure 4.3.1).
- ***Procedures.*** The evidence is very uncertain about the effect of aromatherapy (any mode) on nausea and vomiting among people undergoing procedures (1 trial, 41 participants; very low certainty).

**Table 4.3.1**. Summary of findings for Comparison 1. the effect of aromatherapy (any mode) versus inactive control (usual care, no intervention, placebo) on nausea and vomiting.

| Outcomes | Anticipated absolute effects* (95% CI) | | Relative effect (95% CI) | № of participants (studies) | Certainty of the evidence (GRADE) | Comments |
|---|---|---|---|---|---|---|
| | **Risk with inactive control (usual care, placebo, no intervention)** | **Risk with aromatherapy (any mode)** | | | | |
| Nausea, vomiting or both: All population groups[a] | - | SMD **0.51 SD lower** (0.85 lower to 0.17 lower) | - | 2032 (23 RCTs) | ⨁◯◯◯ Very low[b,c,d,e] | The evidence is very uncertain about the effect of aromatherapy (any mode) on nausea and vomiting. |
| Nausea, vomiting or both in cancer & advanced disease[a] | - | SMD **0.35 SD lower** (0.99 lower to 0.29 higher) | - | 738 (8 RCTs) | ⨁◯◯◯ Very low[f,g,h] | The evidence is very uncertain about the effect of aromatherapy (any mode) on nausea and vomiting among people undergoing chemotherapy for cancer. |
| Nausea, vomiting or both after surgery[a] | - | SMD **0.67 SD lower** (1.37 lower to 0.03 higher) | - | 982 (10 RCTs) | ⨁◯◯◯ Very low[b,i,j] | The evidence is very uncertain about the effect of aromatherapy (any mode) on post-operative nausea and vomiting. |
| Nausea, vomiting or both from procedures[a] | - | SMD **0.09 SD lower** (0.69 lower to 0.52 higher) | - | 41 (1 RCT) | ⨁◯◯◯ Very low[k,l,m] | The evidence is very uncertain about the effect of aromatherapy (any mode) on nausea and vomiting among people undergoing procedures. |
| Nausea, vomiting or both during pregnancy | - | SMD **0.52 SD lower** (1.08 lower to 0.04 higher) | - | 271 (4 RCTs) | ⨁⨁◯◯ Low[b,n,o] | Aromatherapy (any mode) may reduce nausea and vomiting during pregnancy. |

***The risk in the intervention group** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI). **CI:** confidence interval; **SMD:** standardised mean difference
The threshold for an important difference was an SMD of 0.2 (used for interpreting point estimates and confidence intervals). For nausea and vomiting, the resulting interpretation is: < -0.2 is beneficial, -0.2 to 0.2 is trivial or unimportant ("little or no difference" between treatments), > 0.2 is harmful

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

## Explanations

a. Measures varied. Studies measured severity (different scales including VAS, NRS), number of episodes, and proportion of participants with no improvement.

b. Serious risk of bias (-1). All studies in the analysis are at high risk of bias or some concerns, such that the observed benefit may be overestimated.

c. Very serious inconsistency (-2). Confidence intervals do not overlap for many studies, heterogeneity statistics indicate inconsistency, point estimates vary widely, and the prediction interval suggests that the true effect in a new study could range from important benefit to important harm.

d. No serious imprecision. The 95% confidence interval crosses the threshold for a small but important reduction in nausea and vomiting (SMD of -0.2), so the result is compatible with important benefit (SMD 0.85 lower) and little or no difference (SMD 0.17 lower). However, the extent to which the threshold is crossed is modest (likely due to inconsistent effects) and both the upper and lower limit of the confidence interval favours the intervention, so we have not rated down for imprecision.

e. Publication bias strongly suspected (-1). Evidence from contour enhanced funnel plot that there could be missing studies which show effects favouring the control, especially nonsignificant effects (see Appendix D). Applies overall and to population groups with a high proportion of small studies showing large, statistically significant effects favouring aromatherapy (combined effect estimate is moderate to large). Publication bias is not suspected for population groups for which the combined effect estimate is trivial (i.e. an unimportant effect).

f. Very serious risk of bias (-2). Almost all studies in analysis (7/8; majority of weight) are at high risk of bias such that the effect may be overestimated.

g. Serious inconsistency (-1). Heterogeneity statistics indicate inconsistency. Confidence intervals overlap for a majority of studies, but point estimates vary widely (compatible with little or no difference in 4 of 8, important benefit in 3 of 8, important harm in 1 of 8).

h. Serious imprecision (-1). The 95% confidence interval crosses the threshold for both a small but important reduction in nausea and vomiting (SMD -0.2) and a small but important increase (SMD 0.2), so the result is compatible with important benefit (SMD 0.99 lower) and important harm (SMD 0.29 higher). However, we have downgraded by -1 because the imprecision is likely influenced by inconsistent results (rated as serious).

i. Serious inconsistency (-1). Heterogeneity statistics indicate inconsistency. Confidence intervals overlap for a majority of studies, but point estimates vary widely (compatible with little or no difference in 5 of 10, important benefit in 5 of 10).

j. Serious imprecision (-1). The 95% confidence interval crosses the threshold for a small but important benefit (SMD of -0.2), which means the result is compatible with important benefit (SMD 1.37 lower) and little or no difference (SMD 0.03 higher).

k. No serious risk of bias. Single study in analysis with some concerns about risk of bias; however there is little or no difference between treatments so not downgraded for risk of bias.

l. Serious indirectness (-1). Evidence from one small study among people receiving aromatherapy to prevent nausea and vomiting during a procedure. Uncertain whether results apply to other populations undergoing procedures.

m. Very serious imprecision (-2). The 95% confidence interval crosses two thresholds for a small by important effect (SMD of 0.2 and -0.2), so the result is compatible with important benefit (SMD -0.69 lower) and important harm (SMD 0.52 higher).

n. No serious inconsistency. Confidence intervals overlap for all studies, suggesting that any variation in results across studies may be unimportant.

o. Serious imprecision (1). The 95% confidence interval crosses the threshold for a small but important reduction in nausea and vomiting (SMD of -0.2), so the result is compatible with important benefit (SMD 1.08 lower) and unimportant harm (SMD 0.04 higher).

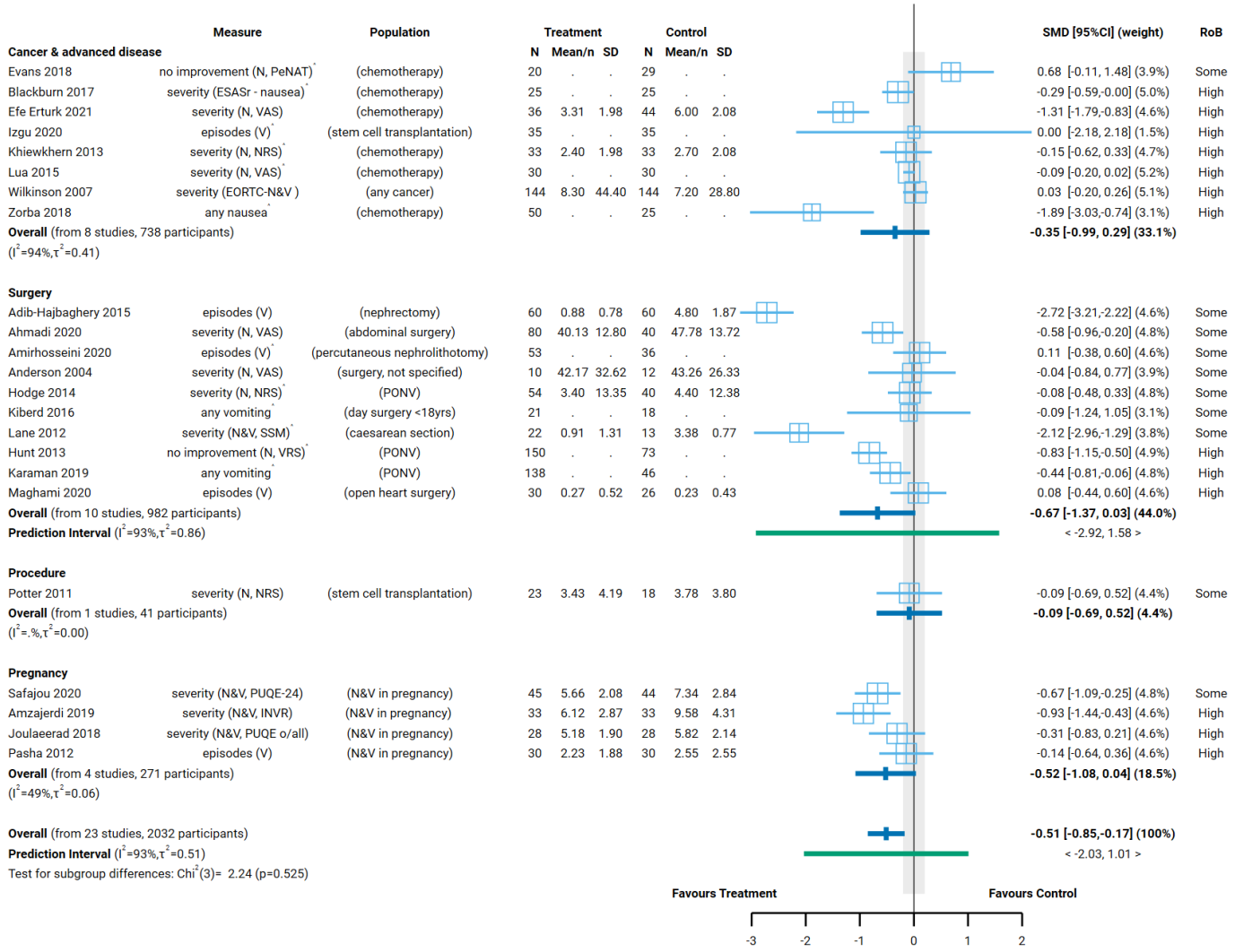


**Fig 4.3.1** | Forest plot for comparison 1. the effect of aromatherapy (any mode) versus inactive control (usual care, no intervention, placebo) on nausea and vomiting. For measures, in bracketed text: N=nausea. V=vomiting. N&V=nausea and vomiting. SMD=standardised mean difference. Blue lines show 95% confidence intervals (CI) and green lines show prediction intervals (PI). The shaded grey area indicates the pre-specified range where the effect of aromatherapy is considered to be no different from control (SMD -0.2 to 0.2 standard units). ^ indicates studies for which data transformation or imputation was required to include the result in the meta-analysis. This may include crossover trials and studies that reported results as a dichotomous or ordinal outcome (identifiable because no mean or SD is reported for the study in the forest plot). *Denotes studies for which the direction of effect was changed to match the overall plot (e.g. positive numbers are beneficial).

## 4.4 Sleep

To be eligible for the sleep analysis, there had to be evidence that participants had insomnia or signs/symptoms of sleep disruption (i.e. either this was part of the trial eligibility criteria or the baseline data indicated sleep disruption; minimal criteria such as self-report of 'disturbed sleep' were accepted).

Twenty-two studies (1397 participants) examined the effect of aromatherapy (any mode) compared to an inactive control on sleep (Comparison 1, Figure 4.4.1). An additional four trials (373 participants) were eligible for this comparison, but either did not report data that could be used in the meta-analysis (1 study each among people with cancer, sleep disruption and in hospital), or the results had errors that could not be reconciled (1 study among people with cancer).

No studies examined the effect of aromatherapy (massage) compared to massage alone on sleep (Comparison 2), except for one of the four trials with missing data (1 trial, 150 participants living with cancer).

### Characteristics of included studies

#### Types of populations

Included studies examined the effect of aromatherapy on sleep quality among

- people living with cancer or advanced disease (during chemotherapy treatment, in palliative care, any cancer),
- post-operatively after any type of surgery or a specific surgery (CABG, colorectal surgery),
- among people who are hospitalised (mainly for cardiovascular disease or events),
- people with chronic insomnia (as a primary diagnosis, in menopause and as a comorbidity of diabetes, each in one trial),
- people with sleep disturbance (mainly those undergoing haemodialysis, postpartum and comorbidity of depression each in one trial).

No studies among people with dementia were included in this analysis.

The specific condition addressed in each of the trials is reported in the forest plot (column 3, Figure 4.4.1) with full details for each study including eligibility criteria, participant characteristics, and ICD 11 codes in Appendix E1.

#### Types of interventions

Of the 22 trials included in Comparison 1, aromatherapy was delivered by inhalation in 19 trials and massage in 3 trials. One of the 22 trials examined the effects of two aromatherapy treatments that we combined prior to inclusion in the meta-analysis (two different doses of rose oil).

Lavender was the most commonly evaluated essential oil (15/22 trials), followed by rose (4 trials) and orange (2 trials). No trials evaluated a blend of essential oils, but one trial gave participants a choice of several oils.

The treatment period varied in length.

- For people with cancer, aromatherapy was administered nightly over two weeks in one trial, one week in another, and two nights in a third.
- For people undergoing surgery or in hospital, the aromatherapy treatment was given over days, although the timing of treatment varies (e.g. before surgery in one trial, after surgery in two others). The exception was a single trial among people in hospital for cardiovascular disease.
- For people with chronic insomnia, all trials involved nightly administration of aromatherapy over a 4-week period.
- For people with symptoms of sleep disruption, the aromatherapy treatment period ranged from one week to eight weeks (3/5 trials were at least one month), with differences in treatment frequency across trials.

#### Types of outcomes

The outcome measure from which data were included for meta-analysis is reported for each trial in the forest plot (column 2, Figure 4.4.1). Full details for each study are in Appendix E1, including the timing of outcome measurement in relation to intervention and details of which outcome was selected when multiple were available (e.g. when both

overall and subscale scores were available). The appendix also reports studies in which sleep was measured, but the population was ineligible for inclusion in the analysis.

All but one study measured sleep quality on a scale, the majority using the Pittsburgh Sleep Quality Index (PSQI; 12/22 trials). Other scales used were the St Mary's Hospital Sleep Questionnaire (SMHSQ; 4/22 trials), Richards-Campbell Sleep Questionnaire (RCSQ; 3/22 trials), Pittsburgh Insomnia Rating Scale (PIRS; 1/22 trials), and a visual analogue scale (VAS; 1/22 trials). Different versions of the same scale were used in some studies (e.g. both the 11 and 14 item versions of the SMHSQ were used). One study measured sleeping time.

All results were reported as a score on the original scale (e.g. sleep quality on the Pittsburgh Sleep Quality Index).

### Effects of aromatherapy on sleep

**Comparison 1: Aromatherapy (any mode) versus inactive control (usual care, placebo, no intervention)**

Aromatherapy may improve sleep quality (all population groups), and for the specific group hospitalisation. However, the effects for other specific population groups are less certain (cancer and advanced disease, chronic insomnia, surgery, sleep disturbance). No studies among people with dementia were included in this analysis.

Factors that reduced our certainty in the combined estimates of effect differed somewhat for each population group, as explained in the GRADE summary of findings table (Table 4.4.1, explanations). In combination, these factors raise concern that any observed benefit could be overestimated (or harm underestimated). Major concerns are as follows.

- ***Publication bias***. There is evidence that there could be studies (or results) missing from the analysis that show effects favouring the control (i.e. selective non-reporting based on the direction and statistical significance of results).
- ***Risk of bias in included trials***. All trials in the analysis have methodological limitations (high risk of bias or some concerns). The absence of trials at low risk of bias meant that it was not possible to examine the impact of these methodological limitations on the size (or direction) of the intervention effect using the approach specified in our protocol (limiting analyses to studies at low risk of bias).

While there is some inconsistency in the effects of aromatherapy on sleep, the results of all studies showed an improvement in sleep quality (an effect estimate greater than the threshold for important benefit, an SMD of 0.2 or higher), and the observed inconsistency is arising because some studies showed much larger benefit than others (Figure 4.4.1; as explained in Appendix D, Section 4.3).

Concerns relating to each finding were considered in the GRADE assessment when interpreting the result. The findings are as follows.

**Results for which an interpretation was made** (low certainty evidence)

- ***Overall (all population groups).*** Aromatherapy (any mode) may improve sleep quality (SMD 1.11 higher, 95% CI 0.72 higher to 1.50 higher; $I^2$ = 90 %; 22 studies, 1397 participants; low certainty, Figure 4.4.1).
- ***Hospitalisation***. Aromatherapy (any mode) may improve sleep quality during hospitalisation for cardiovascular inpatients (excluding surgery) (SMD 0.81 higher, 95% CI 0.12 higher to 1.15 higher; $I^2$ = 89 %; 8 studies, 498 participants; low certainty Figure 4.4.1).

**Results considered too uncertain to interpret**

- ***Cancer or advanced disease.*** The evidence is very uncertain the effect of aromatherapy (any mode) on sleep quality for people living with cancer (3 studies, 163 participants; very low certainty, Figure 4.4.1).
- ***Surgery***. The evidence is very uncertain about the effect of aromatherapy (any mode) on sleep quality after surgery (acute postoperative period) (3 studies, 227 participants; very low certainty Figure 4.4.1).
- ***Chronic insomnia.*** The evidence is very uncertain about the effect of aromatherapy (any mode) on sleep quality among people with chronic insomnia. (3 studies, 131 participants; very low certainty Figure 4.4.1).
- ***Sleep disruption.*** The evidence is very uncertain about the effect of aromatherapy (any mode) on sleep quality among people with signs or symptoms of sleep disruption (378 participants; very low certainty Figure 4.4.4).

**Population groups for which there were no studies.** No studies among people with dementia were included in this analysis.

**Table 4.4.1**. Summary of findings for Comparison 1. the effect of aromatherapy (any mode) versus inactive control (usual care, no intervention, placebo) on sleep quality.

| Outcomes | Anticipated absolute effects* (95% CI) | | Relative effect (95% CI) | № of participants (studies) | Certainty of the evidence (GRADE) | Comments |
|---|---|---|---|---|---|---|
| | **Risk with inactive control (usual care, placebo, no intervention)** | **Risk with aromatherapy (any mode)** | | | | |
| Sleep quality: All population groups[a] | - | SMD **1.11 SD higher** (0.72 higher to 1.50 higher) | - | 1397 (22 RCTs) | ⨁⨁◯◯ Low[b,c,d] | Aromatherapy (any mode) may improve sleep quality. |
| Sleep quality among people with living cancer or advanced disease | - | SMD **1.05 SD higher** (0.31 lower to 2.42 higher) | - | 163 (3 RCTs) | ⨁◯◯◯ Very low[b,c,d,e,f] | The evidence is very uncertain about the effect of aromatherapy (any mode) on sleep quality for people living with cancer or advanced disease. |
| Sleep quality after surgery (acute postoperative period)[a] | - | SMD **0.75 SD higher** (0.40 lower to 1.90 higher) | - | 227 (3 RCTs) | ⨁◯◯◯ Very low[b,c,g] | The evidence is very uncertain about the effect of aromatherapy (any mode) on sleep quality importantly after surgery (acute postoperative period). |
| Sleep quality during hospitalisation (mainly cardiovascular inpatients; not peri-operative)[a] | - | SMD **0.81 SD higher** (0.12 higher to 1.51 higher) | - | 498 (8 RCTs) | ⨁⨁◯◯ Low[b,c,h] | Aromatherapy (any mode) may improve sleep quality during hospitalisation for cardiovascular inpatients (excluding surgery). |
| Sleep quality among people with chronic insomnia (primary diagnosis or as a comorbidity) | - | SMD **1.14 SD higher** (0.05 higher to 2.22 higher) | - | 131 (3 RCTs) | ⨁◯◯◯ Very low[b,c,i,j,] | The evidence is very uncertain about the effect of aromatherapy (any mode) on sleep quality among people with chronic insomnia. |
| Sleep quality among people with signs or symptoms of sleep disruption (primary symptoms or as a comorbidity) | - | SMD **1.88 SD higher** (0.30 higher to 3.47 higher) | - | 378 (5 RCTs) | ⨁◯◯◯ Very low[c,d,e] | The evidence is very uncertain about the effect of aromatherapy (any mode) on sleep quality among people with signs or symptoms of sleep disruption. |
| Sleep quality among people living with dementia - not reported | - | - | - | - | - | Sleep quality was not reported as an outcome for people living with dementia in any of the studies eligible for the synthesis. |

***The risk in the intervention group** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI). **CI:** confidence interval; **SMD:** standardised mean difference

The threshold for an important difference was an SMD of 0.2 (used for interpreting point estimates and confidence intervals). For sleep quality, the resulting interpretation is: > 0.2 is beneficial, -0.2 to 0.2 is trivial or unimportant ("little or no difference" between treatments), < -0.2 is harmful

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

## Explanations

a. Measures varied. Pittsburgh Sleep Quality Inventory (PSQI), Richards-Campbell Sleep Questionnaire (RCSQ), St Mary's Hospital Sleep Questionnaire (SMHSQ), Pittsburgh Insomnia Rating Scale (PIRS-20), sleeping time (hours).

b. Serious risk of bias (-1). All studies in the analysis are at high risk of bias or some concerns, such that the observed benefit may be overestimated.

c. No serious inconsistency. Heterogeneity statistics suggest inconsistent results. However, the confidence intervals overlap for many studies (those that do not are compatible large benefit), and the point estimate for all studies indicates important benefit (SMD of 0.2 or higher). Only two studies in the overall analysis have a confidence limit that is compatible with a very slight increase in harm (SMD of -0.26, both in the hospitalisation subgroup). For this reason, we have not downgraded for inconsistency overall or for any subgroups.

d. Publication bias strongly suspected (-1). Evidence from contour enhanced funnel plot that there could be missing studies which show effects favouring the control, especially nonsignificant effects (see Appendix D). Applies overall and to population groups with a high proportion of small studies showing large, statistically significant effects favouring aromatherapy (combined effect estimate is moderate to large). Publication bias is not suspected for population groups for which the combined effect estimate is trivial (i.e. an unimportant effect).

e. Very serious risk of bias (-2). Studies with the majority of weight in the analysis are at high risk of bias, and these studies show large effects, such that the observed benefit may be overestimated.

f. Very serious imprecision (-2). The 95% confidence interval crosses the threshold for both a small but important improvement in sleep quality (SMD 0.2) and a small but important reduction in sleep quality (SMD -0.2), so the result is compatible with important benefit (SMD 2.42 higher) and important harm (SMD -0.31 lower).

g. Very serious imprecision (-2). The 95% confidence interval crosses the threshold for both a small but important improvement in sleep quality (SMD 0.2) and a small but important reduction in sleep quality (SMD -0.2), so the result is compatible with important benefit (SMD 1.90 higher) and important harm (SMD -0.40 lower).

h. Serious imprecision (-1). The 95% confidence interval crosses the threshold for small but important improvement in sleep quality (SMD of 0.2), which means the result is compatible with important benefit (SMD 1.51 higher) and little or no difference (SMD 0.12 higher).

i. Serious indirectness (-1). Evidence from three small studies among people receiving aromatherapy for chronic insomnia with very different underlying conditions. Uncertain whether results apply to populations with chronic insomnia more generally.

j. Serious imprecision (-1). The 95% confidence interval crosses the threshold for a small but important improvement in sleep quality (SMD of 0.2), which means the result is compatible with important benefit (SMD 2.22 higher) and little or no difference (SMD 0.05 higher).

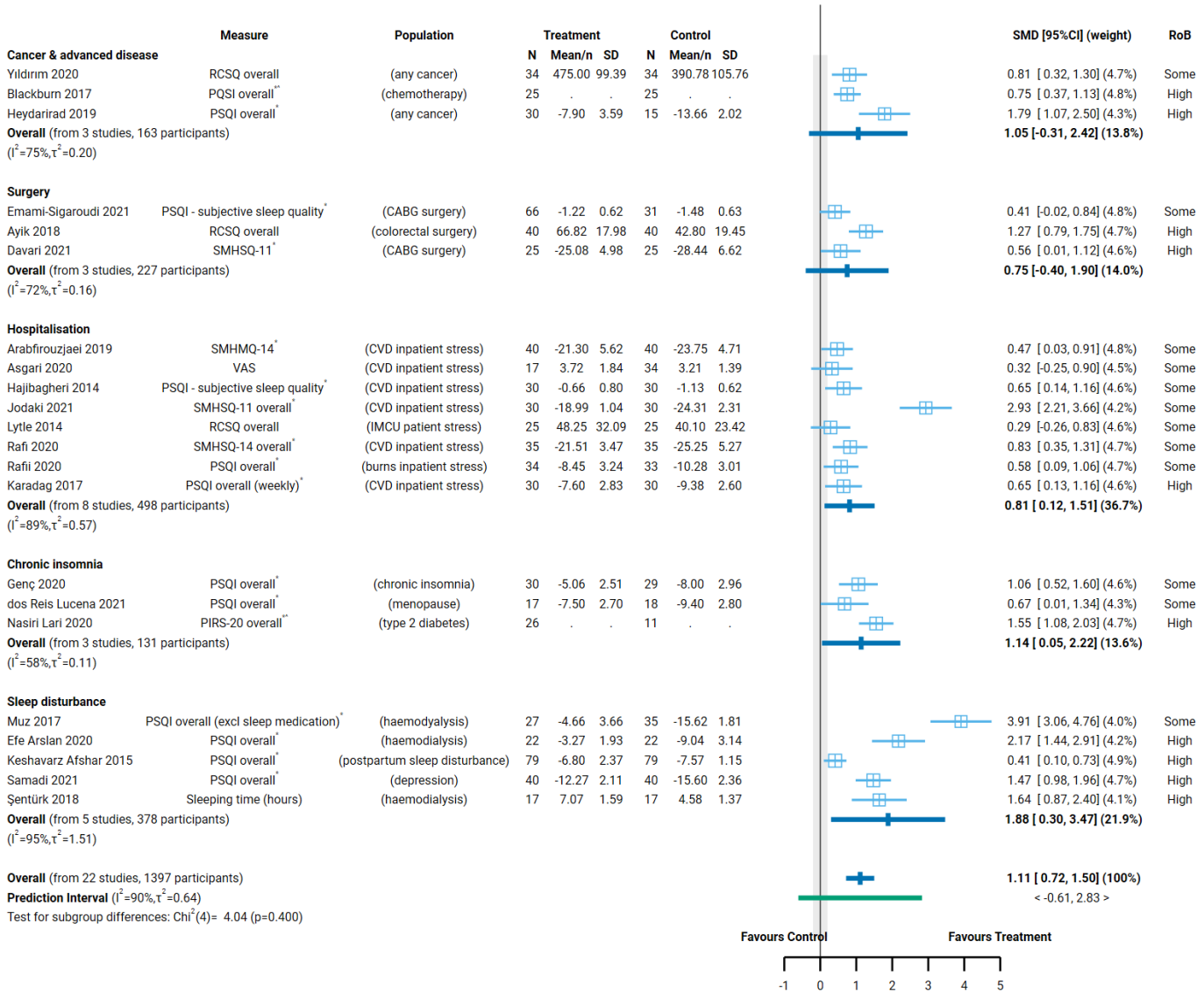


**Fig 4.4.1** | Forest plot for comparison 1. the effect of aromatherapy (any mode) versus inactive control (usual care, no intervention, placebo) on sleep quality. SMD=standardised mean difference. Blue lines show 95% confidence intervals (CI) and green lines show prediction intervals (PI). The shaded grey area indicates the pre-specified range where the effect of aromatherapy is considered to be no different from control (SMD -0.2 to 0.2 standard units). ^ indicates studies for which data transformation or imputation was required to include the result in the meta-analysis. This may include crossover trials and studies that reported results as a dichotomous or ordinal outcome (identifiable because no mean or SD is reported for the study in the forest plot). * Denotes studies for which the direction of effect was changed to match the overall plot (e.g. positive numbers are beneficial).

## 4.5 Fatigue

Overall, nineteen studies that examined the effect of aromatherapy on fatigue were included for meta-analysis, three of these contribute to both Comparison 1 and 2. A further 3 studies (449 participants) were eligible for Comparison 1, but could not be included (see below).

To be considered for the fatigue analysis, trials generally had to administer aromatherapy to an eligible population for longer-term care (i.e. delivering treatment over weeks or longer, not days) and measure fatigue in a time-frame likely to detect meaningful improvement (i.e. not immediately after a single treatment). This led to the exclusion of three trials from the analysis. Two of these studies measured fatigue immediately after either a single aromatherapy treatment or a week of treatment (1 trial), and the third trial measured fatigue the day after a single treatment.

***Comparison 1***.

- Eighteen studies (1316 participants) contribute to the comparison of aromatherapy delivered by any mode compared to an inactive control (Figure 4.5.1).

***Comparison 2***.

- Four studies (252 participants) contribute to the comparison of aromatherapy delivered by massage to an inactive massage control (comparable to massage in the aromatherapy arm) (Figure 4.5.2).

### Characteristics of included studies

#### Types of populations

Included studies examined the effect of aromatherapy on fatigue

- among people with chronic musculoskeletal conditions (rheumatoid arthritis, knee osteoarthritis and knee pain, each in one trial)
- among people living with cancer or advanced disease (chemotherapy in two trials, any cancer in one trial)
- during pregnancy (among those experiencing nausea and vomiting)
- among people with other chronic conditions (mainly haemodialysis in 7 trials, other conditions in single studies)

The specific condition addressed in each trial is reported in the forest plot (column 3, Figure 4.5.1 and Figure 4.5.2) with full details for each study including eligibility criteria, participant characteristics, and ICD 11 codes in Appendix E1.

#### Types of interventions

***Comparison 1 (any mode of aromatherapy delivery).*** Of the 18 trials included in Comparison 1, aromatherapy was delivered by inhalation in 13 trials and massage in five trials. Two of these 18 trials examined the effects of two different aromatherapy interventions; different essential oils were tested in one trial and aromatherapy was delivered by both inhalation and massage in the other trial. In both cases, the two treatment arms were combined prior to inclusion in the meta-analysis).

Nine trials evaluated a single essential oil, eight trials evaluated a blend of two or more oils, and in two trials participants were given a choice of oils from which to select. Lavender was the most commonly evaluated essential oil, either alone (6/19 trials), in a blend (5 trials) or as one of a selection of oils from which participants could choose (1 trial). Other oils evaluated in more than one trial were orange (4 trials, all in blends), peppermint (3 trials, all in blends), chamomile (2 trials), and ginger (2 trials, used alone).

The treatment period varied in length from a week in trials among populations treated for shorter-term fatigue (one trial in pregnancy, two trials among people undergoing chemotherapy), to 2-8 weeks in trials among people with chronic conditions (15 trials) or cancer (1 trial).

***Comparison 2 (aromatherapy delivered by massage)***. All of the trials that delivered aromatherapy by massage administered a single essential oil, although one trial had two aromatherapy groups (one orange, the other lavender). Lavender was evaluated in two trials, and the other oils were chamomile, orange, and ginger. In all four trials, multiple aromatherapy massages were administered over a two- to eight-week period (note, three of these four trials also contribute to Comparison 1).

**Types of outcomes**

The outcome measure from which data were included for meta-analysis is reported for each trial in the forest plots (column 2, Figure 4.5.1 and 4.5.2). Full details for each study are in Appendix E1, including the timing of outcome measurement in relation to intervention and details of which outcome was selected when multiple were available (e.g. when both overall and subscale scores were available). The appendix also reports studies in which fatigue was measured, but the population was ineligible for inclusion in the analysis.

All studies measured fatigue on a scale, the majority using the Fatigue Severity Scale (FSS; 7/19 trials). The EORTC QLQ-C30 and SF-36 vitality scale (also labelled as 'energy and fatigue') were each used in 2 trials. Eight other scales were used in a single trial, the Edmonton Symptom Assessment System (ESAS), the Piper fatigue scale (PFS), the Chalder Fatigue Scale (CFS), the Brief Fatigue Inventory (BFI), Multidimensional Fatigue Symptom Inventory (MFSI), the Numerical rating scale – fatigue (NRS-fatigue), the Multidimensional Fatigue Inventory (MFI) and the Rhoten fatigue scale (RFS).

All results were reported as a score on the original scale (e.g. fatigue on the Fatigue Severity Scale).

### Effects of aromatherapy on fatigue

**Comparison 1: Aromatherapy (any mode) versus inactive control (usual care, placebo, no intervention)**

The effect of aromatherapy on fatigue is uncertain overall (all population groups) and for each population group (chronic musculoskeletal pain, cancer and advanced disease, pregnancy, other chronic conditions).

Factors that reduced our certainty in the combined estimates of effect differed somewhat for each population group, as explained in the GRADE summary of findings table (Table 4.5.1, explanations). In combination, these factors raise concern that any observed benefit could be overestimated (or harm underestimated). Major concerns are as follows.

- ***Publication bias***. There is evidence that there could be studies (or results) missing from the analysis that show effects favouring the control (i.e. selective non-reporting based on the direction and statistical significance of results).
- ***Risk of bias in included trials***. All trials in the analysis have methodological limitations (high risk of bias or some concerns). The absence of trials at low risk of bias meant that it was not possible to examine the impact of these methodological limitations on the estimate of the intervention effect using the approach specified in our protocol (limiting analyses to studies at low risk of bias).
- ***Inconsistent results that lead to different conclusions about the effects of aromatherapy***. There is evidence that the size of the intervention effect differs across studies beyond what would be expected by chance. These differences were not explained by population group (Figure 4.5.1; as explained in Appendix D, Section D.4) or the mode by which aromatherapy was delivered (Appendix D, Section D.4 and Figure D.4.1). This reduced our confidence in the combined estimate because some studies found an important reduction in fatigue (greater than the threshold for important benefit, an SMD of - 0.2 or lower) while others found little or no difference between aromatherapy and control, with no credible evidence to explain whether this reflects true differences in the effects of aromatherapy or methodological problems in some studies.

Concerns relating to each finding were considered in the GRADE assessment when interpreting the result. The findings are as follows.

**Results for which an interpretation was made** (low certainty evidence)

**None**.

**Results considered too uncertain to interpret**

- ***Overall (all population groups).*** The evidence is very uncertain about the effect of aromatherapy (any mode) on fatigue (18 studies, 1316 participants; very low certainty, Figure 4.5.1).
- ***Chronic musculoskeletal conditions***. The evidence is very uncertain about the effect of aromatherapy (any mode) on chronic musculoskeletal conditions (1 trial, 34 participants with rheumatoid arthritis; very low certainty).

- ***Cancer and advanced disease***. The evidence is very uncertain about the effect of aromatherapy (any mode) on fatigue among people living with cancer and advanced disease (3 studies, 398 participants; very low certainty).
- ***Pregnancy.*** The evidence is very uncertain about the effect of aromatherapy (any mode) on fatigue in pregnancy (1 study, 89 participants; very low certainty).
- ***Other chronic conditions***. The evidence is very uncertain about the effect of aromatherapy (any mode) on fatigue among people with other chronic conditions (mainly those undergoing haemodialysis for kidney disease). (13 studies, 795 participants; very low certainty, Figure 4.5.1).

**Table 4.5.1**. Summary of findings for Comparison 1. the effect of aromatherapy (any mode) versus inactive control (usual care, no intervention, placebo) on fatigue.

| Outcomes | Anticipated absolute effects* (95% CI) | | Relative effect (95% CI) | № of participants (studies) | Certainty of the evidence (GRADE) | Comments |
|---|---|---|---|---|---|---|
| | **Risk with inactive control (usual care, placebo, no intervention)** | **Risk with aromatherapy (any mode)** | | | | |
| Fatigue: All population groups | - | SMD **0.78 SD lower** (1.15 lower to 0.41 lower) | - | 1316 (18 RCTs) | ⨁◯◯◯ Very low[a,b,c] | The evidence is very uncertain the effect of aromatherapy (any mode) on fatigue. |
| Fatigue: chronic musculoskeletal conditions | - | SMD **0.96 SD lower** (1.65 lower to 0.26 lower) | - | 34 (1 RCT) | ⨁◯◯◯ Very low[d,e,f] | The evidence is very uncertain about the effect of aromatherapy (any mode) on fatigue for people with chronic musculoskeletal conditions. |
| Fatigue: cancer & advanced disease | - | SMD **0.17 SD lower** (0.90 lower to 0.56 higher) | - | 398 (3 RCTs) | ⨁◯◯◯ Very low[g,h,i] | The evidence is very uncertain about the effect of aromatherapy (any mode) on fatigue among people living with cancer and advanced disease. |
| Fatigue during pregnancy | - | SMD **0.22 SD lower** (0.63 lower to 0.19 higher) | - | 89 (1 RCT) | ⨁◯◯◯ Very low[e,g,j,k] | The evidence is very uncertain about the effect of aromatherapy (any mode) on fatigue during pregnancy. |
| Fatigue: other chronic conditions (mainly people undergoing haemodialysis for kidney disease; also insomnia, pre-diabetes, neuropathic pain) | - | SMD **0.99 SD lower** (1.47 lower to 0.52 lower) | - | 795 (13 RCTs) | ⨁◯◯◯ Very low[g,l,m] | The evidence is very uncertain about the effect of aromatherapy (any mode) on fatigue among people with other chronic conditions (mainly those undergoing haemodialysis for kidney disease). |

***The risk in the intervention group** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI). **CI:** confidence interval; **SMD:** standardised mean difference

The threshold for an important difference was an SMD of 0.2 (used for interpreting point estimates and confidence intervals). For fatigue, the resulting interpretation is: < -0.2 is beneficial, -0.2 to 0.2 is trivial or unimportant ("little or no difference" between treatments), > 0.2 is harmful

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

## Explanations

a. Serious risk of bias (-1). All studies in the analysis are at high risk of bias or some concerns, such that the observed benefit may be overestimated.

b. Serious inconsistency (-1). Confidence intervals do not overlap for many studies, heterogeneity statistics indicate inconsistency, and the prediction interval indicates that the effect in a new study could range from important benefit to important harm. However, the point estimate for most studies (16/18; majority of weight in analysis) indicates important benefit (SMD <-0.2) or trivial effect (SMD -0.2 to 0.2; 2/18) not harm. For this reason, we have downgraded for serious not very serious inconsistency.

c. Publication bias strongly suspected (-1). Evidence from sensitivity analysis and contour enhanced funnel plot that there could be missing studies which show effects favouring the control, especially nonsignificant results (see Appendix D). Applies overall and to population groups with a high proportion of small studies showing large, statistically significant

effects favouring aromatherapy (combined effect estimate is moderate to large). Publication bias is not suspected for population groups for which the combined effect estimate is trivial (i.e. an unimportant effect).

d. Very serious risk of bias (-2). All studies in analysis are at high risk of bias.

e. Inconsistency not assessed: single study

f. Serious indirectness (-1): Evidence from one small study among people with rheumatoid arthritis. Uncertain whether results apply to other populations with chronic musculoskeletal conditions.

g. Serious risk of bias (-1). All studies in analysis are at high risk of bias or some concerns; however there is little or no difference between treatments so downgraded for serious not very serious risk of bias.

h. Serious inconsistency (-1). Heterogeneity statistics indicate inconsistency, confidence intervals do not overlap and the point estimate indicates a trivial effect in 2 of 3 studies (SMD -0.2 to 0.2) and important benefit in the third (SMD -0.2 or lower).

i. Serious imprecision (-1). The 95% confidence interval crosses the threshold for both a small but important reduction in fatigue (SMD -0.2) and a small but important increase in fatigue (SMD 0.2), so the result is compatible with important benefit (SMD 0.9 lower) and important harm (SMD 0.56 higher). However, we have downgraded by -1 because the imprecision is likely influenced by inconsistent results (rated as serious).

j. Serious indirectness (-1): Evidence from one small study among people who are pregnant. Uncertain whether results apply to pregnancy in general.

k. Serious imprecision (-1). The 95% confidence interval crosses the threshold for small but important reduction in fatigue (SMD of -0.2), which means the result is compatible with important benefit (SMD 0.63 lower) and little or no difference (SMD 0.19 higher).

l. Serious inconsistency (-1). Confidence intervals do not overlap for many studies, heterogeneity statistics indicate inconsistency, and the prediction interval indicates that the effect in a new study could range from important benefit to important harm. However, the point estimate for all studies indicates a reduction in fatigue (SMD -0.2 or lower), not a trivial effect or harm. For this reason, we have downgraded for serious not very serious inconsistency.

m. Publication bias strongly suspected (-1). Evidence from contour enhanced funnel plot from the overall analysis that there could be missing studies which show effects favouring the control, especially nonsignificant results (see Appendix D). This subgroup contains the majority of studies (13/18) including multiple small studies showing large effects.

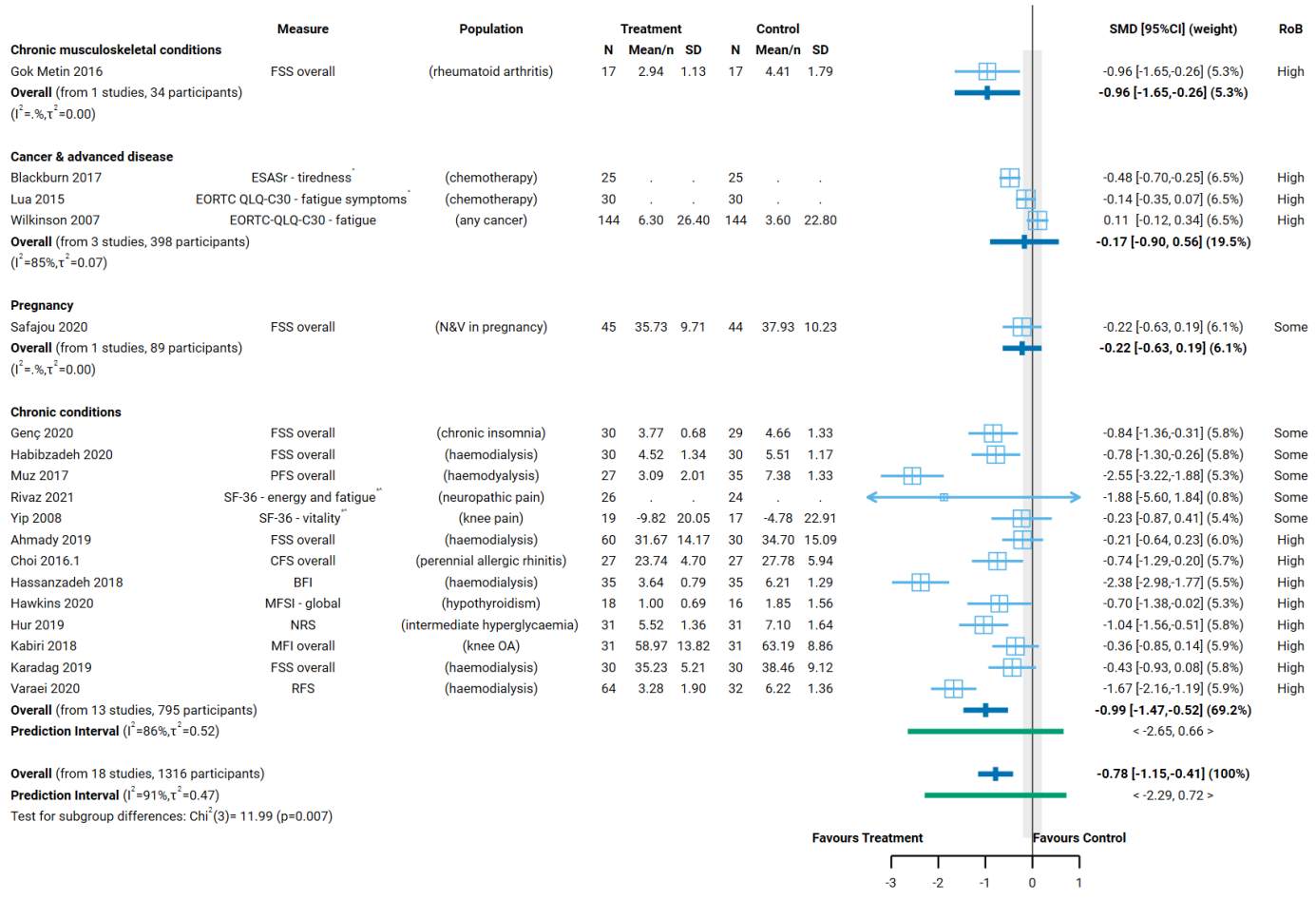


**Fig 4.5.1** | Forest plot for comparison 1. the effect of aromatherapy (any mode) versus inactive control (usual care, no intervention, placebo) on fatigue. SMD=standardised mean difference. Blue lines show 95% confidence intervals (CI) and green lines show prediction intervals (PI). The shaded grey area indicates the pre-specified range where the effect of aromatherapy is considered to be no different from control (SMD -0.2 to 0.2 standard units). ^ indicates studies for which data transformation or imputation was required to include the result in the meta-analysis. This may include crossover trials and studies that reported results as a dichotomous or ordinal outcome (identifiable because no mean or SD is reported for the study in the forest plot). * Denotes studies for which the direction of effect was changed to match the overall plot (e.g. positive numbers are beneficial).

### Comparison 2. Aromatherapy (massage) versus massage

All studies in Comparison 2 were among people with chronic conditions. The evidence is very uncertain about the effect of aromatherapy (massage) on chronic conditions (4 trials, 252 participants; very low certainty). The four small studies are among people receiving aromatherapy for very different underlying conditions (kidney disease, knee pain, neuropathic pain) and it is unclear whether results would be similar for other populations with chronic conditions. Other factors that reduced our certainty in the combined estimate of effect are explained in the GRADE summary of findings table (Table 4.5.2, explanations). There were too few studies to detect publication bias for this analysis.

**Table 4.5.2**. Summary of findings for Comparison 2. the effect of aromatherapy (massage) versus inactive massage control on fatigue.

| Outcomes | Anticipated absolute effects* (95% CI) | | Relative effect (95% CI) | № of participants (studies) | Certainty of the evidence (GRADE) | Comments |
|---|---|---|---|---|---|---|
| | Risk with inactive massage control | Risk with aromatherapy (massage) | | | | |
| Fatigue: chronic conditions (people undergoing haemodialysis for kidney disease, neuropathic pain, knee pain) | - | SMD **0.38 SD lower** (0.93 lower to 0.17 higher) | - | 252 (4 RCTs) | ⨁◯◯◯ Very low[a,b,c] | The evidence is very uncertain about the effect of aromatherapy (massage) on fatigue among people with chronic conditions. |

***The risk in the intervention group** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI). **CI:** confidence interval; **SMD:** standardised mean difference

The threshold for an important difference was an SMD of 0.2 (used for interpreting point estimates and confidence intervals). For fatigue, the resulting interpretation is: < -0.2 is beneficial, -0.2 to 0.2 is trivial or unimportant ("little or no difference" between treatments), > 0.2 is harmful

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

## Explanations

a. Serious risk of bias (-1). All studies in the analysis have some concerns about risk of bias, such that the observed benefit may be overestimated.

b. Serious indirectness (-1). Evidence from four small studies among people receiving aromatherapy for very different underlying conditions (haemodialysis for kidney disease, neuropathic pain, knee pain). Uncertain whether results apply to populations with chronic conditions more generally.

c. Serious imprecision (-1). The 95% confidence interval crosses the threshold for small but important reduction in fatigue (SMD of -0.2), which means the result is compatible with important benefit (SMD 0.93 lower) and little or no difference (SMD 0.17 higher).

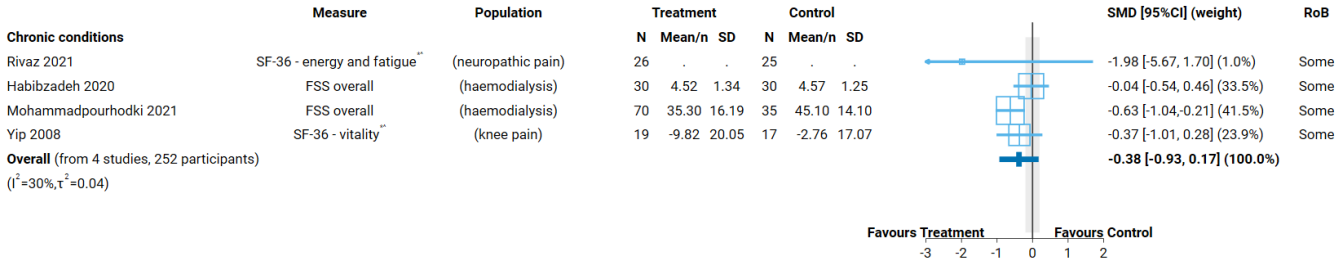


**Fig 4.5.2** | Forest plot for comparison 2. the effect of aromatherapy (massage) versus inactive massage control on fatigue. SMD=standardised mean difference. Blue lines show 95% confidence intervals (CI) and green lines show prediction intervals (PI). The shaded grey area indicates the pre-specified range where the effect of aromatherapy is considered to be no different from control (SMD -0.2 to 0.2 standard units). ^ indicates studies for which data transformation or imputation was required to include the result in the meta-analysis. This may include crossover trials and studies that reported results as a dichotomous or ordinal outcome (identifiable because no mean or SD is reported for the study in the forest plot). * Denotes studies for which the direction of effect was changed to match the overall plot (e.g. positive numbers are beneficial).

## 4.6 Emotional functioning and mental health

To be eligible for this analysis, there had to be either an acute indication (i.e. to prevent perioperative anxiety) or evidence that participants had signs/symptoms of mental distress (i.e. either this was part of the trial eligibility criteria or the baseline data indicated mental distress or a diagnosed mental disorder).

Overall, 92 studies that examined the effect of aromatherapy on emotional functioning and mental health were included for meta-analysis, five of these contribute to both Comparison 1 and 2. A further 17 studies were eligible for one or both of the emotional functioning and mental health meta-analyses, but could not be included (see below).

***Comparison 1***.

- Eighty-six studies (7032 participants) contribute to the comparison of aromatherapy delivered by any mode compared to an inactive control (Figure 4.6.1).
- An additional 13 trials (1318 participants) were eligible for this comparison, but either did not report results that could be included in the meta-analysis, or the results were unavailable or uninterpretable. Most notable in terms of the amount of data unavailable for analysis, were missing results from trials among people undergoing procedures (4 trials, 642 participants) and people living with dementia (4 trials, 186 participants).

***Comparison 2***.

- Eleven studies (664 participants) contribute to the comparison of aromatherapy delivered by massage to an inactive massage control (comparable to massage in the aromatherapy arm) (Figure 4.6.2).
- Five additional trials (260 participants) were eligible for this comparison, but either did not report results that could be included in the meta-analysis, or the results were unavailable or uninterpretable.

### Characteristics of included studies

#### Types of populations

Included studies examined the effect of aromatherapy on emotional functioning and mental health among people

- undergoing surgery (perioperative anxiety; 14 different types of surgery across 17 trials)
- undergoing procedures (periprocedural anxiety, mainly adults; more than 15 different procedures across 33 studies, including angiography, biopsy, haemodialysis, burn dressing changes, gynaecological procedures, dental procedures, catheterisation)
- living with cancer and advanced disease (mainly undergoing chemotherapy)
- in hospital (anxiety, mainly among cardiovascular inpatients)
- during labour and childbirth (anxiety)
- living with behaviour change from dementia (mainly agitation)
- with symptoms of mental distress (mainly depression and menopause)

The specific condition addressed in each trial is reported in the forest plot (column 3, Figure 4.6.1 and Figure 4.6.2) with full details for each study including eligibility criteria, participant characteristics, and ICD 11 codes in Appendix E1.

#### Types of interventions

***Comparison 1 (any mode of aromatherapy delivery).*** Of the 86 trials included in Comparison 1, aromatherapy was delivered by inhalation in 67 trials, by massage in 14 trials and topically in 5 trials. Ten of the 86 trials examined the effects of two or more aromatherapy treatments that we combined prior to inclusion of results in the meta-analysis. These were different essential oils (5 trials), a different dose (1 trial), mode of delivering the same essential oil (1 trial), or aromatherapy with a co-intervention (1 trial).

Lavender was the most commonly evaluated essential oil (50 trials), either alone or in a blend, followed by orange (14 trials, 4 in blends), rose (12 trials), bergamot (8 trials), chamomile (5 trials), and geranium (5 trials). A number of other essential oils were evaluated in a single trial. Six trials evaluated a blend of essential oils, and in two trials participants were given a choice from a selection of oils.

The aromatherapy treatment period varied in length, but this generally reflected the treatment goal (i.e. for an acute or chronic indication).

- In trials addressing anxiety perioperatively, peri-procedurally, during labour and birth, or at the time of cancer treatment, the aromatherapy treatment period was less than a day in 55 trials and up to a week in 14 trials.
- In trials among people undergoing dialysis, living with cancer, behaviour change in dementia, or mental distress, the aromatherapy treatment period was a month or more in 15 trials and 2-3 weeks in two trials.

***Comparison 2 (aromatherapy delivered by massage)***. Of the trials in Comparison 2, lavender (4/11 trials) and chamomile (3/11 trials) were the most commonly evaluated essential oils. The aromatherapy treatment period was days or less for those undergoing surgery or hospitalised, and 3-4 weeks in trials involving people with longer-term or chronic conditions (dementia, cancer, mental distress).

**Types of outcomes**

The outcome measure from which data were included for meta-analysis is reported for each trial in the forest plots (column 2, Figure 4.6.1 and Figure 4.6.2). Full details for each study are in Appendix E1, including the timing of outcome measurement in relation to intervention and details of which outcome was selected when multiple were available.

All studies measured outcomes on a scale. The State-Trait Anxiety Inventory (STAI) was used in the vast majority of studies of aromatherapy for anxiety related to surgery, procedures, labour and childbirth, and hospitalisation (44 trials), although the version and scales used varied (most used the state subscale). In studies involving people with dementia, the Cohen Mansfield Agitation Inventory was used in five of seven trials. In studies among people living with cancer and those experiencing mental distress, the scales used varied considerably, most used only in a single study.

Results for all but one study were reported as a score on the original scale (e.g. state anxiety on the STAI-S scale). An effect estimate was calculated from the dichotomised data from studies that did not report on the scale and transformed to a standardised mean difference (Appendix B).

### Effects of aromatherapy on emotional functioning and mental health

**Comparison 1: Aromatherapy (any mode) versus inactive control (usual care, placebo, no intervention)**

The effect of aromatherapy on emotional functioning and mental health is uncertain overall (all population groups) and for each population group. Aromatherapy may improve emotional functioning and mental health for some population groups (e.g. reducing symptoms of mental distress) and have little or no effect for others (e.g. agitation among people with dementia, distress among people living with cancer); however, there are multiple factors that make these results uncertain.

Factors that reduced our certainty in the combined estimates of effect differed somewhat for each population group, as explained in the GRADE summary of findings table (Table 4.6.1, explanations). In combination, these factors raise concern that any observed benefit could be overestimated (or harm underestimated). Major concerns are as follows.

- ***Publication bias***. There is evidence that there could be studies (or results) missing from the analysis that show effects favouring the control (i.e. selective non-reporting based on the direction and statistical significance of results).
- ***Risk of bias in included trials***. All trials in the analysis have methodological limitations (high risk of bias or some concerns). The absence of trials at low risk of bias meant that it was not possible to examine the impact of methodological limitations on the estimate of the intervention effect using the approach specified in our protocol (limiting analyses to studies at low risk of bias).
- ***Inconsistent results that lead to different conclusions about the effects of aromatherapy***. There is evidence that the size of the intervention effect differs across studies beyond what would be expected by chance. These differences were not explained by population group (Figure 4.6.1; as explained in Appendix D, Section D.5) or the mode by which aromatherapy was delivered (Appendix D, Section D.5 and Figure D.5.1). This reduced our confidence in the combined estimate because some studies found an important improvement in emotional functioning and mental health (below the threshold for important benefit, an SMD of - 0.2 or lower, indicating a reduction in anxiety, for example) while others found little or no difference between aromatherapy and control, with no credible evidence to explain whether this reflects true differences in the effects of aromatherapy or methodological problems in some studies.

Concerns relating to each finding were considered in the GRADE assessment when interpreting the result. The findings are as follows.

**Results for which an interpretation was made** (low certainty evidence)

***Longer-term emotional functioning and mental health***

- ***Cancer and advanced disease***. Aromatherapy (any mode) may make little to no difference to emotional functioning and mental health among people with cancer and advanced disease (SMD 0.11 lower, 95% CI 0.48 lower to 0.26 higher (lower is better, e.g. less distress); $I^2$ = 81 %; 7 studies, 275 participants; low certainty, Figure 4.6.1).
- ***Dementia.*** Aromatherapy (any mode) may make little to no difference to agitation among people with dementia (SMD 0.08 lower, 95% CI 0.38 lower to 0.21 higher (lower is better, e.g. less agitation); $I^2$ = 57 %; 7 studies, 521 participants; low certainty, Figure 4.6.1).
- ***Mental distress***. Aromatherapy (any mode) may improve emotional functioning and mental health among people with symptoms of mental distress (mainly depression symptoms) (SMD 1.14 lower, 95% CI 1.94 lower to 0.34 lower (lower is better, e.g. less agitation); $I^2$ = 85 %; 5 studies, 440 participants; low certainty, Figure 4.6.1).

**Results considered too uncertain to interpret**

- ***Overall (all population groups).*** The evidence is very uncertain about the effect of aromatherapy (any mode) on emotional functioning and mental health (86 studies, 7032 participants; very low certainty, Figure 4.6.1).

***Short-term anxiety arising from stressful situations***

- ***Surgery***. The evidence is very uncertain about the effect of aromatherapy (any mode) on perioperative anxiety (i.e. in the period immediately before surgery) (17 studies, 1428 participants; very low certainty, Figure 4.6.1).
- ***Procedures***. The evidence is very uncertain about the effect of aromatherapy (any mode) on periprocedural anxiety (i.e. before or during a procedure) (33 studies, 2854 participants; very low certainty, Figure 4.6.1).
- ***Hospitalisation.*** The evidence is very uncertain about the effect of aromatherapy (any mode) on anxiety during hospitalisation for people admitted for cardiovascular conditions (12 studies, 1030 participants; very low certainty, Figure 4.6.1).
- ***Labour and childbirth***. The evidence is very uncertain about the effect of aromatherapy (any mode) on anxiety during labour and childbirth (5 studies, 484 participants; very low certainty, Figure 4.6.1).

None of the studies included for this analysis examined the effects of aromatherapy (any mode) among people with a diagnosed mental disorder (e.g. generalised anxiety disorder or depressive disorder).

**Table 4.6.1**. Summary of findings for Comparison 1. the effect of aromatherapy (any mode) versus inactive control (usual care, no intervention, placebo) on emotional functioning and mental health.

| Outcomes | Anticipated absolute effects* (95% CI) | | Relative effect (95% CI) | № of participants (studies) | Certainty of the evidence (GRADE) | Comments |
|---|---|---|---|---|---|---|
| | **Risk with inactive control (usual care, placebo, no intervention)** | **Risk with aromatherapy (any mode)** | | | | |
| Emotional functioning and mental health: any population | - | SMD **0.90 SD lower** (1.18 lower to 0.61 lower) | - | 7032 (86 RCTs) | ⨁◯◯◯ Very low[a,b,c] | The evidence is very uncertain about the effect of aromatherapy (any mode) on emotional functioning and mental health. |
| Emotional functioning and mental health: perioperative anxiety (surgery) | - | SMD **0.83 SD lower** (1.30 lower to 0.36 lower) | - | 1428 (17 RCTs) | ⨁◯◯◯ Very low[a,c,d] | The evidence is very uncertain about the effect of aromatherapy (any mode) on perioperative anxiety. |

| Outcomes | Anticipated absolute effects* (95% CI) | | Relative effect (95% CI) | № of participants (studies) | Certainty of the evidence (GRADE) | Comments |
|---|---|---|---|---|---|---|
| | **Risk with inactive control (usual care, placebo, no intervention)** | **Risk with aromatherapy (any mode)** | | | | |
| Emotional functioning and mental health: periprocedural anxiety (procedures) | - | SMD **0.89 SD lower** (1.23 lower to 0.54 lower) | - | 2854 (33 RCTs) | ⨁◯◯◯ Very low[a,c,e] | The evidence is very uncertain about the effect of aromatherapy (any mode) on periprocedural anxiety. |
| Emotional functioning and mental health among people with cancer and advanced disease | - | SMD **0.11 SD lower** (0.48 lower to 0.26 higher) | - | 275 (7 RCTs) | ⨁⨁◯◯ Low[f,g,h] | Aromatherapy (any mode) may make little to no difference to emotional functioning and mental health among people with cancer and advanced disease. |
| Emotional functioning and mental health: anxiety during hospitalisation (mainly cardiovascular inpatients) | - | SMD **0.82 SD lower** (1.08 lower to 0.56 lower) | - | 1030 (12 RCTs) | ⨁◯◯◯ Very low[c,ij,] | The evidence is very uncertain about the effect of aromatherapy (any mode) on anxiety during hospitalisation. |
| Emotional functioning and mental health: anxiety during labour and childbirth | - | SMD **3.71 SD lower** (8.73 lower to 1.31 higher) | - | 484 (5 RCTs) | ⨁◯◯◯ Very low[c,i,k,l] | The evidence is very uncertain about the effect of aromatherapy (any mode) on anxiety during labour and childbirth. |
| Emotional functioning and mental health among people with dementia (mainly agitation) | - | SMD **0.08 SD lower** (0.38 lower to 0.21 higher) | - | 521 (7 RCTs) | ⨁⨁◯◯ Low[f,g,m] | Aromatherapy (any mode) may make little to no difference to agitation among people with dementia. |
| Emotional functioning and mental health among people with symptoms of mental distress (mainly depression) | - | SMD **1.14 SD lower** (1.94 lower to 0.34 lower) | - | 440 (5 RCTs) | ⨁⨁◯◯ Low[a,c,g] | Aromatherapy (any mode) may improve emotional functioning and mental health among people with symptoms of mental distress (mainly depression). |
| Emotional functioning and mental health among people with a diagnosed mental disorder (e.g. depression, anxiety) - not reported | - | - | - | - | - | None of the studies included for this analysis examined the effects of aromatherapy on mental health among people with a diagnosed mental disorder. |

***The risk in the intervention group** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI). **CI:** confidence interval; **SMD:** standardised mean difference

The threshold for an important difference was an SMD of 0.2 (used for interpreting point estimates and confidence intervals). For emotional functioning and mental health, the resulting interpretation is: < -0.2 is beneficial, -0.2 to 0.2 is trivial or unimportant ("little or no difference" between treatments), > 0.2 is harmful

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

## Explanations

a. Serious risk of bias (-1). All studies in the analysis are at high risk of bias or some concerns, such that the observed benefit may be overestimated.

b. Serious inconsistency (-1). Confidence intervals do not overlap for many studies, heterogeneity statistics indicate inconsistency, and the prediction interval indicates that the effect in a new study could range from important benefit to important harm. However, the point estimate for most studies (majority of weight in analysis) indicates important benefit (SMD <-0.2) or trivial effect (SMD -0.2 to 0.2) not harm. For this reason, we have downgraded for serious not very serious inconsistency.

c. Publication bias strongly suspected (-1). Evidence from contour enhanced funnel plot that there could be missing studies which show effects favouring the control, especially nonsignificant effects (see Appendix D). Applies overall and to population groups with a high proportion of small studies showing large, statistically significant effects favouring aromatherapy (combined effect estimate is moderate to large). Publication bias is not suspected for population groups for which the combined effect estimate is trivial (i.e. an unimportant effect).

d. Serious inconsistency (-1). Confidence intervals do not overlap for many studies, heterogeneity statistics indicate inconsistency, and the prediction interval indicates that the effect in a new study could range from important benefit to important harm. However, the point estimate for most studies (11/17; majority of weight in analysis) indicates important benefit (SMD <-0.2) or trivial effect (SMD -0.2 to 0.2; 6/17) not harm. For this reason, we have downgraded for serious not very serious inconsistency.

e. Serious inconsistency (-1). Confidence intervals do not overlap for many studies, heterogeneity statistics indicate inconsistency, and the prediction interval indicates that the effect in a new study could range from important benefit to important harm. However, the point estimate for most studies (26/33; majority of weight in analysis) indicates important benefit (SMD <-0.2) or trivial effect (SMD -0.2 to 0.2; 5/33) not harm. For this reason, we have downgraded for serious not very serious inconsistency.

f. Serious risk of bias (-1). Most studies in analysis are at high risk of bias (majority of weight); however there is little or no difference between treatments so downgraded for serious not very serious risk of bias.

g. No serious inconsistency. Heterogeneity statistics suggest inconsistent results, however the confidence intervals overlap for the majority of studies.

h. Serious imprecision (-1). The 95% confidence interval crosses two thresholds for a small by important effect (SMD of 0.2 and -0.2), so the result is compatible with important benefit (SMD -0.48 lower) and important harm (SMD 0.26 higher). However, the extent to which the threshold for harm is crossed is modest, so we have rated down for serious not very serious imprecision.

i. Very serious risk of bias (-2). Most studies in analysis are at high risk of bias (majority of weight)

j. No serious inconsistency. Heterogeneity statistics suggest inconsistent results, however the confidence intervals overlap for the majority of studies, and the prediction interval indicates that the effect in a new study could range from important benefit to unimportant benefit.

k. Very serious inconsistency (-2). Confidence intervals do not overlap for many studies, heterogeneity statistics indicate inconsistency, and point estimates vary widely.

l. Extremely serious imprecision (-3). The 95% confidence interval crosses the threshold for both important benefit (SMD -0.2) and important harm (SMD 0.2), and is too wide for the result to be interpretable (SMD -8.73 to 1.31).

m. Serious imprecision (-1). The 95% confidence interval crosses two thresholds for a small by important effect (SMD of 0.2 and -0.2), so the result is compatible with important benefit (SMD -0.38 lower) and important harm (SMD 0.21 higher). However, the extent to which the threshold for harm is crossed is modest, so we have rated down for serious not very serious imprecision.

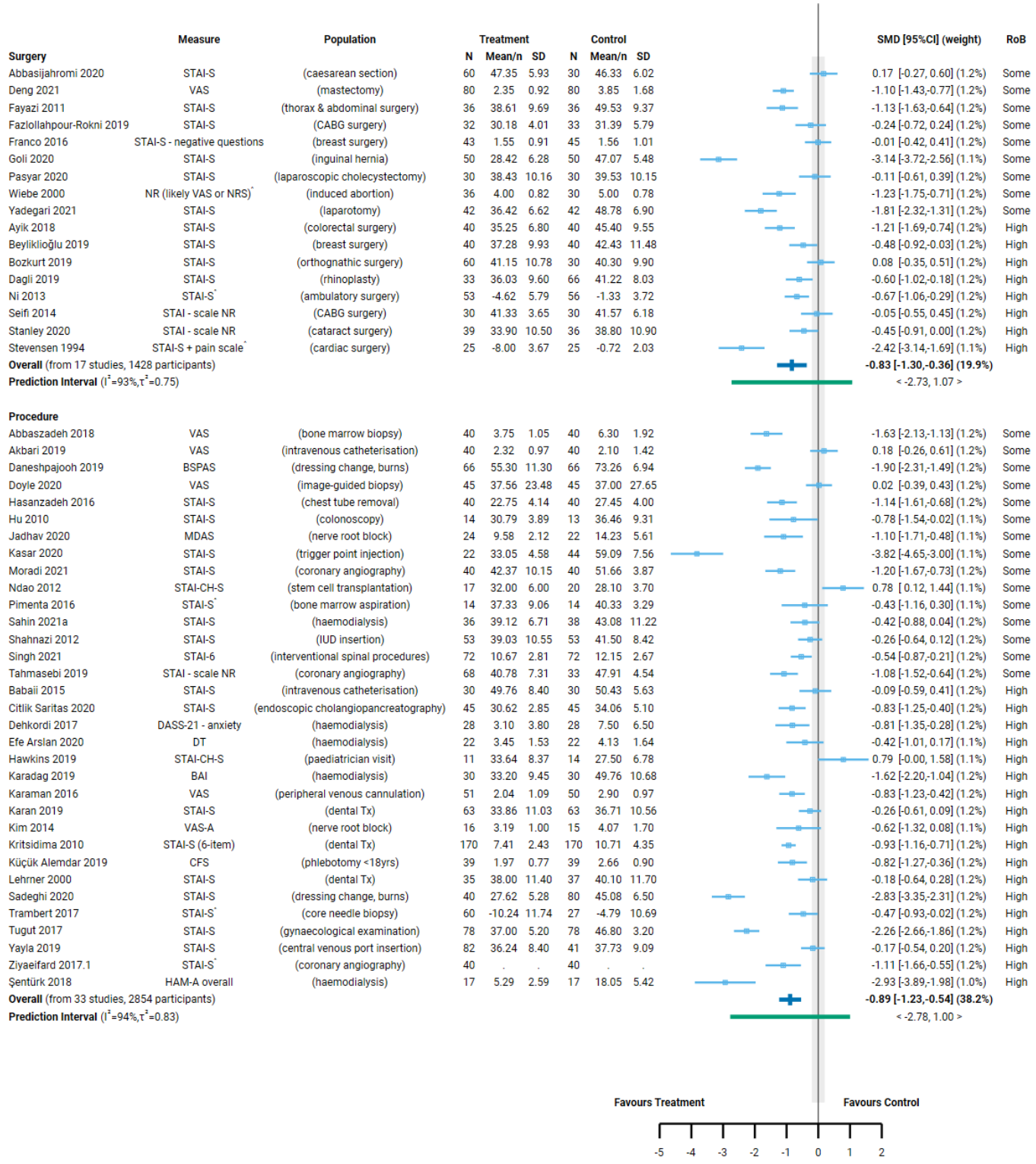


**Fig 4.6.1** | Forest plot for comparison 1. the effect of aromatherapy (any mode) versus inactive control (usual care, no intervention, placebo) on emotional functioning and mental health. See next page for continuation of plot and figure caption.

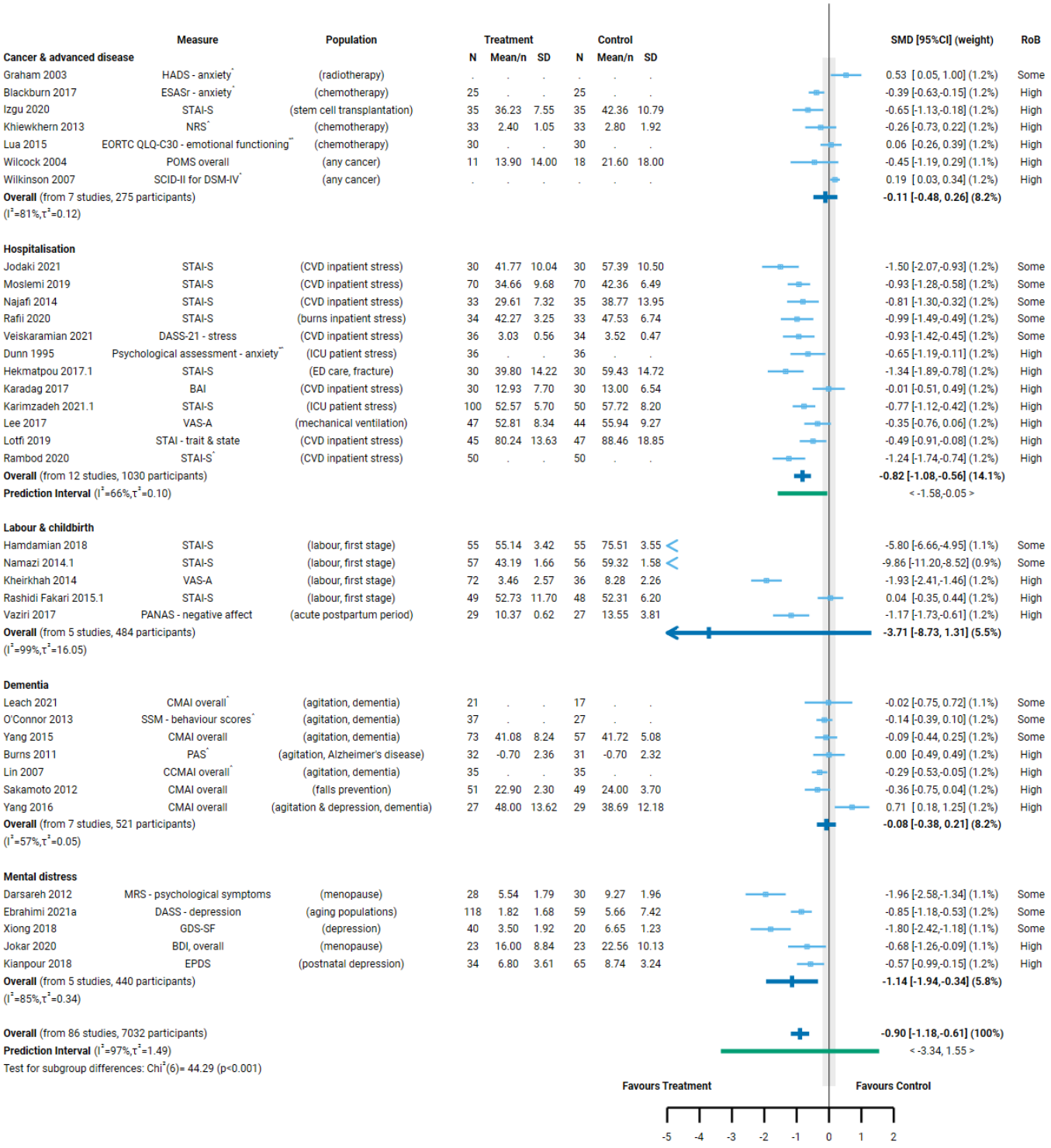


**Fig 4.6.1** | Forest plot for comparison 1. the effect of aromatherapy (any mode) versus inactive control (usual care, no intervention, placebo) on emotional functioning and mental health. SMD=standardised mean difference. Blue lines show 95% confidence intervals (CI) and green lines show prediction intervals (PI). The shaded grey area indicates the pre-specified range where the effect of aromatherapy is considered to be no different from control (SMD -0.2 to 0.2 standard units). ^ indicates studies for which data transformation or imputation was required to include the result in the meta-analysis. This may include crossover trials and studies that reported results as a dichotomous or ordinal outcome (identifiable because no mean or SD is reported for the study in the forest plot). * Denotes studies for which the direction of effect was changed to match the overall plot (e.g. positive numbers are beneficial).

**Comparison 2. Aromatherapy (massage) versus massage**

The effect of aromatherapy (massage) compared to massage alone on emotional functioning and mental health is uncertain overall (all population groups) and for each population group.

Factors that reduced our certainty in the combined estimates of effect differed somewhat for each population group, as explained in the GRADE summary of findings table (Table 4.6.2, explanations). In combination, these factors raise concern that any observed benefit could be overestimated (or harm underestimated). Major concerns are as follows.

- ***Publication bias***. There were too few studies to detect publication bias in this analysis; however, given the evidence for analyses with more studies, we cannot rule out the possibility that there could be studies (or results) missing from the analysis that show effects favouring the control.
- ***Risk of bias in included trials***. All trials in the analysis have methodological limitations (high risk of bias or some concerns). The absence of trials at low risk of bias meant that it was not possible to examine the impact of these methodological limitations on the estimate of the intervention effect using the approach specified in our protocol (limiting analyses to studies at low risk of bias).
- ***Inconsistent results that lead to different conclusions about the effects of aromatherapy***. There is evidence that the size of the intervention effect differs across studies beyond what would be expected by chance. These differences were partly explained by population group, so inconsistent effects are mainly a concern for the overall analysis (Figure 4.6.2; as explained in Appendix D, Section 4.5).

Concerns relating to each finding were considered in the GRADE assessment when interpreting the result. The findings are as follows.

**Results for which an interpretation was made** (low certainty evidence)

- ***Hospitalisation.*** Aromatherapy (massage) may reduce anxiety during hospitalisation slightly compared to massage alone (SMD 0.21 lower, 95% CI 0.47 lower to 0.06 higher (lower means less anxiety); $I^2$ = 0 %; 3 studies, 232 participants; low certainty, Figure 4.6.2).

**Results considered too uncertain to interpret**

- ***Overall (all population groups).*** The evidence is very uncertain about the effect of aromatherapy (massage) compared to massage alone on emotional functioning and mental health (all populations) (11 studies, 664 participants; very low certainty, Figure 4.6.2).
- ***Surgery***. The evidence is very uncertain about the effect of aromatherapy (massage) compared to massage alone on anxiety during labour and childbirth (2 studies, 130 participants; very low certainty). The result is too imprecise to interpret (compatible with both large harm and large benefit).
- ***Cancer and advanced disease***. The evidence is very uncertain about the effect of aromatherapy (massage) compared to massage alone on emotional functioning and mental health among people with cancer and advanced disease (2 studies, 134 participants; very low certainty).
- ***Dementia.*** The evidence is very uncertain about the effect of aromatherapy (massage) compared to massage alone on emotional functioning and mental health among people with dementia (agitation or other behaviour change) (2 studies, 85 participants; very low certainty).
- ***Mental distress***. The evidence is very uncertain about the effect of aromatherapy (massage) compared to massage alone on emotional functioning and mental health among people with symptoms of mental distress. (1 study, 57 participants; very low certainty).

None of the studies included for this analysis examined the effects of aromatherapy (massage) compared to massage alone among people undergoing procedures, during labour and childbirth, or among people with a diagnosed mental disorder.

**Table 4.6.2**. Summary of findings for Comparison 2. the effect of aromatherapy (massage) versus inactive massage control on emotional functioning and mental health.

| Outcomes | Anticipated absolute effects* (95% CI) | | Relative effect (95% CI) | № of participants (studies) | Certainty of the evidence (GRADE) | Comments |
|---|---|---|---|---|---|---|
| | Risk with inactive massage control | Risk with aromatherapy (massage) | | | | |
| Emotional functioning and mental health: any population | - | SMD **0.3 SD lower** (0.64 lower to 0.04 higher) | - | 637 (10 RCTs) | ⨁◯◯◯ Very low[a,b,c] | The evidence is very uncertain about the effect of aromatherapy (massage) compared to massage alone on emotional functioning and mental health (all populations). |
| Emotional functioning and mental health: perioperative anxiety (surgery) | - | SMD **0.45 SD lower** (6.42 lower to 5.52 higher) | - | 130 (2 RCTs) | ⨁◯◯◯ Very low[d,e,f] | The evidence is very uncertain about the effect of aromatherapy (massage) compared to massage alone on perioperative anxiety compared to massage alone. |
| Emotional functioning and mental health: periprocedural anxiety (procedures) - not reported | - | - | - | - | - | No studies included in this analysis report effects of aromatherapy compared to massage alone on people undergoing procedures. |
| Emotional functioning and mental health among people with cancer and advanced disease | - | SMD **0.22 SD lower** (1.48 lower to 1.93 higher) | - | 133 (2 RCTs) | ⨁◯◯◯ Very low[a,e,g,h] | The evidence is very uncertain about the effect of aromatherapy (massage) compared to massage alone on emotional functioning and mental health among people with cancer and advanced disease. |
| Emotional functioning and mental health: anxiety during hospitalisation (cardiovascular, burns and ICU patients) | - | SMD **0.21 SD lower** (0.47 lower to 0.06 higher) | - | 232 (3 RCTs) | ⨁⨁◯◯ Low[a,e,i] | Aromatherapy (massage) may reduce anxiety during hospitalisation slightly compared to massage alone. |
| Emotional functioning and mental health: anxiety during labour and childbirth - not reported | - | - | - | - | - | No studies included in this analysis report effects of aromatherapy (massage) compared to massage alone during labour and childbirth. |
| Emotional functioning and mental health among people with dementia (agitation or behavioural change) | - | SMD **0.46 SD lower** (3.03 lower to 2.12 higher) | - | 85 (2 RCTs) | ⨁◯◯◯ Very low[a,e,j,k] | The evidence is very uncertain about the effect of aromatherapy massage compared to massage alone on emotional functioning and mental health among people with dementia (or other behaviour change). |
| Emotional functioning and mental health among people with symptoms of mental distress (menopause only) | - | SMD **1.2 SD lower** (1.76 lower to 0.64 lower) | - | 57 (1 RCT) | ⨁◯◯◯ Very low[a,l,m,n] | The evidence is very uncertain about the effect of aromatherapy (massage) compared to massage alone on emotional functioning and mental health among people with symptoms of mental distress. |
| Emotional functioning and mental health among people with a diagnosed mental disorder (e.g. depression, anxiety) - not reported | - | - | - | - | - | No studies included in this analysis report effects of aromatherapy (massage) compared to massage alone among people with a diagnosed mental disorder. |

| Outcomes | Anticipated absolute effects* (95% CI) | | Relative effect (95% CI) | № of participants (studies) | Certainty of the evidence (GRADE) | Comments |
|---|---|---|---|---|---|---|
| | **Risk with inactive massage control** | **Risk with aromatherapy (massage)** | | | | |

***The risk in the intervention group** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI). **CI:** confidence interval; **SMD:** standardised mean difference

The threshold for an important difference was an SMD of 0.2 (used for interpreting point estimates and confidence intervals). For emotional functioning and mental health, the resulting interpretation is: < -0.2 is beneficial, -0.2 to 0.2 is trivial or unimportant ("little or no difference" between treatments), > 0.2 is harmful

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

## Explanations

a. Serious risk of bias (-1). All studies in the analysis are at high risk of bias or some concerns, such that the observed benefit may be overestimated.

b. Serious inconsistency (-1). Heterogeneity statistics indicate inconsistency. Confidence intervals overlap for the majority of studies (most compatible with both benefit and little or no difference), however the interpretation of the point estimate varies in direction and magnitude across studies.

c. Serious imprecision (-1). The 95% confidence interval crosses the threshold for small but important improvement in emotional functioning and mental health (SMD of -0.2), which means the result is compatible with important benefit (SMD -0.64 lower) and little or no difference (SMD 0.04 higher).

d. Very serious risk of bias (-2). All studies in analysis are at high risk of bias.

e. No serious inconsistency. Confidence intervals overlap for all studies, suggesting that any variation in results across studies may be unimportant.

f. Extremely serious imprecision (-3). The 95% confidence interval crosses the threshold for both small but important benefit (SMD -0.2) and small but important harm (SMD 0.2), and is too wide for the result to be interpretable (SMD -6.42 to 5.52).

g. Serious indirectness (-1). Evidence from two small studies among people with cancer. Uncertain whether results would be generalisable.

h. Very serious imprecision (-2). The 95% confidence interval crosses the threshold for both a small but important improvement in emotional functioning and mental health (SMD -0.2) and a small but important reduction (SMD 0.2), so the result is compatible with important benefit and important harm.

i. Serious imprecision (-1). The 95% confidence interval crosses the threshold for small but important improvement in emotional functioning and mental health (SMD of -0.2), which means the result is compatible with important benefit (SMD -0.47 lower) and little or no difference (SMD 0.06 higher).

j. Serious indirectness (-1). Evidence from two small studies among people living with dementia. Uncertain whether results would be generalisable.

k. Extremely serious imprecision (-3). The 95% confidence interval crosses the threshold for both small but important benefit (SMD -0.2) and small but important harm (SMD 0.2), and is too wide for the result to be interpretable (SMD -3.03 to 2.12).

l. Inconsistency not assessed: single study

m. Very serious indirectness (-2). Evidence from one small study among women during menopause. Uncertain whether results apply to populations with symptoms of mental distress (especially, symptoms of anxiety or depression) more generally.

n. No serious imprecision. Both the upper and lower limits of the 95% confidence interval are compatible with important benefit. Data are from a single small study with large effect.

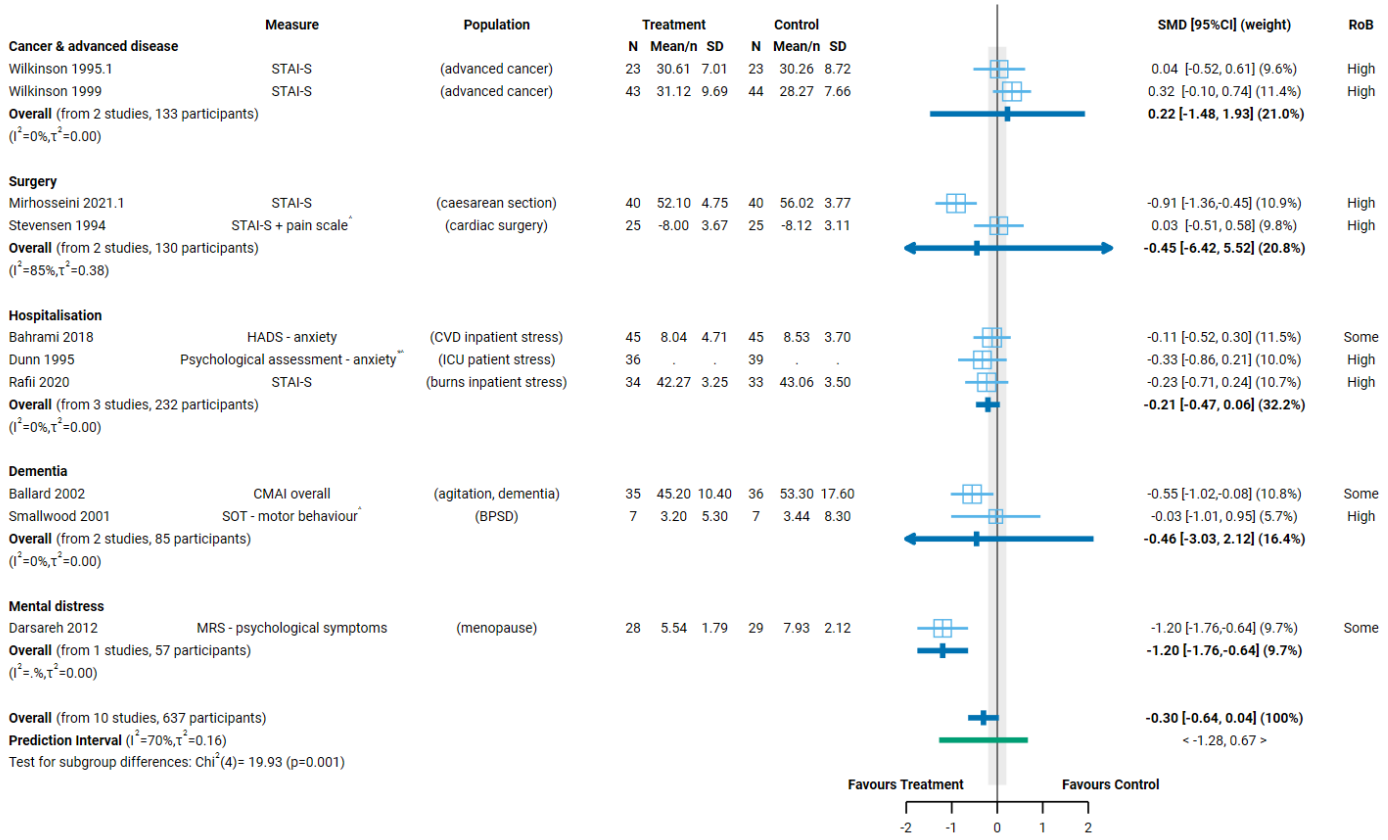


**Fig 4.6.2** | Forest plot for comparison 2. the effect of aromatherapy (massage) versus inactive massage control on emotional functioning and mental health. SMD=standardised mean difference. Blue lines show 95% confidence intervals (CI) and green lines show prediction intervals (PI). The shaded grey area indicates the pre-specified range where the effect of aromatherapy is considered to be no different from control (SMD -0.2 to 0.2 standard units). ^ indicates studies for which data transformation or imputation was required to include the result in the meta-analysis. This may include crossover trials and studies that reported results as a dichotomous or ordinal outcome (identifiable because no mean or SD is reported for the study in the forest plot). * Denotes studies for which the direction of effect was changed to match the overall plot (e.g. positive numbers are beneficial).

## 4.7 Health-related quality of life (HRQoL)

Overall, twenty studies that examined the effect of aromatherapy on HR-QoL were included for meta-analysis, six of these contribute to both Comparison 1 and 2. A further 4 studies (353 participants) were eligible for one or both of the HR-QoL meta-analyses, but could not be included (see below).

To be considered for the HR-QoL analysis, trials had to administer aromatherapy to a population living with cancer or a chronic condition for longer-term care (i.e. delivering treatment over weeks or longer, not days) and measure HR-QoL in a time-frame likely to detect meaningful improvement (i.e. generally 4 weeks or more from commencement treatment).

***Comparison 1***.

- Fourteen studies (1048 participants) contribute to the comparison of aromatherapy delivered by any mode compared to an inactive control (Figure 4.7.1).
- An additional 4 trials (353 participants) were eligible for this comparison, but either did not report results that could be included in the meta-analysis, or the results were unavailable or uninterpretable.

***Comparison 2***.

- Twelve studies (851 participants) contribute to the comparison of aromatherapy delivered by massage to an inactive massage control (comparable to massage in the aromatherapy arm) (Figure 4.7.2).
- One additional trial (118 participants) was eligible for this comparison, but it reported results that were unsuitable for analysis.

### Characteristics of included studies

#### Types of populations

Included studies examined the effect of aromatherapy on health-related quality of life (HR-QoL) among people living with

- Cancer and advanced disease (two involving people with advanced cancer, one involving people undergoing chemotherapy, one any cancer)
- chronic conditions (behavioural change in dementia in three trials; menopause, musculoskeletal conditions, and chronic kidney disease each in two trials; and single trials involving people with allergic rhinitis, diabetes, neuropathic pain, multiple sclerosis, lymphoedema, and chronic prostatitis)

The specific condition addressed in each trial is reported in the forest plot (column 3, Figure 4.7.1 and Figure 4.7.2) with full details for each study including eligibility criteria, participant characteristics, and ICD 11 codes in Appendix E1.

#### Types of interventions

***Comparison 1 (any mode of aromatherapy delivery).*** Of the 14 trials included in Comparison 1, aromatherapy was delivered by massage in 8 trials, by inhalation in 5 trials, and topically in 1 trial. One of these 14 trials examined the effects of two different aromatherapy interventions, testing aromatherapy delivered by both inhalation and massage. The two treatment arms were combined prior to inclusion in the meta-analysis).

Lavender was the most commonly evaluated essential oil, either alone (4/14 trials) or in a blend (3 trials), followed by chamomile (2 trials, one in a blend with lavender) and ginger (2 trials, one in a blend). Other essential oils were evaluated in a single trial, most as part of a blend of essential oils (3 trials). One trial tailored essential oil blends for individuals, and another gave participants a choice from a selection of essential oils.

In all trials among people with cancer, and a majority of trials among people with chronic conditions (7 of 11 trials), aromatherapy was administered over a period of four to twelve weeks. We included two trials among people with chronic conditions that administered aromatherapy multiple times over three weeks (since this was in keeping with our intent to include trials of longer-term care) and one trial among people with allergic rhinitis in which aromatherapy was delivered daily for seven days. In the latter case, we made an exception to the general rule because the core outcome set for this population indicated that seven day follow up was sufficient for measurement of HR-QoL outcomes.

***Comparison 2 (aromatherapy delivered by massage)***.

Chamomile and lavender were the most commonly evaluated essential oils used for aromatherapy massage (4 trials each), either alone or in a blend, followed by ginger and rosemary (2 trials, also used in blends). Other essential oils were evaluated in a single trial, alone or in a blend. Four trials evaluated a blend of essential oils. The treatment period was similar in studies that delivered aromatherapy by massage and by other modes (length of treatment from three to 12 weeks).

**Types of outcomes**

The outcome measure from which data were included for meta-analysis is reported for each trial in the forest plots (column 2, Figure 4.7.1 and Figure 4.7.2). Full details for each study are in Appendix E1, including the timing of outcome measurement in relation to intervention and details of which outcome was selected when multiple were available.

All but one study measured HR-QoL on a scale, the exception being a study among people with behaviour change in dementia for which % time in social withdrawal was measured. Across the nineteen studies that measured HR-QoL on a scale, thirteen different scales were used. The only scales used in more than one trial were the SF-36 general health scale (4 trials), the Rotterdam symptom checklist (3 trials among people living with cancer), and the Menopause Rating Scale (MRS; 2 trials). Many of the studies used a condition-specific measure. All results were reported by trialists as a score on the original scale.

### Effects of aromatherapy on health-related quality of life

**Comparison 1: Aromatherapy (any mode) versus inactive control (usual care, placebo, no intervention)**

Aromatherapy may improve health-related quality of life overall (all population groups) and for people with chronic conditions. Effects are uncertain for people living with cancer and advanced disease.

Factors that reduced our certainty in the combined estimates of effect differed somewhat for each population group, as explained in the GRADE summary of findings table (Table 4.7.1, explanations). In combination, these factors raise concern that any observed benefit could be overestimated (or harm underestimated). Major concerns are as follows.

- ***Publication bias***. There is evidence that there could be studies (or results) missing from the analysis that show effects favouring the control (i.e. selective non-reporting based on the direction and statistical significance of results).
- ***Risk of bias in included trials***. All trials in the analysis have methodological limitations (high risk of bias or some concerns). The absence of trials at low risk of bias meant that it was not possible to examine the impact of these methodological limitations on the estimate of the intervention effect using the approach specified in our protocol (limiting analyses to studies at low risk of bias).

Concerns relating to each finding were considered in the GRADE assessment when interpreting the result. The findings are as follows.

**Results for which an interpretation was made** (low certainty evidence)

- ***Overall (all population groups).*** Aromatherapy (any mode) may improve health-related quality of life. (SMD 0.54 higher, 95% CI 0.13 higher to 0.94 higher; $I^2$ = 87 %; 14 studies, 1048 participants; low certainty, Figure 4.7.1).

**Results considered too uncertain to interpret**

- ***Cancer and advanced disease***. The evidence is very uncertain about the effect of aromatherapy (any mode) on cancer and advanced disease (3 studies, 527 participants; very low certainty).
- ***Chronic conditions***. The evidence is very uncertain about the effect of aromatherapy (any mode) on health-related quality of life for people with chronic conditions (11 studies, 521 participants; very low certainty Figure 4.7.1).

**Table 4.7.1**. Summary of findings for Comparison 1. the effect of aromatherapy (any mode) versus inactive control (usual care, no intervention, placebo) on health-related quality of life.

| Outcomes | Anticipated absolute effects* (95% CI) | | Relative effect (95% CI) | № of participants (studies) | Certainty of the evidence (GRADE) | Comments |
|---|---|---|---|---|---|---|
| | **Risk with inactive control (usual care, placebo, no intervention)** | **Risk with aromatherapy (any mode)** | | | | |
| Health-related quality of life: all populations | - | SMD **0.54 SD higher** (0.13 higher to 0.94 higher) | - | 1048 (14 RCTs) | ⨁⨁◯◯ Low[a,b,c,d] | Aromatherapy (any mode) may improve health-related quality of life. |
| Health-related quality of life among people with cancer and advanced disease | - | SMD **0.34 SD higher** (0.90 lower to 1.57 higher) | - | 527 (3 RCTs) | ⨁◯◯◯ Very low[e,f,g] | The evidence is very uncertain about the effect of aromatherapy (any mode) on HR-QoL among people living with cancer and advanced disease. |
| Health-related quality of life among people with chronic conditions | - | SMD **0.59 SD higher** (0.06 higher to 1.11 higher) | - | 521 (11 RCTs) | ⨁◯◯◯ Very low[a,b,d,h] | The evidence is very uncertain about the effect of aromatherapy (any mode) on health-related quality of life for people with chronic conditions. |

*__The risk in the intervention group__ (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI). **CI:** confidence interval; **SMD:** standardised mean difference

The threshold for an important difference was an SMD of 0.2 (used for interpreting point estimates and confidence intervals). For HR-QoL, the resulting interpretation is: > 0.2 is beneficial, -0.2 to 0.2 is trivial or unimportant ("little or no difference" between treatments), < -0.2 is harmful

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

## Explanations

a. Serious risk of bias (-1). All studies in the analysis are at high risk of bias or some concerns, such that the observed benefit may be overestimated.

b. No serious inconsistency. Confidence intervals overlap for most studies in the overall analysis and in the analysis for the chronic conditions subgroup, suggesting that any variation in results across studies may be unimportant.

c. No serious imprecision. The 95% confidence interval crosses the threshold for a small but important improvement in HR-QoL (SMD of 0.2), so the result is compatible with important benefit (SMD 0.94 higher) and little or no difference (SMD 0.13 higher). However, the extent to which the threshold is crossed is modest (likely due to inconsistent effects) and both the upper and lower limit of the confidence interval favours the intervention, so we have not rated down for imprecision.

d. Publication bias strongly suspected (-1). Evidence from contour enhanced funnel plot that there could be missing studies which show effects favouring the control, especially nonsignificant effects (see Appendix D).

e. Very serious risk of bias (-2). All studies in analysis are at high risk of bias.

f. Serious inconsistency (-1). Heterogeneity statistics indicate inconsistency, although there are too few studies to interpret the statistics with confidence. Confidence intervals do not overlap for 2/3 studies, and the effects in these two studies is conflicting (little or no difference in one study; important benefit in the other).

g. Very serious imprecision (-2). The 95% confidence interval crosses the threshold for both a small but important improvement in HR-QoL (SMD 0.2) and a small but important reduction in HR-QoL (SMD -0.2), so the result is compatible with important harm (SMD -0.90 lower) and important benefit (SMD 1.57 higher). While the confidence interval is too wide to interpret the result, it is likely affected by inconsistency for which we have downgraded.

h. Serious imprecision (-1). The 95% confidence interval crosses the threshold for a small but important improvement in HR-QoL (SMD of 0.2), so the result is compatible with important benefit (SMD 1.11 higher) and little or no difference (SMD 0.06 higher).

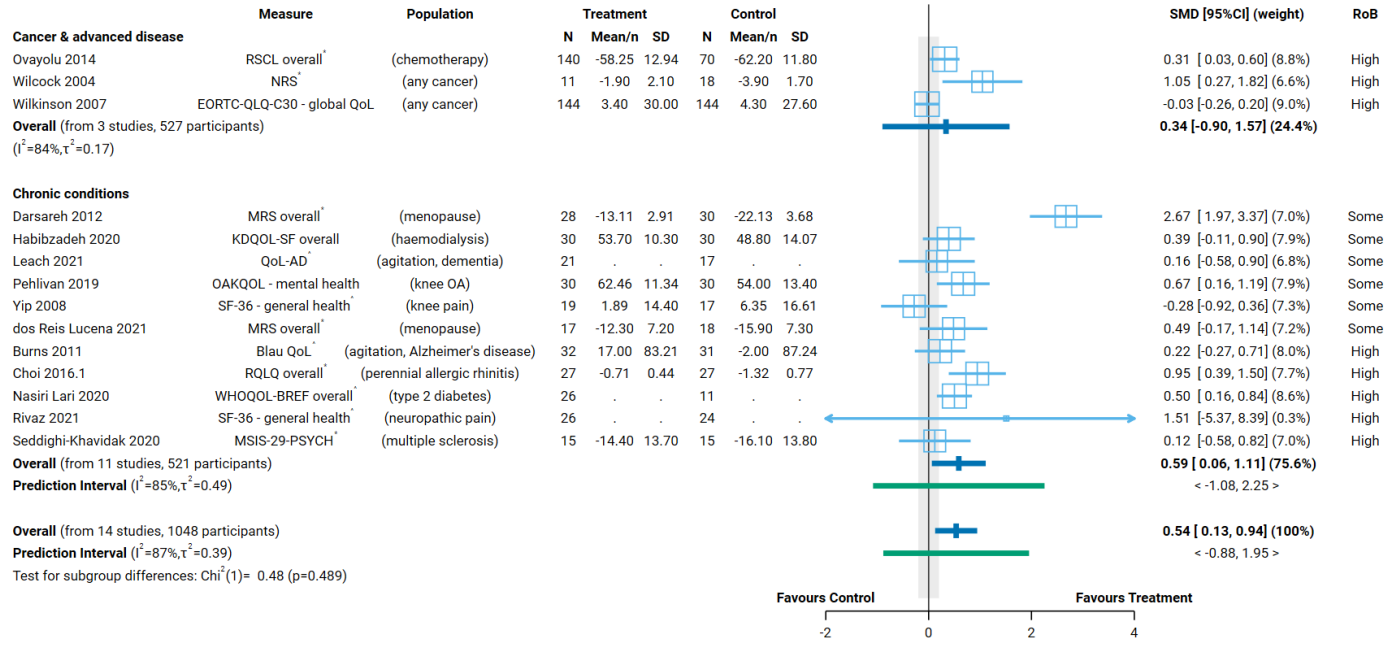


**Fig 4.7.1** | Forest plot for Comparison 1. the effect of aromatherapy (any mode) versus inactive control (usual care, no intervention, placebo) on health-related quality of life (HR-QoL). SMD=standardised mean difference. Blue lines show 95% confidence intervals (CI) and green lines show prediction intervals (PI). The shaded grey area indicates the pre-specified range where the effect of aromatherapy is considered to be no different from control (SMD -0.2 to 0.2 standard units). ^ indicates studies for which data transformation or imputation was required to include the result in the meta-analysis. This may include crossover trials and studies that reported results as a dichotomous or ordinal outcome (identifiable because no mean or SD is reported for the study in the forest plot). * Denotes studies for which the direction of effect was changed to match the overall plot (e.g. positive numbers are beneficial).

### Comparison 2. Aromatherapy (massage) versus massage

The evidence is uncertain about the effects of aromatherapy massage on health-related quality of life overall (all population groups) and for people living cancer and advanced disease. Aromatherapy massage may improve health-related quality of life for people living with chronic conditions.

Factors that reduced our certainty in the combined estimates of effect differed somewhat for each population group, as explained in the GRADE summary of findings table (Table 4.7.1, explanations). In combination, these factors raise concern that any observed benefit could be overestimated (or harm underestimated). Major concerns are as follows.

- ***Publication bias.*** There were too few studies to detect publication bias in this analysis; however, given the evidence from analyses with more studies, we cannot rule out the possibility that there could be studies (or results) missing from the analysis that show effects favouring the control.
- ***Risk of bias in included trials***. All trials in the analysis have methodological limitations (high risk of bias or some concerns). The absence of trials at low risk of bias meant that it was not possible to examine the impact of these methodological limitations on the estimate of the intervention effect using the approach specified in our protocol (limiting analyses to studies at low risk of bias).
- ***Inconsistent results that lead to different conclusions about the effects of aromatherapy***. There is evidence that the size of the intervention effect differs across studies beyond what would be expected by chance. There is some evidence that this may be partly explained by population group (Figure 4.7.1; as explained in Appendix D, Section D.6). Our confidence in the combined estimate was still reduced for analyses in which some studies found an important improvement in health-related quality of life (greater than the threshold for important benefit, an SMD of 0.2 or higher) and others found little or no difference between aromatherapy and control, with no credible evidence to explain whether this reflects true differences in the effects of aromatherapy or methodological problems in some studies.

Concerns relating to each finding were considered in the GRADE assessment when interpreting the result. The findings are as follows.

**Results for which an interpretation was made** (low certainty evidence)

***Chronic conditions***. Aromatherapy (massage) may improve health-related quality of life for people with chronic conditions. (SMD 0.53 higher, 95% CI 0.02 higher to 1.04 higher; $I^2$ = 82 %; 9 studies, 581 participants; low certainty Figure 4.7.2).

**Results considered too uncertain to interpret**

- ***Overall (all population groups).*** The evidence is very uncertain about the effect of aromatherapy (massage) on health-related quality of life. (12 studies, 851 participants; very low certainty, Figure 4.7.2).
- ***Cancer and advanced disease***. The evidence is very uncertain about the effect of aromatherapy (massage) on health-related quality of life for people living with cancer and advanced disease. (3 studies, 270 participants; very low certainty Figure 4.7.2).

**Table 4.7.2**. Summary of findings for Comparison 2. the effect of aromatherapy (massage) versus inactive massage control on health-related quality of life (HR-QoL).

| Outcomes | Anticipated absolute effects* (95% CI) | | Relative effect (95% CI) | № of participants (studies) | Certainty of the evidence (GRADE) | Comments |
|---|---|---|---|---|---|---|
| | Risk with inactive massage control | Risk with aromatherapy (massage) | | | | |
| Health-related quality of life: all populations | - | SMD **0.34 SD higher** (0.07 lower to 0.75 higher) | - | 851 (12 RCTs) | ⨁◯◯◯ Very low[a,b,c,d] | The evidence is very uncertain about aromatherapy (massage) on health-related quality of life. |
| Health-related quality of life: cancer and advanced disease | - | SMD **0.15 SD lower** (0.44 lower to 0.13 higher) | - | 270 (3 RCTs) | ⨁◯◯◯ Very low[d,e,f,g] | The evidence is very uncertain about aromatherapy (massage) on health-related quality of life for people living with cancer and advanced disease. |
| Health-related quality of life: chronic conditions | - | SMD **0.53 SD higher** (0.02 higher to 1.04 higher) | - | 581 (9 RCTs) | ⨁⨁◯◯ Low[a,d,h,i] | Aromatherapy (massage) may improve health-related quality of life for people with chronic conditions. |

***The risk in the intervention group** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI). **CI:** confidence interval; **SMD:** standardised mean difference

The threshold for an important difference was an SMD of 0.2 (used for interpreting point estimates and confidence intervals). For HR-QoL, the resulting interpretation is: > 0.2 is beneficial, -0.2 to 0.2 is trivial or unimportant ("little or no difference" between treatments), < -0.2 is harmful

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

## Explanations

a. Serious risk of bias (-1). All studies in the analysis are at high risk of bias or some concerns, such that the observed benefit may be overestimated.

b. Serious inconsistency (-1). Confidence intervals do not overlap for many studies, heterogeneity statistics indicate inconsistency, and the prediction interval indicates that the effect in a new study could range from important benefit to important harm. However, the point estimate for the majority of studies indicates important benefit (SMD > 0.2) or trivial effect (SMD -0.2 to 0.2) not harm. In addition, the subgroup analysis provides some evidence that population subgroup may explain some of the inconsistency (Appendix D for explanation). For this reason, we have downgraded for serious not very serious inconsistency.

c. Serious imprecision (-1). The 95% confidence interval crosses the threshold for a small but important improvement in HR-QoL (SMD of 0.2), which means the result is compatible with important benefit (SMD 0.75 higher) and little or no difference (SMD -0.07 lower).

d. Publication bias not detected. Too few studies in contour enhanced funnel plot to detect missing studies (see Appendix D).

e. Very serious risk of bias (-2). All studies in analysis are at high risk of bias.

f. No serious inconsistency. Confidence intervals overlap for all studies, suggesting that any variation in results across studies may be unimportant.

g. Serious imprecision (-1). The 95% confidence interval crosses the threshold for a small but important reduction in HR-QoL (SMD of -0.2), which means the result is compatible with important harm (SMD -0.44 lower) and little or no difference (SMD 0.13 higher).

h. No serious inconsistency. Heterogeneity statistics indicate inconsistency. However, confidence intervals overlap for most studies and the point estimate for the majority of studies (7/9) indicates important benefit (SMD > 0.2). The other two studies show a trivial effect (close to an SMD -0.2 to 0.2) not harm. For this reason, we have not downgraded for inconsistency.

i. Serious imprecision (-1). The 95% confidence interval crosses the threshold for a small but important improvement in HR-QoL (SMD of 0.2), which means the result is compatible with important benefit (SMD 1.04 higher) and little or no difference (SMD 0.02 higher).

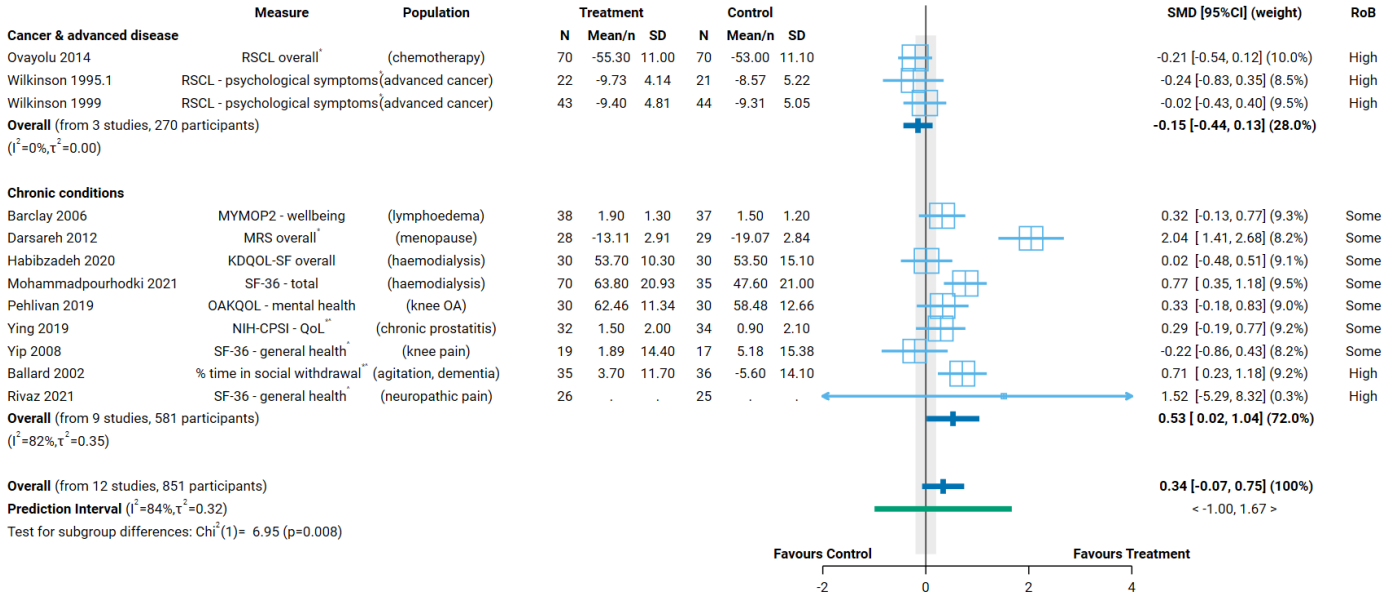


**Fig 4.7.2** | Forest plot for comparison 2. the effect of aromatherapy (massage) versus inactive massage control on health-related quality of life (HR-QoL). SMD=standardised mean difference. Blue lines show 95% confidence intervals (CI) and green lines show prediction intervals (PI). The shaded grey area indicates the pre-specified range where the effect of aromatherapy is considered to be no different from control (SMD -0.2 to 0.2 standard units). ^ indicates studies for which data transformation or imputation was required to include the result in the meta-analysis. This may include crossover trials and studies that reported results as a dichotomous or ordinal outcome (identifiable because no mean or SD is reported for the study in the forest plot). * Denotes studies for which the direction of effect was changed to match the overall plot (e.g. positive numbers are beneficial).

## 4.8 Physical function

Overall, thirteen studies that examined the effect of aromatherapy on physical function were included for meta-analysis, four of these contribute to both Comparison 1 and 2. A further two studies (238 participants) were eligible for one or both of the physical function meta-analyses, but could not be included (see below).

To be considered for the physical function analysis, trials had to administer aromatherapy to an eligible population for longer-term care (i.e. delivering treatment over weeks or longer, not days) and measure physical function in a time-frame likely to detect meaningful improvement (i.e. not immediately after a single treatment). No studies were excluded from the analysis on this basis.

***Comparison 1***.

- Ten studies (527 participants) contribute to the comparison of aromatherapy delivered by any mode compared to an inactive control (Figure 4.8.1).
- Two additional trials (238 participants) were eligible for this comparison, but did not report results that could be included in the meta-analysis.

***Comparison 2***.

- Seven studies (434 participants) contribute to the comparison of aromatherapy delivered by massage to an inactive massage control (comparable to massage in the aromatherapy arm) (Figure 4.8.2).
- An additional trial (118 participants) was eligible for this comparison, but did not report results that could be included in the meta-analysis.

### Characteristics of included studies

#### Types of populations

Included studies examined the effect of aromatherapy on physical function among people living with

- chronic musculoskeletal conditions (knee osteoarthritis/pain in six trials; neck pain and carpal tunnel syndrome, one trial each),
- cancer and advanced disease (people undergoing chemotherapy in one trial), and
- chronic conditions (neuropathic pain in two trials; chronic kidney disease and multiple sclerosis, one trial each).

The specific condition addressed in each trial is reported in the forest plot (column 3, Figure 4.8.1 and Figure 4.8.2) with full details for each study including eligibility criteria, participant characteristics, and ICD 11 codes in Appendix E1.

#### Types of interventions

***Comparison 1 (any mode of aromatherapy delivery).*** Of the 10 trials included in Comparison 1, aromatherapy was delivered by massage in 5 trials, topically in 3 trials and by inhalation in 2 trials. Lavender was the most commonly evaluated essential oil (5 trials), followed by ginger (3 trials, one in a blend with rosemary), nutmeg and chamomile (one trial each).

In the majority of trials involving people with chronic musculoskeletal conditions and other chronic conditions, aromatherapy was administered over a period of three weeks (6 of 9 trials; 4 weeks in 2 trials, 8 weeks in 1 trial). In the single trial involving people with cancer, aromatherapy was delivered continuously for five days after chemotherapy.

***Comparison 2 (aromatherapy delivered by massage)***. In Comparison 2, the essential oils evaluated were ginger (3 trials, one in a blend with rosemary), lavender (2 trials), orange (1 trial) and a blend of four oils (1 trial). The aromatherapy treatment period ranged from three to five weeks.

#### Types of outcomes

The outcome measure from which data were included for meta-analysis is reported for each trial in the forest plots (column 2, Figure 4.8.1 and Figure 4.8.2). Full details for each study are in Appendix E1, including the timing of outcome measurement in relation to intervention and details of which outcome was selected when multiple were available.

The Western Ontario McMaster Osteoarthritis Index (WOMAC) was used in all but one study among people with knee osteoarthritis/pain (5 of 6 trials, of which 4 reported results from the physical function scale). The SF-36 physical function scale was used in two trials (one on neuropathic pain, one on haemodialysis). Condition specific measures were used in other trials, for example the EORTC QLQ-C30 physical functioning scale in the trial involving people with cancer, the Neck Disability Index (NDI), the Boston Carpal Tunnel Syndrome Questionnaire (BCTQ), the Multiple Sclerosis Impact Scale (MSIS-29), and the Brief Pain Inventory for diabetic peripheral neuropathy (BPI-DPN).

All results were reported as a score on the original scale (e.g. physical function on the WOMAC).

### Effects of aromatherapy on physical function

**Comparison 1: Aromatherapy (any mode) versus inactive control (usual care, placebo, no intervention)**

Aromatherapy may improve physical function in general (all population groups) and for chronic musculoskeletal conditions (mainly knee osteoarthritis/pain), however the results are uncertain. The evidence is very uncertain about effects on physical function for people with cancer and advanced disease and with other chronic conditions.

Factors that reduced our certainty in the combined estimates of effect differed somewhat for each population group, as explained in the GRADE summary of findings table (Table 4.8.1, explanations). In combination, these factors raise concern that any observed benefit could be overestimated (or harm underestimated). Major concerns are as follows.

- ***Publication bias.*** There were too few studies to detect publication bias in this analysis; however, given the evidence for analyses with more studies, we cannot rule out the possibility that there could be studies (or results) missing from the analysis that show effects favouring the control.
- ***Risk of bias in included trials***. All trials in the analysis have methodological limitations (high risk of bias or some concerns). The absence of trials at low risk of bias meant that it was not possible to examine the impact of these methodological limitations on the estimate of the intervention effect using the approach specified in our protocol (limiting analyses to studies at low risk of bias).
- ***Inconsistent results that lead to different conclusions about the effects of aromatherapy***. There is evidence that the size of the intervention effect differs across studies beyond what would be expected by chance. These differences were not explained by population group (Figure 4.8.1; as explained in Appendix D, Section D.7). There is some evidence that the mode by which aromatherapy was delivered may partly explain the inconsistency but the results are inconclusive (Appendix D, Section D.7 and Figure D.7.1). Overall, the inconsistency reduced our confidence in the combined estimate because some studies found an improvement in physical function (greater than the threshold for important benefit, an SMD of 0.2 or higher) while others found little or no difference between aromatherapy and control, with no credible evidence to explain whether this reflects true differences in the effects of aromatherapy or methodological problems in some studies.

Concerns relating to each finding were considered in the GRADE assessment when interpreting the result. The findings are as follows.

**Results for which an interpretation was made** (low certainty evidence)

- ***Overall (all population groups).*** Aromatherapy (any mode) may improve physical function, (SMD 0.50 higher, 95% CI 0.15 higher to 0.85 higher; $I^2$ = 75 %; 10 studies, 527 participants; low certainty, Figure 4.8.1).
- ***Chronic musculoskeletal conditions***. Aromatherapy (any mode) may improve physical function among people with knee osteoarthritis, but the effects are very uncertain for other chronic musculoskeletal conditions (SMD 0.61 higher, 95% CI 0.15 higher to 1.07 higher; $I^2$ = 59 %; 6 studies, 313 participants, 265 with knee osteoarthritis; low certainty Figure 4.8.1).

**Results considered too uncertain to interpret**

- ***Cancer and advanced disease***. The evidence is very uncertain about the effect of aromatherapy (any mode) on physical function among people living with cancer and advanced disease (1 trial, 60 participants; very low certainty).
- ***Other chronic conditions.*** The evidence is very uncertain about the effect of aromatherapy (any mode) on physical function among people with other chronic conditions (3 studies, 154 participants; very low certainty).

No studies were included in this analysis among people with headache or migraine (chronic or episodic).

**Table 4.8.1**. Summary of findings for Comparison 1. the effect of aromatherapy (any mode) versus inactive control (usual care, no intervention, placebo) on physical function.

| Outcomes | Anticipated absolute effects* (95% CI) | | Relative effect (95% CI) | № of participants (studies) | Certainty of the evidence (GRADE) | Comments |
|---|---|---|---|---|---|---|
| | Risk with inactive control (usual care, placebo, no intervention) | Risk with aromatherapy (any mode) | | | | |
| Physical function: all population groups[a] | - | SMD **0.50 SD higher** (0.15 higher to 0.85 higher) | - | 527 (10 RCTs) | ⨁⨁◯◯ Low[b,c,d,e] | Aromatherapy (any mode) may improve on physical function. |
| Physical function: chronic musculoskeletal conditions (mainly knee OA; also knee pain, carpal tunnel syndrome)[f] | - | SMD **0.61 SD higher** (0.15 higher to 1.07 higher) | - | 313 (6 RCTs) | ⨁⨁◯◯ Low[e,g,h,i] | Aromatherapy (any mode) may improve physical function among people with knee osteoarthritis. The effects are very uncertain for other chronic musculoskeletal conditions. |
| Physical function: cancer and advanced disease (chemotherapy)[j] | - | SMD **0.01 SD lower** (0.16 lower to 0.15 higher) | - | 60 (1 RCT) | ⨁◯◯◯ Very low[e,k,l,m,n] | The evidence is very uncertain about the effect of aromatherapy (any mode) on physical function among people living with cancer and advanced disease. |
| Physical function: other chronic conditions (neuropathy, multiple sclerosis)[o] | - | SMD **0.61 SD higher** (1.23 lower to 2.46 higher) | - | 154 (3 RCTs) | ⨁◯◯◯ Very low[b,e,p,q] | The evidence is very uncertain about the effect of aromatherapy (any mode) on physical function among people with other chronic conditions. |
| Physical function: migraine or headache | | | | | No studies included in this analysis report effect of aromatherapy (any mode) for people with migraine or headache. | |

***The risk in the intervention group** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI). **CI:** confidence interval; **SMD:** standardised mean difference

The threshold for an important difference was an SMD of 0.2 (used for interpreting point estimates and confidence intervals). For physical function, the resulting interpretation is: > 0.2 is beneficial, -0.2 to 0.2 is trivial or unimportant ("little or no difference" between treatments), < -0.2 is harmful

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

## Explanations

a. Measures varied: WOMAC physical function, BCTQ - function, ISK overall, WOMAC overall, EORTC QLQ C30 - physical function, BPI-DPN walking, SF-36 physical function, MSIS-29-PHYS

b. Serious risk of bias (-1). All studies in the analysis are at high risk of bias or some concerns, such that the observed benefit may be overestimated.

c. Serious inconsistency (-1). Heterogeneity statistics suggest inconsistent results, however the confidence intervals overlap for most studies, and the majority of point estimate indicates important benefit (SMD of 0.2 or higher) or little or no difference, not harm. We have downgraded for serious not very serious inconsistency.

d. No serious imprecision. The 95% confidence interval crosses the threshold for a small but important improvement in function (SMD of 0.2), so the result is compatible with important benefit (SMD 0.85 higher) and little or no difference (SMD 0.15 higher). However, the extent to which the threshold is crossed is modest (likely due to inconsistent effects, which is rated down) and both the upper and lower limit of the confidence interval favours the intervention, so we have not rated down for imprecision.

e. Publication bias undetected. Evidence from sensitivity analysis and contour enhanced funnel plot for the overall analysis indicated that there could be missing studies which show effects favouring the control, but the number of studies is small so this could be due to reasons other than reporting bias (see Appendix D).

f. Measures varied: WOMAC physical function, BCTQ - function, ISK overall, WOMAC overall

g. Very serious risk of bias (-2). The majority of studies in analysis (most weight) are at high risk of bias.

h. No serious inconsistency. Heterogeneity statistics suggest inconsistent results, however the confidence intervals overlap for all studies, and all but one point estimate indicates important benefit (SMD of 0.2 or higher). For this reason we have not downgraded for inconsistency.

i. No serious imprecision. The 95% confidence interval crosses the threshold for a small but important improvement in function (SMD of 0.2), so the result is compatible with important benefit (SMD 1.07 higher) and little or no difference (SMD 0.15 higher). However, the extent to which the threshold is crossed is modest (likely due some inconsistency) and both the upper and lower limit of the confidence interval favours the intervention, so we have not rated down for imprecision.

j. Measure: EORTC QLQ C30 - physical function

k. Serious risk of bias (-1). Single study at high risk of bias; however there is little or no difference between treatments so downgraded for serious not very serious risk of bias.

l. Inconsistency not assessed: single study

m. Very serious indirectness (-2): Evidence from one small study among people receiving care during chemotherapy. Uncertain whether results apply to populations with cancer more generally.

n. No serious imprecision. Both the upper and lower limits of the 95% confidence interval (SMD -0.16 to 0.15) are compatible with little or no effect on function (SMD -0.2 to 0.2).

o. Measures varied: BPI-DPN walking, SF-36 physical function, MSIS-29-PHYS

p. Serious inconsistency (-1). Confidence intervals overlap marginally and effect estimates vary widely.

q. Very serious imprecision (-2). The 95% confidence interval crosses the threshold for both an important increase in function (SMD 0.2) and important reduction in function (SMD -0.2), so the result is compatible with important benefit (SMD 2.46 higher) and important harm (SMD 1.23 lower).

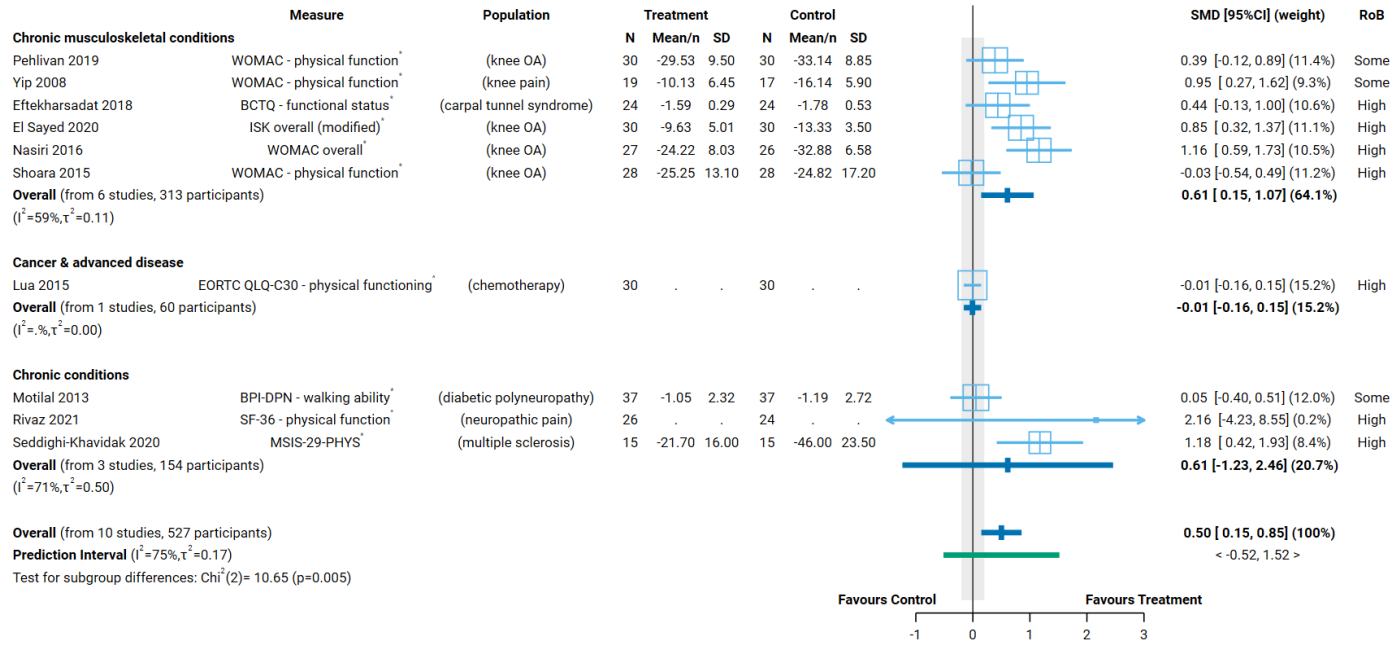


**Fig 4.8.1** | Forest plot for comparison 1. the effect of aromatherapy (any mode) versus inactive control (usual care, no intervention, placebo) on physical function. SMD=standardised mean difference. Blue lines show 95% confidence intervals (CI) and green lines show prediction intervals (PI). The shaded grey area indicates the pre-specified range where the effect of aromatherapy is considered to be no different from control (SMD -0.2 to 0.2 standard units). ^ indicates studies for which data transformation or imputation was required to include the result in the meta-analysis. This may include crossover trials and studies that reported results as a dichotomous or ordinal outcome (identifiable because no mean or SD is reported for the study in the forest plot). * Denotes studies for which the direction of effect was changed to match the overall plot (e.g. positive numbers are beneficial).

### Comparison 2. Aromatherapy (massage) versus massage

The evidence is very uncertain about effects of aromatherapy on physical function overall (any population) and for people with chronic musculoskeletal conditions and other chronic conditions.

Factors that reduced our certainty in the combined estimates of effect differed somewhat for each population group, as explained in the GRADE summary of findings table (Table 4.8.2, explanations). In combination, these factors raise concern that any observed benefit could be overestimated (or harm underestimated). Major concerns are as follows.

- ***Publication bias.*** There were too few studies to detect publication bias in this analysis; however, given the evidence for analyses with more studies, we cannot rule out the possibility that there could be studies (or results) missing from the analysis that show effects favouring the control.
- ***Risk of bias in included trials***. All trials in the analysis have methodological limitations (high risk of bias or some concerns). The absence of trials at low risk of bias meant that sensitivity analysis assessing risk of bias by looking at results only for low risk of bias were not conducted (as per the protocol).
- ***Inconsistent results that lead to different conclusions about the effects of aromatherapy***. There is evidence that the size of the intervention effect differs across studies beyond what would be expected by chance. These differences were not explained by population group (Figure 4.8.2; as explained in Appendix D, Section D.7). Overall, the inconsistency reduced our confidence in the combined estimate because some studies found an improvement in physical function (greater than the threshold for important benefit, an SMD of 0.2 or higher) while others found little or no difference between aromatherapy and control, with no credible evidence to explain whether this reflects true differences in the effects of aromatherapy or methodological problems in some studies.

Concerns relating to each finding were considered in the GRADE assessment when interpreting the result. The findings are as follows.

**Results for which an interpretation was made** (low certainty evidence)

**None**.

**Results considered too uncertain to interpret**

- ***Overall (all population groups).*** The evidence is very uncertain about the effect of aromatherapy (massage) on physical function (7 studies, 434 participants; very low certainty, Figure 4.8.2).
- ***Chronic musculoskeletal conditions***. The evidence is very uncertain about the effect of aromatherapy (massage) on physical function among people with chronic musculoskeletal conditions (5 studies, 278 participants; very low certainty Figure 4.8.2).
- ***Other chronic conditions.*** The evidence is very uncertain about the effect of aromatherapy (massage) on physical function among people with other chronic conditions (2 studies, 156 participants; very low certainty).

No studies were included in the analysis that examined the effect of aromatherapy massage on physical function among people living with cancer and advanced disease or headache and migraine.

**Table 4.8.2**. Summary of findings for Comparison 2. the effect of aromatherapy (massage) versus inactive massage control on physical function.

| Outcomes | Anticipated absolute effects* (95% CI) | | Relative effect (95% CI) | № of participants (studies) | Certainty of the evidence (GRADE) | Comments |
|---|---|---|---|---|---|---|
| | Risk with inactive control (massage) | Risk with aromatherapy (massage) | | | | |
| Physical function: all population groups[a] | - | SMD **0.45 SD higher** (0.09 higher to 0.80 higher) | - | 434 (7 RCTs) | ⨁◯◯◯ Very low[b,c,d,e] | The evidence is very uncertain about the effect of aromatherapy (massage) physical function. |
| Physical function: chronic musculoskeletal conditions (neck pain, knee OA, knee pain)[f] | - | SMD **0.39 SD higher** (0.12 lower to 0.91 higher) | - | 278 (5 RCTs) | ⨁◯◯◯ Very low[b,c,e,g] | The evidence is very uncertain about the effect of aromatherapy (massage) physical function among people with chronic musculoskeletal conditions. |
| Physical function: other chronic conditions (haemodialysis, neuropathic pain)[h] | - | SMD **0.63 SD higher** (0.60 lower to 1.86 higher) | - | 156 (2 RCTs) | ⨁◯◯◯ Very low[b,e,i,j] | The evidence is very uncertain about the effect of aromatherapy (massage) on physical function among people with other chronic conditions. |

| Outcomes | Anticipated absolute effects* (95% CI) | | Relative effect (95% CI) | № of participants (studies) | Certainty of the evidence (GRADE) | Comments |
|---|---|---|---|---|---|---|
| | **Risk with inactive control (massage)** | **Risk with aromatherapy (massage)** | | | | |
| Physical function: migraine or headache | | | | | | No studies included in this analysis report effect of aromatherapy (massage) for people with migraine or headache. |
| Physical function: cancer or advanced disease | | | | | | No studies included in this analysis report effect of aromatherapy (massage) among people living with cancer or advanced disease. |

***The risk in the intervention group** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI). **CI:** confidence interval; **SMD:** standardised mean difference

The threshold for an important difference was an SMD of 0.2 (used for interpreting point estimates and confidence intervals). For physical function, the resulting interpretation is: > 0.2 is beneficial, -0.2 to 0.2 is trivial or unimportant ("little or no difference" between treatments), < -0.2 is harmful

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

## Explanations

a. Measures varied: NDI, WOMAC - physical function, WOMAC - overall, SF-36 physical function.

b. Serious risk of bias (-1). All studies in the analysis are at high risk of bias or some concerns, such that the observed benefit may be overestimated.

c. Serious inconsistency (-1). Heterogeneity statistics suggest inconsistent results. The confidence intervals overlap for all studies, suggesting compatible results, however, in the overall analysis the point estimate indicates important benefit (SMD of 0.2 or higher) in 4/7 studies and little or no difference in 2/7 (the CI is too wide to interpret the last study). In the chronic MsK subgroup, 3/5 point estimates indicate benefit and 2/5 no difference. For this reason we have rated down for inconsistency.

d. Serious imprecision (-1). The 95% confidence interval crosses the threshold for a small but important improvement in function (SMD of 0.2), so the result is compatible with important benefit (SMD 0.80 higher) and little or no difference (SMD 0.09 higher).

e. Publication bias not detected. The number of studies in the analysis was too small to identify whether missing results are likely from the contour enhanced funnel plot (see Appendix D).

f. Measures varied: NDI, WOMAC - physical function, WOMAC - overall.

g. Serious imprecision (-1). The 95% confidence interval crosses the threshold for a small but important improvement in function (SMD of 0.2), so the result is compatible with important benefit (SMD 0.91 higher) and little or no difference (SMD -0.12 lower).

h. Measures: SF-36 physical function.

i. No serious inconsistency. Heterogeneity statistics do not suggest inconsistent results, the confidence intervals overlap completely, and the point estimate in both studies indicates important benefit (SMD of 0.2 or higher).

j. Extremely serious imprecision (-3). The 95% confidence interval crosses the threshold for both an important increase in function (SMD 0.2) and important reduction in function (SMD -0.2), so the result is compatible with important harm (SMD 0.60 lower) and important benefit (SMD 1.86 higher). The confidence interval is too wide to interpret the result.

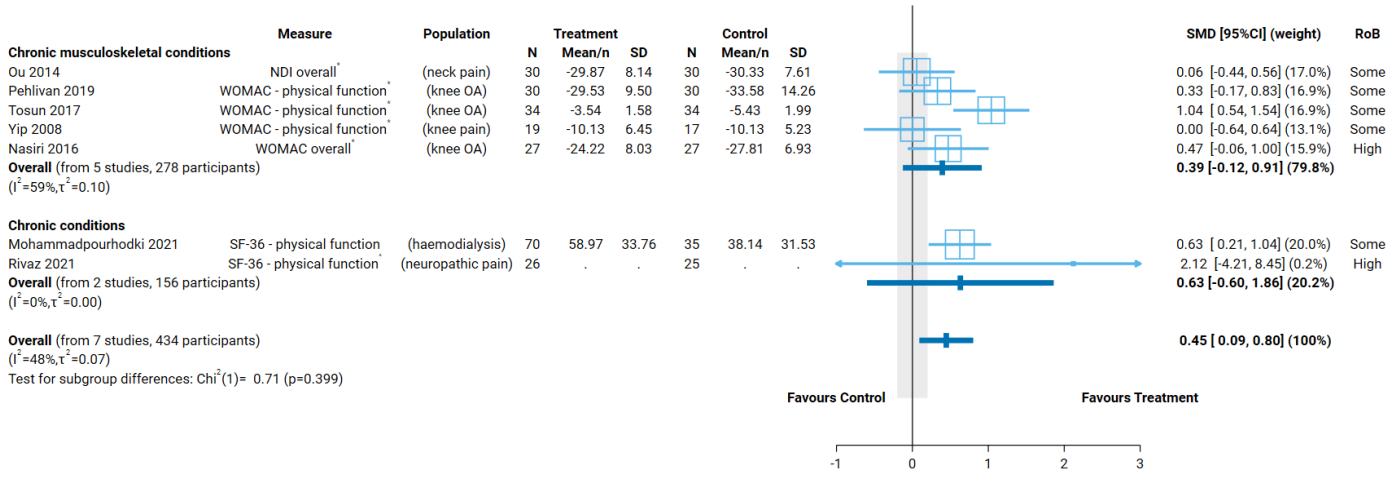


**Fig 4.8.2** | Forest plot for comparison 2. the effect of aromatherapy (massage) versus inactive massage control on physical function. SMD=standardised mean difference. Blue lines show 95% confidence intervals (CI) and green lines show prediction intervals (PI). The shaded grey area indicates the pre-specified range where the effect of aromatherapy is considered to be no different from control (SMD -0.2 to 0.2 standard units). ^ indicates studies for which data transformation or imputation was required to include the result in the meta-analysis. This may include crossover trials and studies that reported results as a dichotomous or ordinal outcome (identifiable because no mean or SD is reported for the study in the forest plot). * Denotes studies for which the direction of effect was changed to match the overall plot (e.g. positive numbers are beneficial).

# 5. Discussion

## Summary of the main results

This review was limited to assessment of the evidence for certain conditions and groups of people to inform the Australian Government about health policy decisions for private health insurance rebates. This review was not designed to assess all the reasons that people use aromatherapy, or the reasons practitioners prescribe aromatherapy and was not intended to inform individual choices about using aromatherapy.

This systematic review included a large body of research from trials of aromatherapy. For each outcome, we examined the effects of aromatherapy overall (across multiple conditions) and for specific population groups. This approach makes use of the body of evidence to evaluate whether there is evidence that aromatherapy works across multiple population groups or whether any effects might be limited to specific population groups.

Of the 323 studies included in the review, 234 were eligible for at least one synthesis (the remaining 89 studies were included in the evidence inventory). Across all syntheses, we were able to include data from 201 studies. The largest syntheses examined the effect of aromatherapy delivered by any mode on pain (7193 participants in 82 trials) and emotional functioning and mental health (7032 participants in 86 trials). Only two of the seven analyses for this comparison had fewer than 1000 participants (physical function, 527 participants; fatigue, 795 participants).

We found that across multiple conditions, compared to an inactive control (placebo, no intervention, usual care), aromatherapy (delivered by inhalation, massage, or topically):

- may improve sleep quality (low certainty evidence; no trials among people living with dementia and behaviour change),
- may improve health-related quality of life (low certainty evidence),
- may improve physical function (low certainty evidence).

For other outcomes (pain, nausea and vomiting, fatigue, emotional functioning and mental health) we are uncertain about the effects of aromatherapy across multiple conditions. For these outcomes, the effects varied importantly across studies; some studies showed benefit, others showed little or no effect on the outcome. These inconsistent effects were not explained by differences in the population receiving aromatherapy nor by the way in which aromatherapy was delivered (mode of delivery). However, the evidence was considered more certain for some population groups as follows.

There was low certainty that aromatherapy may improve

- pain among people with chronic musculoskeletal conditions,
- acute or episodic pain conditions (mainly dysmenorrhea)
- nausea and vomiting during pregnancy,
- sleep during hospitalisation (not surgery, mainly cardiovascular)
- mental health among people with symptoms of mental distress,
- physical function among people with chronic musculoskeletal conditions.

There was also low certainty evidence that aromatherapy may have little or no effect on

- mental distress among people living with cancer or advanced disease
- mental health among people living with dementia (mainly agitation)

Fewer studies compared aromatherapy massage to an inactive massage control (comparable to that used in the aromatherapy arm), with no studies on nausea and vomiting or sleep quality. This comparison was included to isolate the effect of aromatherapy from those of massage. There was low certainty evidence that health-related quality of life improved with aromatherapy massage, but it was uncertain whether there was benefit or little or no effect on other outcomes. Consistent with this finding, our exploratory (subgroup) analyses did not provide credible evidence that delivering aromatherapy by massage had larger effects than by inhalation or topically.

### Comparability of these findings with other systematic reviews

A 2019 mapping review examined the effects of aromatherapy on health outcomes [51]. The scope of the mapping review was similar to our review, but the evidence came from existing systematic reviews, so many questions were not covered and the evidence was less up-to-date. Consistent with our findings, the authors found evidence that aromatherapy inhalation (including massage) improved sleep quality in various populations (low confidence evidence), reduced pain during labour and childbirth (moderate confidence), had no effect on anxiety in palliative care (moderate confidence), and reduced pain in dysmenorrhea (moderate confidence). Our rating of certainty was low for these outcomes. This is unsurprising because existing systematic reviews rarely provide complete information to assess certainty (the authors state "our measure of the level of confidence cannot approach the rigor represented by standardized approaches", p6, [51]. For example, the authors did not examine the impact of publication bias, which is an important difference because concerns about publication reduced our certainty in many findings (e.g. all results for pain) and the body of evidence overall. The mapping review also found 'unclear/insufficient' evidence about effects on pain and psychological outcomes in other populations (including perioperative), nausea and vomiting (all studied populations), and 'global symptoms' (encompassing HR-QoL, fatigue, physical function). There was no evidence for many populations we considered because they had not been covered by existing systematic reviews.

## Overall completeness and applicability of evidence

There is an extensive body of evidence examining the effects of aromatherapy on health outcomes, in particular on pain and emotional functioning and mental health. Included studies addressed outcomes or conditions identified in the PRACI survey as most often treated in Australia (i.e. stress and mental health, musculoskeletal condition associated with chronic pain, sleep disruption, and cancer and palliative care). Headache and migraine and sports injury were exceptions. The evidence is dominated by studies evaluating aromatherapy for acute indications such as control of pain, anxiety and other outcomes perioperatively (46 of 201 trials in the analysis), periprocedurally (45 trials), and during labour and childbirth (11 trials). Fewer studies have examined the effects of aromatherapy among people living with chronic or life-limiting conditions, for whom the use of aromatherapy for supportive care is of interest [13, 31, 35]. Of the 201 trials included in analyses, 30 were among people with chronic conditions (covering a diversity of conditions), 16 were among people living with cancer or advanced disease, and nine among people living with dementia and behaviour change.

The vast majority of studies included in the analysis were conducted in Iran (91 of 201 trials), followed by Turkey (37 of 201), the United States of America (15 trials) and the United Kingdom (11 trials). Other countries in which multiple trials were conducted included Taiwan (7 trials), Korea (7 trials), China (4 trials) and Japan (3 trials). Many of the trials in Iran and Turkey were in hospital settings.

## Certainty of the evidence

The certainty of evidence was considered when interpreting each result by applying the GRADE approach. Despite the large body of trials research on the effects of aromatherapy, the evidence arising from this review is of low or very low certainty for all results. Overarching concerns that reduce confidence in all findings arise from methodological limitations of included trials (for all 201 studies in the analysis there was either a high risk of bias or some concerns), missing results (evidence that results may be missing for studies for which results favoured the control), and inconsistent results across studies (some showing benefit, others showing little or no effect). Additional concerns applied to many findings. Methodological limitations of the included studies included a risk of bias arising from the randomised process, unblinded outcome assessment; and selection of the reported results. Many of these limitations in the conduct and reporting of trials were preventable.

In addition to factors addressed in the GRADE assessment, there were major problems with the quality of reporting in the included studies. Incomplete and ambiguous reporting affected our ability to understand the study design and confirm design features related to bias. It also led to exclusion of a large amount of data from the analyses; 41 trials (4415 participants) had data missing from at least one analysis for which it was eligible. In a high proportion of these studies, the problems with the data were so concerning that the results were not trusted. We chose not to report or synthesise results for studies that could not be included in meta-analyses.

### Potential biases in the review process

In this review we applied methods recommended in the Cochrane handbook for systematic reviews of interventions and the GRADE approach, as per the detailed protocol that was prospectively registered on PROSPERO after undergoing independent methodological review. The populations and outcomes eligible for the synthesis were finalised after studies were identified for inclusion in the review. To minimise bias in this process, a pre-specified prioritisation process was implemented in which NTWC, with input from NTREAP, prioritised the populations and outcomes eligible for the review without knowledge of the included studies or results of those studies. An initial analytic framework for the synthesis was included in the protocol to inform these decisions, which provided an a priori rationale for the final synthesis questions, criteria and structure.

While data extraction for each study was performed by a single reviewer, the selection of outcomes and coding of studies for inclusion in meta-analyses was performed independently by a second experienced review author. All data was checked by a second experienced author, with input from a biostatistician, and all data manipulation and analyses performed by a biostatistician. These steps minimised the risk of errors or misinterpretation. Risk of bias assessments were performed for each study by a single reviewer following detailed guidance developed for the review and training in the assessment of design features relevant to this review (to promote consistency between different reviewers).

While we endeavoured to include all available studies in the analyses (including any outcome measure and applying all suggested methods from the Cochrane handbook to included data), many studies reported data from which the required statistics could not be calculated or imputed, or presented results that could not be interpreted. The large number of studies in the review meant it was not feasible to contact trialists for additional information, nor was it possible to review trial registry entries to conduct a thorough assessment of missing results from the synthesis. For most analyses, this did not lessen our certainty in the evidence because we were able to examine and address the impact of missing results in our GRADE assessment through other methods (contour enhanced funnel plots, sensitivity analyses).

Finally, we screened and reported citations for studies in languages other than English but did not include these studies in the synthesis (as per protocol). There is no reason to expect that the results of these studies would differ systematically from those reported in English and, in turn, that exclusion of these studies would bias the results of the review. Given the amount of data contributing to most analyses, addition of these studies is unlikely to change the review conclusions.

# 6. Conclusions

There is a large and growing body of evidence examining the effects of aromatherapy on health. Despite this, it is not possible to draw conclusions about the effects of aromatherapy with confidence for any condition or outcome. Unlike many reviews, this is not due to a lack of evidence from randomised trials, with all but two of the syntheses for the first comparison containing data from over 1000 participants (i.e. the precision of effect estimates is not an overriding concern). Instead, the uncertainty reflects significant methodological problems with the evidence base. Although an interpretation is made for many of the results from meta-analyses, the evidence is of low or very low certainty, meaning that the true effects of aromatherapy may be substantially different from the estimated effects. Many factors contribute to this uncertainty. Of greatest concern is that results that show large beneficial effects from aromatherapy (beyond what would be seen for many first line therapies) may have been published selectively, while results that show little or no effect are not reported. Together with biases in the conduct of studies (e.g. bias arising from unblinded outcome assessment), this may be one of the underlying reasons for the inconsistent results observed across studies. In addition, the absence of any studies at low risk of bias means it is not possible to examine the impact that bias in the included studies has on the results.

### Implications for health policy

The evidence is of low or very low certainty for all outcomes and populations considered in this review. This means that our confidence in the estimate of effect for each outcome is limited, and the true effect may be substantially different. Major concerns about inconsistent results (some studies showing benefit, and others little

or no effect) without a credible explanation, and the likelihood that results that show large beneficial effects from aromatherapy may have been selectively published by trialists, should be considered when deciding whether there is any credible evidence to support the use of aromatherapy.

## Implications for future research

Given the extent of concerns about bias in included studies and bias due to missing results (reporting bias), it is unlikely that systematic reviews will be able to answer questions about the effects of aromatherapy with any certainty by building on the very large body of existing evidence. Although a thorough investigation of the integrity of existing research in this field may provide evidence about the extent of reporting bias, our examination of trial registry entries suggests that there may not be sufficient information to conduct these studies using methods proposed for research-on-research integrity. Improving the conduct and, at a minimum the reporting, of trials in this field is an imperative. Any future trials must address preventable limitations in the conduct and reporting of trials of aromatherapy (including, but not limited to, bias arising from the randomised process, the method of outcome assessment; and the reporting of results). Adhering to relevant reporting guidelines such as CONSORT (https://www.equator-network.org/reporting-guidelines/consort/), and addressing errors that rendered results unusable is essential. The value of conducting more trials on aromatherapy would need to be carefully assessed to avoid further research waste.

# 7. Author contributions and declaration of interest

| | |
|---|---|
| Sue Brennan[1*]<br><br>sue.brennan@monash.edu<br>*(contact author) | Senior Evidence Officer responsible for leading the review. Led the design of the review and data extraction systems, and the implementation of risk of bias assessment. Wrote the review report with contributions from other authors as described. |
| Max Murano[1] | Implemented and managed electronic systems for screening studies and data extraction, and associated work processes. Managed and coordinated study selection and data extraction (including training data extractors), selected studies, and performed data checking of extracted studies and cleaned data. Prepared material for the report and technical appendices, and contributed to the writing of methods and results in Appendix A. |
| Simon Turner[2] | Provided advice on extraction of results data, prepared the data set for meta-analysis (including transformations and manipulations required to include results in analysis), conducted all meta-analyses (including sensitivity and subgroup analyses), prepared figures and results tables for the report. Documented analysis methods. |
| Steve McDonald[1] | Developed, wrote and implemented the search strategy. Screened studies for inclusion in the review and piloted data collection and risk of bias methods. Prepared material for the report and technical appendices. Wrote the search methods and results, and study selection. |
| Anneliese Synnot[1] | Selected, appraised and extracted data from studies. Contributed to data checking. |
| Phoebe Nguyen[2] | Appraised and extracted data from studies. |
| Kimberley Jones[3] | Appraised and extracted data from studies. |
| Isabella Freijah[3] | Appraised and extracted data from studies. |
| Joanne McKenzie[2] | Wrote the analysis plan and method for reporting treatment effects. Wrote the section on *Assessment of biases due to missing results*. Designed the data collection form for quantitative results data. Provided statistical advice on risk of bias assessment, data extraction/transformation/manipulations and interpretation. Provided oversight for the conduct and interpretation of the analysis. Reanalysed data from crossover and cluster trials. |

[1] Cochrane Australia, Evidence Synthesis Qualitative & Implementation Methods Unit, School of Public Health and Preventive Medicine, Monash University

[2] Methods in Evidence Synthesis Unit, School of Public Health and Preventive Medicine, Monash University

[3] Indigenous Health Equity Unit, Melbourne School of Population and Global Health, University of Melbourne

## Declarations of interest

All authors declare they have no financial, personal or professional interests that could be construed to influence the conduct or results of this systematic review.

## Funding

This review is funded by the Australian Government Department of Health and Aged Care via the NHMRC, under Official Order 2020-21P030.

# 8. References


1. Brennan SE, McDonald S, Murano M, McKenzie JE: **Effectiveness of aromatherapy for prevention or treatment of disease, medical or preclinical conditions, and injury: protocol for a systematic review and meta-analysis.** *Systematic Reviews* 2022, **11:**148.
2. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page M, Welch V (Eds.): **Cochrane Handbook for Systematic Reviews of Interventions version 6.1 (updated September 2020). Cochrane. Available from www.training.cochrane.org/handbook**; 2020.
3. Brennan SE, Johnston RV: **Research Note: Interpreting findings of a systematic review using GRADE methods.** *Journal of Physiotherapy* 2023.
4. Schünemann HJ, Higgins J, Vist GE, Glasziou P, Akl EA, Skoetz N, Guyatt GH, On behalf of the Cochrane GRADEing Methods Group (formerly Applicability and Recommendations Methods Group) and the Cochrane Statistical Methods Group: **Chapter 14: Completing 'Summary of findings' tables and grading the certainty of the evidence.** In *Cochrane Handbook for Systematic Reviews of Interventions.* 2nd edition. Edited by Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Welch V. Chichester (UK): John Wiley & Sons; 2019
5. Schünemann HJ, Vist GE, Higgins J, Santesso N, Deeks JJ, Glasziou P, Akl EA, Guyatt GH, On behalf of the Cochrane GRADEing Methods Group: **Chapter 15: Interpreting results and drawing conclusions.** In *Cochrane Handbook for Systematic Reviews of Interventions.* 2nd edition. Edited by Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Welch V. Chichester (UK): John Wiley & Sons; 2019
6. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, et al: **The PRISMA 2020 statement: an updated guideline for reporting systematic reviews.** *BMJ* 2021, **372:**n71.
7. Page MJ, Moher D, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, et al: **PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews.** *BMJ* 2021, **372:**n160.
8. National Health and Medical Research Council: **Statement of requirement: Evidence evaluations for review of natural therapies (tranche two).** 2020.
9. Posadzki P, Watson LK, Alotaibi A, Ernst E: **Prevalence of use of complementary and alternative medicine (CAM) by patients/consumers in the UK: systematic review of surveys.** *Clin Med (Lond)* 2013, **13:**126-131.
10. Harnett JE, McIntyre E, Steel A, Foley H, Sibbritt D, Adams J: **Use of complementary medicine products: a nationally representative cross-sectional survey of 2019 Australian adults.** *BMJ Open* 2019, **9:**e024198.
11. Steel A, McIntyre E, Harnett J, Foley H, Adams J, Sibbritt D, Wardle J, Frawley J: **Complementary medicine use in the Australian population: Results of a nationally-representative cross-sectional survey.** *Sci Rep* 2018, **8:**17325.
12. National Health and Medical Research Council: **Aromatherapy description developed in conversation with the National Health and Medical Research Council's Natural Therapies Working Committee Chair and the Department of Health's Natural Therapies Review Expert Advisory Panel (February 2020).** 2020.
13. PDQ Integrative Alternative Complementary Therapies Editorial Board: **Aromatherapy With Essential Oils (PDQ®): Health Professional Version.** In *PDQ Cancer Information Summaries*. Bethesda (MD): National Cancer Institute (US); 2019.
14. International Federation of Professional Aromatherapists: **International Federation of Professional Aromatherapists (IFPA). Available from: http://www.ifparoma.org/ Access date: 4 February 2021.** 2021.
15. Canadian Federation of Aromatherapists (CFA): **About Canadian Federation of Aromatherapists (CFA). Available from: http://www.cfacanada.com/ Access date: 4 February 2021.** 2021.
16. International Aromatherapy and Aromatic Medicine Association: **About Aromatherapy. Available from https://www.iaama.org.au/about-aromatherapy.html Access date: 4 February 2021.** 2021.
17. National Association for Holistic Aromatherapy: **About National Association for Holistic Aromatherapy (NAHA). Available from: https://naha.org Access date: 4 February 2021.** 2021.
18. Tisserand R, Young R: **Essential oil safety: a guide for health care professionals.** 2nd edition. pp. 795. Edinburgh: Elsevier Limited; 2014:795.
19. Orchard A, van Vuuren SF: **Carrier oils in dermatology.** *Arch Dermatol Res* 2019, **311:**653-672.
20. Australian Health Practitioner Regulatory Association: **What we do. Available from: https://www.ahpra.gov.au/ Access date: 4 February 2021.** 2021.

21. Steel A, Leach M, Wardle J, Sibbritt D, Schloss J, Diezel H, Adams J: **The Australian Complementary Medicine Workforce: A Profile of 1,306 Practitioners from the PRACI Study.** *J Altern Complement Med* 2018, **24:**385-394.
22. International Aromatherapy and Aromatic Medicine Association: **About the International Aromatherapy and Aromatic Medicine Association (IAAMA). Available from https://www.iaama.org.au Access date: 4 February 2021.** 2021.
23. Australian Traditional Medicine Society: **About the Australian Traditional Medicine Society (ATMS). Available from: https://www.atms.com.au/about-us Access date: 4 February 2021.** 2021.
24. Australian Government Department of Health: **Therapeutic Goods Administration: An overview of the regulation of complementary medicines in Australia. Available from: Access date: 4 February 2021.** 2013.
25. Sanger GJ, Andrews PLR: **A History of Drug Discovery for Treatment of Nausea and Vomiting and the Implications for Future Research.** *Frontiers in pharmacology* 2018, **9:**913-913.
26. Koyama S, Heinbockel T: **The Effects of Essential Oils and Terpenes in Relation to Their Routes of Intake and Application.** *International journal of molecular sciences* 2020, **21:**1558.
27. Block E: **What's that smell? A controversial theory of olfaction deemed implausible.** In *The Conversation*; 2015.
28. Orchard A, van Vuuren S: **Commercial Essential Oils as Potential Antimicrobials to Treat Skin Diseases.** *Evid Based Complement Alternat Med* 2017, **2017:**4517971.
29. Peterfalvi A, Miko E, Nagy T, Reger B, Simon D, Miseta A, Czéh B, Szereday L: **Much More Than a Pleasant Scent: A Review on Essential Oils Supporting the Immune System.** *Molecules* 2019, **24:**4530.
30. Ball EL, Owen-Booth B, Gray A, Shenkin SD, Hewitt J, McCleery J: **Aromatherapy for dementia.** *Cochrane Database of Systematic Reviews* 2020.
31. Candy B, Armstrong M, Flemming K, Kupeli N, Stone P, Vickerstaff V, Wilkinson S: **The effectiveness of aromatherapy, massage and reflexology in people with palliative care needs: A systematic review.** *Palliat Med* 2020, **34:**179-194.
32. Hines S, Steels E, Chang A, Gibbons K: **Aromatherapy for treatment of postoperative nausea and vomiting.** *Cochrane Database of Systematic Reviews* 2018.
33. Stea S, Beraudi A, De Pasquale D: **Essential oils for complementary treatment of surgical patients: state of the art.** *Evid Based Complement Alternat Med* 2014, **2014:**726341.
34. Steel A: **PRACI Study: unpublished data.** 2021.
35. Armstrong M, Flemming K, Kupeli N, Stone P, Wilkinson S, Candy B: **Aromatherapy, massage and reflexology: A systematic review and thematic synthesis of the perspectives from people with palliative care needs.** *Palliat Med* 2019, **33:**757-769.
36. Steel A, Schloss J, Diezel H, Palmgren PJ, Maret JB, Filbet M: **Complementary medicine visits by palliative care patients: a cross-sectional survey.** *BMJ Support Palliat Care* 2020.
37. Smith CA, Collins CT, Crowther CA: **Aromatherapy for pain management in labour.** *Cochrane Database of Systematic Reviews* 2011.
38. Johnson JR, Rivard RL, Griffin KH, Kolste AK, Joswiak D, Kinney ME, Dusek JA: **The effectiveness of nurse-delivered aromatherapy in an acute care setting.** *Complement Ther Med* 2016, **25:**164-169.
39. Steel A, Schloss J, Leach M, Adams J: **The naturopathic profession in Australia: A secondary analysis of the Practitioner Research and Collaboration Initiative (PRACI).** *Complement Ther Clin Pract* 2020, **40:**101220.
40. Green S, Hill M, Kim M, Kramer S, McDonald S, McKenzie J, Marion V, Wasiak J: **Evaluation of evidence about the effectiveness of aromatherapy: an overview of systematic reviews. Report prepared for the National Health and Medical Research Council by Cochrane Australia. 20 February.** Melbourne, Australia: Monash University; 2014.
41. Koo M: **A bibliometric analysis of two decades of aromatherapy research.** *BMC Res Notes* 2017, **10:**46.
42. Higgins JPT, Savovic J, Page MJ, Sterne JAC: **Revised Cochrane risk-of-bias tool for randomized trials (RoB 2). Available from https://www.riskofbias.info/welcome/rob-2-0-tool/current-version-of-rob-2. Access date: 8 February 2021.** 2019.
43. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, McLeod L, Delacqua G, Delacqua F, Kirby J, Duda SN: **The REDCap consortium: Building an international community of software platform partners.** *J Biomed Inform* 2019, **95:**103208.

44. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG: **Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support.** *J Biomed Inform* 2009, **42:**377-381.
45. Sterne JAC, Savovic J, Page MJ, Elbers RG, Blencowe NS, Boutron I, Cates CJ, Cheng HY, Corbett MS, Eldridge SM, et al: **RoB 2: a revised tool for assessing risk of bias in randomised trials.** *BMJ* 2019, **366:**l4898.
46. Higgins JPT, Eldridge SM, Li T: **Chapter 23: Including variants on randomized trials. Cochrane Handbook for Systematic Reviews of Interventions version 6.1 (updated September 2020). Cochrane. Available from www.training.cochrane.org/handbook.** (Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page M, Welch V eds.); 2020.
47. Zeng L, Brignardello-Petersen R, Hultcrantz M, Siemieniuk RAC, Santesso N, Traversy G, Izcovich A, Sadeghirad B, Alexander PE, Devji T, et al: **GRADE guidelines 32: GRADE offers guidance on choosing targets of GRADE certainty of evidence ratings.** *J Clin Epidemiol* 2021, **137:**163-175.
48. Schunemann HJ, Brozek J, Guyatt G, Oxman AD (Eds.): **Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach**. Hamilton, Canada: McMaster University; 2013.
49. Santesso N, Carrasco-Labra A, Langendam M, Brignardello-Petersen R, Mustafa RA, Heus P, Lasserson T, Opiyo N, Kunnamo I, Sinclair D, et al: **Improving GRADE Evidence Tables part 3: Guidance for useful GRADE certainty in the evidence judgments through explanatory footnotes.** *J Clin Epidemiol* 2016.
50. Santesso N, Glenton C, Dahm P, Garner P, Akl EA, Alper B, Brignardello-Petersen R, Carrasco-Labra A, De Beer H, Hultcrantz M, et al: **GRADE guidelines: 26. informative statements to communicate the findings of systematic reviews of interventions.** *Journal of Clinical Epidemiology* 2019.
51. Freeman M, Ayers C, Peterson C, Kansagara D: **Aromatherapy and Essential Oils: A Map of the Evidence. VA ESP Project #05-225. Available at: https://www.hsrd.research.va.gov/publications/esp/reports.cfm.** Washington, DC: Evidence Synthesis Program, Health Services Research and Development Service, Office of Research and Development, Department of Veterans Affairs; 2019.