Cochrane Australia

Systematic review of evidence on the clinical effectiveness of reflexology

Report prepared by

Cochrane Australia

14 November 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). This report was endorsed by NTWC on 20 November 2024.

Copyright

Attribution

Use of images

Contact

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 reflexology in preventing and/or treating injury, disease, medical conditions or preclinical conditions. Reflexology is a non-invasive touch-based therapy in which practitioners use their hands, thumbs and fingers to apply pressure to specific reflex points on the extremities (feet, hands, outer ears, face and lower legs) that are believed to stimulate a response in a corresponding organ or system in the body.
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 reflexology, 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 reflexology, nor is it intended to inform decisions about whether an individual or practitioner should use reflexology.

Key messages

- There is a large body of evidence examining the effects of reflexology on health. Despite this, it is not possible to draw conclusions about the effects of reflexology with confidence for any condition or outcome.
- The evidence is of low certainty for four (4) of the outcomes examined in the review and of very low certainty for all other outcomes (34 of 38), meaning that the true effect of reflexology may be substantially different.
- We are uncertain about the effects of reflexology 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 reflexology 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 reflexology 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 reflexology on outcome domains:

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

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 reflexology 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 reflexology 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 [1] 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 CRD42023394291) that underwent independent review by methods specialists, the Department's panel (NTREAP) and was endorsed by NHMRC's Natural Therapies Working Committee (NTWC) [2]. The review is reported in accordance with the PRISMA 2020 statement [3, 4].

What were the main results of the review?

Following screening of 1239 records, we retrieved 386 full-text reports from which we included 174 studies in the review (123 for the synthesis and 51 in the evidence inventory), of which 101 studies contribute results to at least one synthesis of evidence. Twenty-two (22) of the 123 studies did not contribute to any of the meta-analyses for which they were eligible because the outcome data were not reported, were reported incompletely or were uninterpretable. For similar reasons, a further 7 studies were missing from at least one analysis for which an outcome had been measured. In total, 47 out of 210 eligible results (22%) could not be included in the synthesis.

Across multiple conditions and compared to an inactive control (placebo, no intervention, usual care), the evidence provides low certainty that reflexology may improve sleep quality (12 trials, 782 participants).

For pain, fatigue, emotional functioning and mental health, health related quality of life, physical function and global symptoms 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.

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

There was low certainty that reflexology may:

- improve sleep quality for people with symptoms of sleep disruption (6 trials, 376 participants),
- reduce fatigue for people with certain chronic conditions (8 trials, 535 participants),
- improve health-related quality of life for people with certain chronic and longer-term conditions (12 trials, 777 participants).

Effects were very uncertain for:

- pain for surgery, procedures, labour and childbirth, other acute pain, cancer or advanced disease, chronic musculoskeletal conditions or other chronic pain,
- sleep quality for surgery, hospitalisation (not procedures) or cancer or advanced disease,
- fatigue for cancer or advanced disease, chronic musculoskeletal conditions or pregnancy,
- emotional functioning and mental health for surgery, procedures, hospitalisation (not procedures), labour or childbirth, mental distress, cancer or advanced disease, mental disorders, dementia,
- health related quality of life for cancer or advanced disease,
- physical function for cancer or advanced disease, chronic musculoskeletal conditions or other chronic conditions
- global symptoms for cancer or advanced disease, other chronic conditions or chronic respiratory conditions

No studies were found comparing reflexology to inactive controls for pain for people with acute musculoskeletal conditions or migraine / headache, sleep for people with chronic insomnia or dementia, or physical function for people with migraine or headache (chronic or episodic).

The effects of reflexology compared to other active comparators was not examined, as pre-specified criteria for synthesis were not met (i.e. no two studies at low risk of bias evaluated the same evidence-based treatment). Studies that only contributed active comparators are listed in an inventory (Appendix C3 and E3).

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 reflexology, or the reasons practitioners prescribe reflexology and was not intended to inform individual choices about using reflexology.

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 reflexology on each outcome is limited, and the true effect may be substantially different. Major concerns about study design limitations and the likelihood that results that show large beneficial effects from reflexology may have been selectively published by trialists should be considered when deciding whether there is any credible evidence to support the use of reflexology. This review listed, but did not assess studies that compared reflexology to other interventions, so no conclusions can be drawn on whether reflexology is as effective as other interventions. Studies published in a language other than English were listed, but not included in the evaluation. The period over which reflexology was delivered varied across studies, although this generally reflected the reasons why reflexology was used (e.g. single treatments for relief of acute effects of surgery and procedures; treatment over a month or longer for chronic conditions). The timing of follow up also varied. In most studies, effects were measured immediately after the end of the reflexology treatment period. Longer-term effects were generally not reported and, as such, were not examined in the review so it is unknown whether any effects are sustained.

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 reflexology with any certainty. While further investigation of published and unpublished trials of reflexology 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. For example, implementing study design features that minimise the risk of bias, considering power of trials, measuring outcomes that are well established and patient relevant (e.g. as identified in consensus-based core outcome sets), reporting all measured outcomes, and ensuring trials are registered and reported in accordance with relevant reporting guidelines. The value of conducting more trials on reflexology would need 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 3 February 2023. Studies published after this date are not included in this review.

Executive summary

Background

Reflexology is a non-invasive touch-based therapy in which practitioners use their hands, thumbs and fingers to apply pressure to specific reflex points on the extremities (feet, hands, outer ears, face and lower legs) that are believed to stimulate a response in a corresponding organ or system in the body [5]. 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 reflexology 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 reflexology was effective. The current systematic review considered primary evidence and a wider range of publication dates.

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

Objectives

Primary objective to address the following question was

1. What is the effect of *reflexology* compared to an inactive control (no intervention, sham, placebo, wait list control, or a co-intervention offered to both groups, or continuation of usual care) among people with any condition, pre-condition, injury or risk factor on each of the outcomes for which reflexology is commonly used (pain, sleep quality, fatigue, emotional functioning and mental health, health-related quality of life, physical function and global symptoms)?

Secondary objectives related to the following questions

2. What is the effect of *reflexology* compared to an inactive control (no intervention, sham, placebo, wait list control, or a co-intervention offered to both groups, or continuation of usual care) on outcomes of importance for each underlying condition, pre-condition, injury or risk factor (for example, what is the effect on fatigue for people with cancer or advanced disease)?
3. What are the effects of *reflexology* compared to '*evidence-based' treatments* (active comparators) on outcomes for each underlying condition, pre-condition, injury or risk factor?
4. What evidence exists examining the effects of reflexology compared to other active comparators? (for inclusion in evidence inventory only, not the synthesis)

For objective 3, it was agreed that the planned comparison of the effects of reflexology to evidence-based treatments was not feasible because of the large volume of evidence contributing to objectives 1 and 2. Subsequent inspection of trials with an active comparator showed that the prespecified criteria for synthesis were not met (comparable PICO criteria, low risk of bias). For these reasons, active comparators are listed in appendices (E1 and E3). Other objectives were as stated in the protocol, with editing to include the outcome domains agreed through the prioritisation process.

Methods

This review was prospectively registered on the international prospective register of systematic reviews (PROSPERO ID CRD42023394291) and the methods pre-specified in a protocol published on the register and in a peer reviewed journal article [2]. The methods were based on the Cochrane Handbook for Systematic Reviews of Interventions [1]. The Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach was used to summarise and assess the certainty of evidence arising from this review [6-8]. The review is reported in accordance with the PRISMA 2020 statement [3, 4] 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 reflexology to an inactive control (no intervention, sham, placebo, wait list control, or a co-intervention offered to both groups, or continuation of usual care). Reflexology treatments were eligible irrespective of the method of reflexology, whether applied to feet or hands or other extremities (as above), the dose and duration of treatment, the setting in which it was delivered, and the training or qualifications of the practitioner.
- ***Types of populations***. Any condition, pre-condition, injury or risk factor (excluding healthy participants without clearly identified risk factors for the condition reflexology was used to prevent). Through the prioritisation process, it was agreed to include all conditions for which evidence was found except hypertension (studies are included in the evidence inventory for the review).
- ***Types of outcomes***. Any patient-important outcome for which reflexology is indicated was eligible for the review. Through the prioritisation process, outcomes determined to be critical or important for the synthesis were pain, sleep, fatigue, emotional functioning and mental health, health-related quality of life, physical function and global symptoms.
- ***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 (Cochrane Library, Issue 2, 2023), PubMed, Emcare, AMED and CINAHL on 3 February 2023. Searches were not limited by language, year of publication or publication status. Due to the volume of evidence, it was not feasible to search and screen registry records.

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, The National Health and Medical Research Council's Natural Therapies Working Committee (NTWC) with input from the Department's Natural Therapy Review Expert Advisory Panel (NTREAP) prioritised outcomes and confirmed the population groups proposed 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 all citations and full text reports was completed by two authors, independently. Data extraction and risk of bias assessment (ROB 2.0) methods were completed for each study by one of three authors; extracted data were checked by a senior author and a subset of ROB 2.0 assessments verified by the lead author. The data extraction and risk of bias tools had been applied and calibrated across authors in other reviews in the suite of natural therapies reviews, which ensured consistency across reviewers.

Comparisons were based on outcome domains (pain, sleep, fatigue, emotional functioning and mental health, health-related quality of life, physical function and global symptoms), 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. 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 about 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

Following screening 1239 records and 386 full-text reports we included 174 studies in the review. Of these, 123 studies (169 reports) were included in the evidence synthesis and 51 studies (56 reports) were included in the evidence inventory (mainly studies with active comparators or ineligible outcomes). No unique studies were identified from the public submissions. Eighty-three (83) studies (95 reports) are awaiting classification (63 studies in languages other than English of which 52 are likely to be eligible and 11 are unclear; the remainder were published in abstracts only).

Of the 123 studies included in the synthesis, 101 studies contributed results to at least one meta-analysis. Twenty-two (22) of the 123 studies did not contribute to any of the meta-analyses for which they were eligible because the outcome data were not reported, were reported incompletely or were uninterpretable. For similar reasons, a further 7 studies were missing from at least one analysis for which an outcome had been measured. In total, 47 out of 210 eligible results (22%) could not be included in the synthesis.

Effects of reflexology compared to an inactive control

Pain

Across conditions, the evidence about the effect of reflexology on pain is of very low certainty (46 trials, 3187 participants).

Effects on pain among the following groups are also very uncertain

- Surgery (acute postoperative) (8 trials, 603 participants)
- Procedures (during or after) (11 trials, 805 participants)
- Labour and childbirth (7 trials, 490 participants)
- Other acute pain (3 trials, 178 participants)
- Cancer or advanced disease (8 trials, 630 participants)
- Chronic musculoskeletal conditions (6 trials, 334 participants)
- Other chronic pain (3 trials, 147 participants)

No studies examined the effect of reflexology on pain for people with acute musculoskeletal conditions or migraine / headache.

Sleep quality

Across conditions, reflexology may improve sleep quality in general (12 trials, 782 participants; low certainty evidence).

There was low certainty evidence that Reflexology may improve sleep quality for people with symptoms of sleep disruption (6 trials, 376 participants).

Effects on sleep for the following groups are very uncertain

- Surgery (acute postoperative period) (2 trials, 110 participants)
- Hospitalisation (not for surgery) (2 trials, 164 participants)
- Cancer and advanced disease (2 trials, 132 participants)

No studies examined the effect of reflexology on sleep for people with chronic insomnia or dementia.

Fatigue

Across conditions, the evidence about the effect of reflexology on fatigue is of very low certainty (20 trials, 1590 participants).

Reflexology may reduce fatigue for people with other chronic conditions (8 trials, 535 participants; low certainty evidence).

Effects on fatigue for the following groups are very uncertain

- Cancer and advanced disease (7 trials, 741 participants)
- Chronic musculoskeletal conditions (3 trials, 157 participants)
- Pregnancy (2 trials, 157 participants)

Emotional functioning and mental health

Across conditions, the evidence about the effect of reflexology on emotional functioning and mental health is of very low certainty (40 trials, 3220 participants).

Effects for the following groups are also very uncertain

- Surgery (perioperative anxiety) (4 trials, 473 participants)
- Procedures (periprocedural anxiety) (7 trials, 551 participants)
- Hospitalisation (3 trials, 270 participants)
- Labour and childbirth (5 trials, 503 participants)
- Mental distress (8 trials, 447 participants)
- Cancer and advanced disease (10 trials, 826 participants)
- Mental disorders (2 trials, 130 participants)
- Dementia (1 trial, 20 participants).

Health-related quality of life

Across conditions, the evidence about the effect of reflexology on health-related quality of life is of very low certainty (20 trials, 1575 participants).

- Reflexology may improve health-related quality of life for people with other chronic and longer-term conditions. (12 trials, 777 participants)
- Effects are very uncertain for people with cancer and advanced disease. (8 trials, 798 participants).

Physical function

Across conditions, the evidence about the effect of reflexology on physical function is of very low certainty (10 trials, 877 participants).

Effects in the following groups are also very uncertain

- Cancer and advanced disease (3 trials, 475 participants)
- Chronic musculoskeletal conditions (3 trials, 172 participants)
- Other chronic conditions (4 trials, 230 participants)

No studies examined the effect of reflexology on physical function for people with migraine or headache (chronic or episodic).

Global symptoms

Across conditions, the evidence about the effect of reflexology on global symptoms is of very low certainty (18 trials, 1284 participants).

Effects in the following groups are also very uncertain

- Cancer and advanced disease (6 trials, 591 studies)
- Other chronic conditions (10 trials, 603 participants)
- Chronic respiratory conditions (2 trials, 90 participants)

No studies examined the effect of reflexology for people with chronic musculoskeletal conditions.

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).

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: 29 trials of which 22 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. Effects were often measured immediately after the end of the reflexology treatment period. Longer-term effects were generally not reported and, as such, were not examined in the review so it is unknown whether any effects are sustained.

Of the review process

In this review steps were taken to address potential limitations. 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 synthesis questions could not be fully specified at protocol stage; however, the final list of populations and outcomes eligible for the review were determined through a pre-specified prioritisation process, performed by NTWC with input from NTREAP, and 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, with safeguards implemented to ensure consistent application of the risk of bias tool. These included providing worked examples 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 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.

Assessments of cost-effectiveness, safety and studies of healthy populations were out of scope.

Conclusions

There is a large body of evidence about the effects of reflexology on health. Despite this, it is not possible to draw conclusions about the effects of reflexology with confidence for any condition or outcome. Many factors contribute to this uncertainty. Of greatest concern is that results that show large beneficial effects from reflexology (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 or are reported incompletely. 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

We found 123 trials that evaluated the effects of the reflexology compared to a sham intervention, usual care or no intervention on a prioritised outcome. 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 study design limitations and the likelihood that results that show large beneficial effects from reflexology may have been selectively published by trialists should be considered when deciding whether there is any credible evidence to support the use of reflexology. This review listed, but did not assess studies that compared the effects of reflexology to other interventions, so no conclusions can be drawn on whether reflexology is as effective as other interventions. Studies published in a language other than English were listed, but not included in the assessment.

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 reflexology 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 for other trials of natural therapies 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 reflexology (including, but not limited to, attention should be given to the power of the trial, bias arising from the randomisation process, the method of outcome assessment; and the reporting of results). Ensuring trials are registered and reported in accordance with relevant reporting guidelines (most importantly the CONSORT statement [9], but also TIDIER for reporting interventions [10]) is essential, as is reporting all measured outcomes and avoiding common errors that render results unusable. The value of conducting more trials on reflexology 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 reflexology 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 reflexology 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) [11].

Reflexology is reported to be a widely used natural therapy 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 reflexology was the fifth most popular CM from among 28 different therapies [12]. In Australia, a 2017 cross sectional survey examining consultation with complementary therapists and use of complementary medicine products found that about a third of all respondents (36%; 726/2019 adults) had consulted a complementary medicine practitioner in the last 12 months [13, 14]. Reflexologists were not among the most commonly consulted complementary medicine practitioners, and data are lacking on the prevalence and frequency of consultations with these practitioners [14].

1.1 Description of the intervention

Reflexology is a non-invasive touch-based therapy in which practitioners use their hands, thumbs and fingers to apply pressure to specific points on the feet, hands, face and outer ears [5]. The practice is based on the belief that there are "areas on the feet, hands, face, lower legs and ears" that correspond to other parts of the body, including organs and glands [15]. Manual touch techniques "such as thumb- and finger-walking, hook and backup and rotating-on-a-point" are used to apply pressure to these 'reflex points' in order to stimulate a therapeutic response [5]. The process is guided by a reflex map that shows the different zones of the body associated with each reflex point (for examples, see interactives maps [15]).

The use of specific touch techniques, and a reflex map of the body to guide touch, are the two components that differentiate reflexology from similar natural therapies, such as massage and acupressure [16]. Whereas massage involves manipulation of muscles or fascia to release tension, reflexology claims that working specific reflex points will stimulate the release of tension in other parts of the body [17, 18]. Reflexology and acupressure are both "reflex" therapies, in which practitioners use touch techniques on one part of the body to stimulate a 'reflex' in another part of the body. However, acupressure differs in that the reflex points run the entire length of the body (along meridians) rather than being located at the extremities [16, 17].

The Reflexology Association of Australia identifies four main methods of reflexology. They suggest that the most widely used is the Ingham method "in which the thumb (or finger) bends and straightens whilst maintaining a constant pressure across the area of the foot being worked" (thumbwalking) [19]. The method is named after American Eunice Ingham who is credited with developing the "foot maps and reflexology charts still in use today" and introducing "reflexology practices to the non-medical community in the 1930s" [17]. Other methods include Rwo Shur, which is practised in parts of Asia, and newer approaches that integrate concepts or principles from acupressure and Ayurveda [19].

Mode of administration and frequency of treatment

Reflexology is usually delivered by a trained practitioner in a one-to-one session. Sessions are typically of 40 minutes to an hour duration, and follow-up sessions may be scheduled for those with long term illness or ongoing health concerns. Some reflexology professional associations suggest weekly sessions over a 6-to-8-week period, with intermittent follow-up as required [17].

Practitioners of reflexology and regulation

Reflexology is practised by natural therapists, primarily reflexologists. Reflexology practice is not regulated by the Australian Health Practitioner Regulation National Law, which means there is no requirement for professional

registration of practitioners of reflexology [20, 21]. The Reflexology Association of Australia (RAoA) and the Australian Traditional Medicine Society (ATMS) offer membership to reflexology practitioners in Australia who have completed certificate- or diploma-level training accredited according to the requirements of the Australian Qualifications Framework [22, 23]. Both organisations set standards for practice and ethical conduct, but there appears to be no formal requirements for continuing professional education [22, 23]. Some professional associations have guidelines and position statements aimed at preventing the use of contraindicated practices and ensuring client safety (for examples, see [22, 24-26]). In the United States, reflexologists undergo testing and certification by the American Reflexology Certification Board (ARCB), an independent agency that aims to protect the public by ensuring the competency of reflexology practitioners and the legitimacy of the practice more generally [5].

1.2 How reflexology might work

Underlying all reflexology practice is the belief that applying pressure to specific reflex points on the extremities (feet, hands, outer ears, face and lower legs) will stimulate a response in a corresponding organ or system in the body. Several theories have been proposed to explain the mechanism by which reflexology may act [17]. The simplest of descriptions align with mainstream biopsychosocial models of pain suggesting that reflexology "promotes deep relaxation" which may reduce stress and improve mood, in turn reducing associated health effects such as the perception of pain [17, 25]. Other theories of how reflexology might work suggest that the application of pressure might stimulate the central nervous system leading to relaxation with potential effects on respiration and others systems [17]. Empirical evidence is lacking to support these theories or a scientific basis for the mechanism by which reflexology might act [15].

The Association of Reflexologists in the United Kingdom describe the mechanism as an "energetic effect" that encourages the "body towards a state of homeostasis and self-healing" [15]. In other descriptions, pressure on the reflex points is claimed to stimulate "the body's own natural healing process and that the body starts progressively clearing blockages, re-establishing energy flows and balancing itself, resulting in better health" [27]. The mention of "energetic" effects and "energy flows" references a belief in a vital force or *qi* that originates from Chinese and East Asian culture. The belief that there is an energy field ('vital energy') in the human body that supports healing is common to other therapies such as Reiki and Qigong [28]. Practices such as reflexology are claimed to restore the energy flow needed for optimum health and to combat illness.

1.3 Description of conditions for which reflexology is used

Professional associations for reflexologists emphasise that reflexology aims to treat "whole people not symptoms" and many have explicit statements in their code of practice that reflexologists must not claim to treat specific conditions (see for example [24, 25]). Those seeking reflexology may do so to complement other treatments and help them manage symptoms of a condition. Examples include to alleviate pain, symptoms of anxiety (that occur as a reaction to stress), low mood, sleep disturbance, and fatigue [29-31]. These indications align with the most commonly treated conditions reported in a 2015 survey completed by 61 practising reflexologists in Australia [21, 32]. Stress was the condition most frequently reported as 'often treated' (by 79% of reflexologists). Musculoskeletal conditions associated with chronic pain were also frequently reported as often treated, especially neck pain (75% of reflexologists), knee pain (58%), sciatica (54%), arthritis (52%) and foot problems (44%). Other conditions that were reported as 'often treated' were headache and migraine (53%), mental health conditions (46%), insomnia (45%), cancer (32%), sports injury (27%), diabetes (23%), and palliative care (20%).

There is particular interest in using reflexology 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) [33-35]. In Australia, reflexology is offered in some major cancer hospitals and facilities as part of integrative oncology services [36]. Among people with cancer and advanced disease, reflexology is used as a form of supportive care to enhance physical and emotional wellbeing, in addition to alleviating specific symptoms [33-35]. In other cases, reflexology is used as an alternative or adjunctive therapy by those seeking to avoid pharmacological or invasive treatment, especially for conditions associated with stress. For example, reflexology has been used to ameliorate anxiety and sleep disturbances [29], to relieve symptoms of multiple sclerosis [37], and for functional constipation [38]. Treatments are predominantly delivered in healthcare settings (primary care and secondary care cancer services linked to hospitals), with varying levels of integration with conventional providers [36].

Because reflexology is often sought for wellbeing and relief of symptoms, those receiving reflexology for the same indication may have very different underlying conditions (e.g. cancer, neck pain, stress). Examining the effects of reflexology 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 reflexology would differ importantly by condition. Where this is the case, a broad synthesis across conditions would address 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 reflexology 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 reflexology). The conclusion from the evidence evaluation conducted on reflexology for the *2015 Review* was that the "effect of reflexology on improving outcomes is uncertain" [39]. The evidence evaluation used overview methods, synthesising results from 18 systematic reviews published up to June 2013. Of the primary studies included in these systematic reviews (N=31), all but one were published before 2010. Since the completion of the original evidence evaluation, there has been steady growth in published trials of reflexology. A scoping search of PubMed indicates that about 100 randomised trials have been published since the *2015 Review* was conducted. In contrast to the 2015 reflexology evidence evaluation, this review will examine 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 reflexology in preventing and/or treating injury, disease, medical conditions or preclinical conditions [11]. The review focused on outcomes (and underlying conditions) for which reflexology is commonly sought or prescribed in Australia, to inform 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 objective to address the following question was

1. What is the effect of *reflexology* compared to an inactive control (no intervention, sham, placebo, wait list control, or a co-intervention that was offered to both groups, or continuation of usual care) among people with any condition, pre-condition, injury or risk factor on each of the outcomes for which reflexology is commonly used (pain, sleep quality, fatigue, emotional functioning and mental health, health-related quality of life, physical function and global symptoms)?

Secondary objectives related to the following questions

2. What is the effect of *reflexology* compared to an inactive control (no intervention, sham, placebo, wait list control, or a co-intervention offered to both groups, or continuation of usual care) on outcomes of importance for each underlying condition, pre-condition, injury or risk factor (for example, what is the effect on fatigue for people with cancer or advanced disease)?
3. What are the effects of *reflexology* compared to '*evidence-based' treatments* (active comparators) on outcomes for each underlying condition, pre-condition, injury or risk factor?
4. What evidence exists examining the effects of reflexology compared to active comparators? (for inclusion in evidence inventory only, not the synthesis)

For objective 3, it was agreed that the planned comparison of the effects of reflexology compared to evidence-based treatments was not feasible because of the large volume of evidence contributing to objectives 1 and 2. Subsequent inspection of trials with an active comparator showed that the prespecified criteria for synthesis were not met (comparable PICO criteria, low risk of bias). For these reasons, active comparators are listed in Appendix E1 (for studies that also contributed to objectives 1 and 2) and Appendix E3 (for studies that only contributed to the

evidence inventory). Other objectives were as stated in the protocol, with editing to include the outcome domains agreed through the prioritisation process. The final synthesis questions and criteria for including studies in each synthesis are presented in Figure 3.5.1.

3. Summary of methods

This review followed methods pre-specified in the protocol endorsed by NTWC. The protocol was prospectively registered on the International prospective register of systematic reviews (PROSPERO ID. CRD42023394291). The methods were based on the Cochrane Handbook for Systematic Reviews of Interventions [1]. The Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach was used to summarise and assess the certainty of evidence arising from this review [7, 8]. The review is reported in accordance with the PRISMA 2020 statement [3, 4].

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).

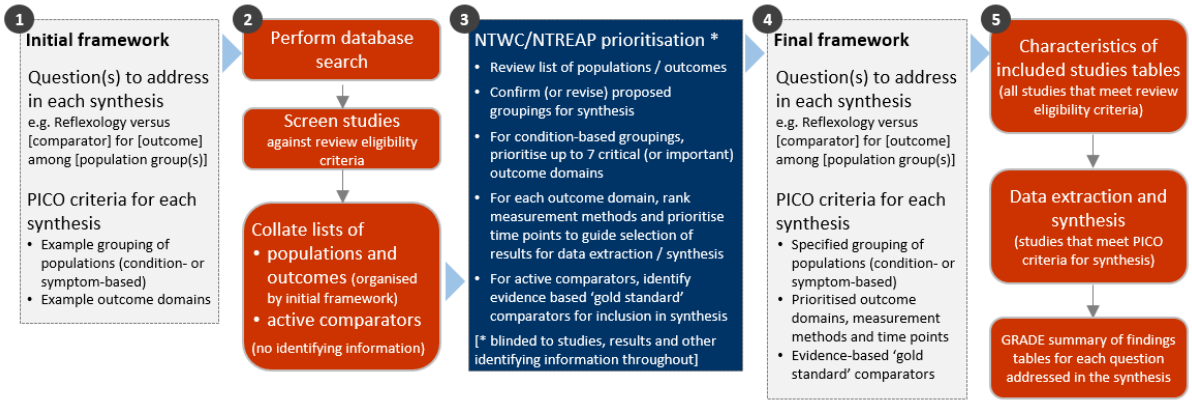

Fig 3.1 | Staged approach for developing the questions and analytic framework for this review. Active comparators were not considered in the prioritisation process because of the large number of studies comparing reflexology to inactive comparators. Separate tables are presented for studies included for the evidence synthesis (Appendix E1 and E2) and those in the evidence inventory (Appendix E3). Studies with ineligible populations, outcomes or active comparators are reported in the evidence inventory.

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) [40].

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 reflexology 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 reviewed and accepted a proposal to exclude one condition 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 reflexology.

3.1.3 Types of interventions

Reflexology was defined as a system of applying pressure to the outer extremities of the body (feet, hands, outer ears, and sometimes the lower limbs or face) on reflexes located within reflex maps of the body [27, 41]. Pressure is usually applied to the reflexes using the practitioner's hand, fingers and thumbs and may involve a range of specific touch techniques (e.g. thumb- and finger-walking, hook and backup and rotating-on-a-point). The modality is non-invasive.

Because of the close similarity of reflexology with related modalities, and the likelihood of identifying studies in which the defining components of reflexology are incompletely reported, studies were included if (a) the therapy is described as reflexology, or (b) one of the recognised synonyms for reflexology (reflex therapy, zone therapy, reflex point therapy) was used, the description of the intervention in the study report included the defining features of reflexology (immediately above), and the intervention was clearly not another modality (e.g. not massage or acupressure).

Reflexology treatments were eligible irrespective of the method of reflexology, whether applied to feet or hands or other extremities (as above), whether provided by a reflexologist or another practitioner, the setting in which reflexology was delivered, the training or qualifications of the practitioner, the dose and duration of treatment, or with or without the use of neutral oils, talc or cream (i.e. without an active ingredient such as an essential oil).

Comparisons

1. Reflexology versus any inactive control (no intervention, sham, placebo, wait list control, or a co-intervention offered to both groups, or continuation of usual care)

2. Reflexology versus '*evidence-based' treatments* (active comparators) (included in the evidence inventory, not the synthesis, due to large volume of studies for comparison 1)

3. Reflexology versus other active comparators (for inclusion in evidence inventory only, not the synthesis – See below).

Any co-intervention was eligible (i.e. pharmacological or non-pharmacological). Usual care comparators were eligible if there was an explicit statement that indicated that participants could continue to access their routine care or therapy (including self-care). If a comparator labelled as 'usual care' involved a defined intervention (i.e. specific treatments and processes selected by the researchers), this was deemed to be either an active intervention (if restricted to the comparator group) or a co-intervention (if able to be accessed by both groups, e.g. continuation of a specific medication).

We sought advice from NTWC on the categorisation of comparators that involved massage or reflex points. Based on this advice, we categorised comparators that involved massage as 'inactive' if the description indicated a placebo for massage used in the reflexology group (e.g. "placebo heel massage", "control group received simple touch without pressure"). If the description indicated an active massage intervention (e.g. Swedish massage with description of specific techniques), the comparator was categorised as active. We categorised comparators that involved reflex points as 'inactive' if the description indicated that the reflex points were not specific to the condition/symptoms (as identified by the trialists).

Active comparators eligible for the review were any pharmacological or non-pharmacological intervention, except natural therapies in other evidence evaluations. For comparison 2, a decision was made during the prioritisation step to include active comparators in the evidence inventory only (not the synthesis). This was initially due to the large volume of studies for comparison 1, but on closer inspection, the criteria for synthesis were not met for any evidence-based treatments (at least two low risk of bias studies with the same comparator, population and outcome).

We excluded head-to-head comparisons of reflexology (e.g. another reflexology touch technique or reflexology on a different part of the body, such as hand versus foot).

3.1.4 Types of outcomes

Any patient-important outcome that aligned with the reasons why reflexology 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 endorsed as critical or important for the synthesis through the prioritisation process were:

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

Nausea and vomiting was also endorsed as a critical/important outcome domain. Very few studies measured this outcome, and so these measures were considered under global symptoms rather than as a stand-alone domain.

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 measures were available), timing of outcome measurement (first measure after end of reflexology 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 (Cochrane Library, Issue 2, 2023), PubMed, Emcare (Ovid), AMED (Ovid) and CINAHL (EBSCOhost) on 3 February 2023. Searches were not limited by language, year of publication or publication status.

Citations from the 2015 evidence review and citations received from the Department's public call for evidence, NTREAP or the Committee were cross-checked against the citations retrieved from the database searches and any unique citations added to Covidence for screening.

3.3 Selection of studies

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 flow 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 Department's call for evidence, 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 deidentified, aggregate information about the populations and outcomes addressed in included studies (no bibliographic information, titles, details about the number of studies, participants, methodological quality or results). We proposed a list of outcome domains relevant to each population, and possible exclusions to limit scope and ensure a focus on patient-relevant outcomes. NTWC, with input from NTREAP, prioritised outcome domains and population groups 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 reflexology 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

Figure 3.5.1, panel A shows the final analytic framework for the evidence summary and synthesis. The framework provides a guide to the structure of the synthesis and reporting of results (see caption for details). 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). We separated acute conditions or indications from chronic and longer-term conditions, to provide greater clarity about which outcomes were relevant. For example, for people undergoing spinal 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. Hypertension was the only condition excluded from the synthesis.

Prioritised outcomes. The outcome domains specified in the initial framework were endorsed, and the outcomes relevant to each population groups were agreed. A hierarchy of outcome measures developed for another natural therapies review of aromatherapy was endorsed for use in the reflexology review (to be applied for selecting outcomes when multiple measures of an outcome domain were reported in a study). The main outcome domains endorsed as ineligible were (1) physiological function, signs and symptoms, (2) biomarkers of stress, and (3) biomechanical outcomes.

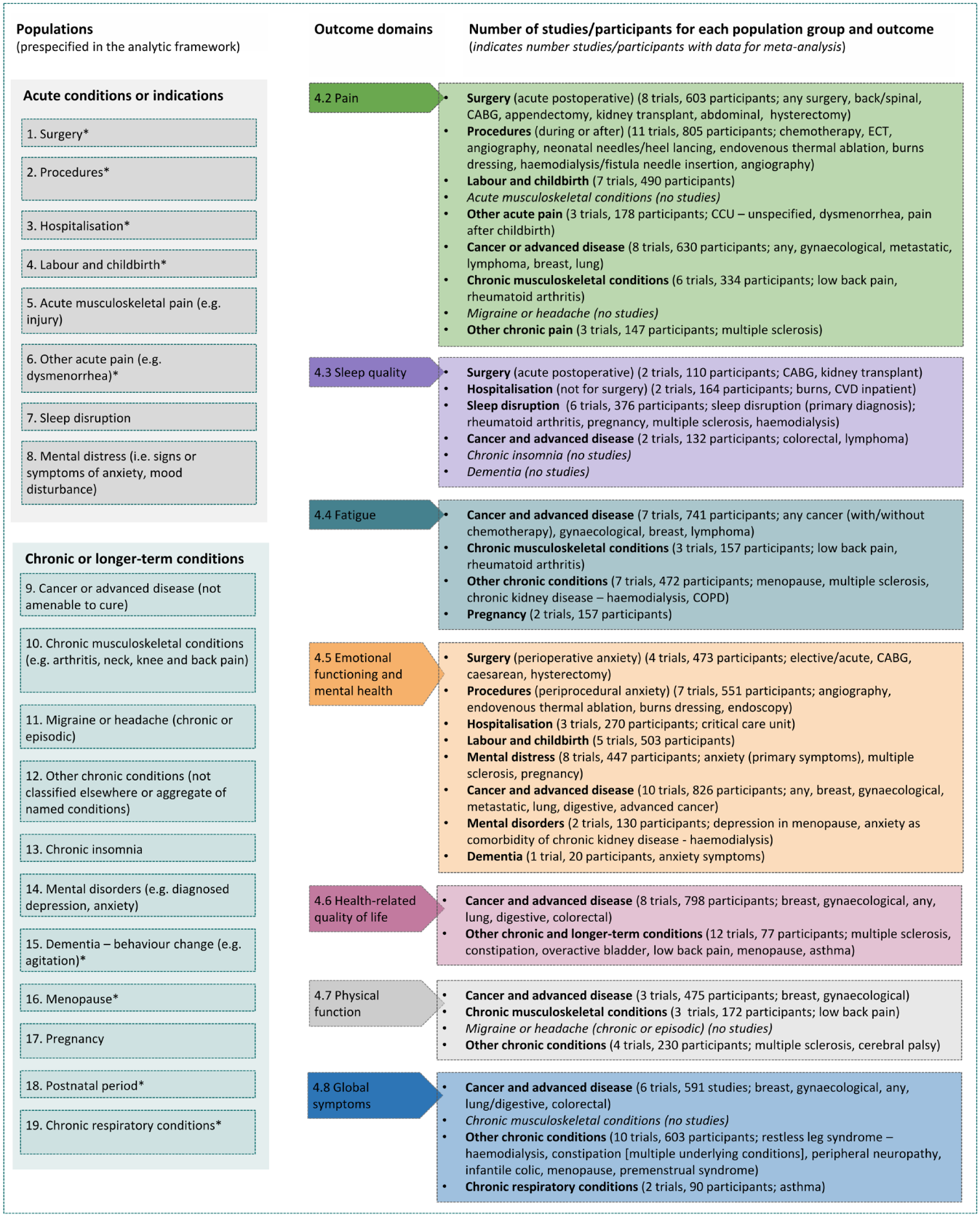

Fig 3.5.1 | Final analytic framework for the review as agreed through the prioritisation process (Appendix A5). Columns 1 to 2 show the populations and outcome domains eligible for the evidence synthesis. Column 3 shows the populations and outcome domains for which studies were available. Results are reported for each population group in the section indicated in column 2. Study-level data and meta-analyses are presented in the corresponding forest plot in each section. Population groups are those reported by practitioners of reflexology as often treated in the PRACI survey. *conditions not reported in PRACI survey but included for completeness.

3.6 Data extraction and management

3.6.1 Data extraction

Study data were collected and managed using REDCap electronic data capture tools [42, 43]. 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 (PN, LK, KJ) 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 [40, 44, 45]. After piloting of the tools by senior authors (SB, MM, SM, PN), we developed worked examples and supporting materials to ensure appropriate and consistent application of the ROB 2 guidance between assessors (PN, EK, KJ, MM). This guidance had been used by the author team to assess over 200 natural therapies studies prior to application in the current review. One review author (PN, LK, KJ, MM) then applied the tool to the selected results from each study following the RoB 2 guidance [40], and a second author (SB) checked a subset of 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 [40].

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 reflexology using the standardised mean difference (SMD).

Our interpretation was based on whether there was an important effect or not [6, 46], 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 reflexology 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, health-related quality of life, and physical function) and harm for other outcomes (an increase in pain, anxiety or agitation, fatigue, global symptoms).

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 reflexology for the outcome (review objectives 1), as well as estimates within each population group (review objective 2). Forest plots were used to visually depict the intervention effect estimates and their confidence intervals. Forest plots are stratified by condition and studies ordered by 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 [7, 46, 47], 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 [6].

- **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 is important, unexplained inconsistency in results across studies (size or direction of effect; interpreted in relation to the thresholds for an important effect).
- **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 because of selective non-reporting of results (or studies) that showed unfavourable effects.

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

- estimates of the effects of reflexology reported as standardised mean differences.
- the overall GRADE (rating of certainty) and an explanation of the reason(s) for rating down (or borderline decisions) [48].
- 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) [49].

3.7.3 Interpretation of findings (evidence statements)

When interpreting results, we followed GRADE guidance for writing informative statements [49]. 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, 'Reflexology 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 reflexology 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 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.

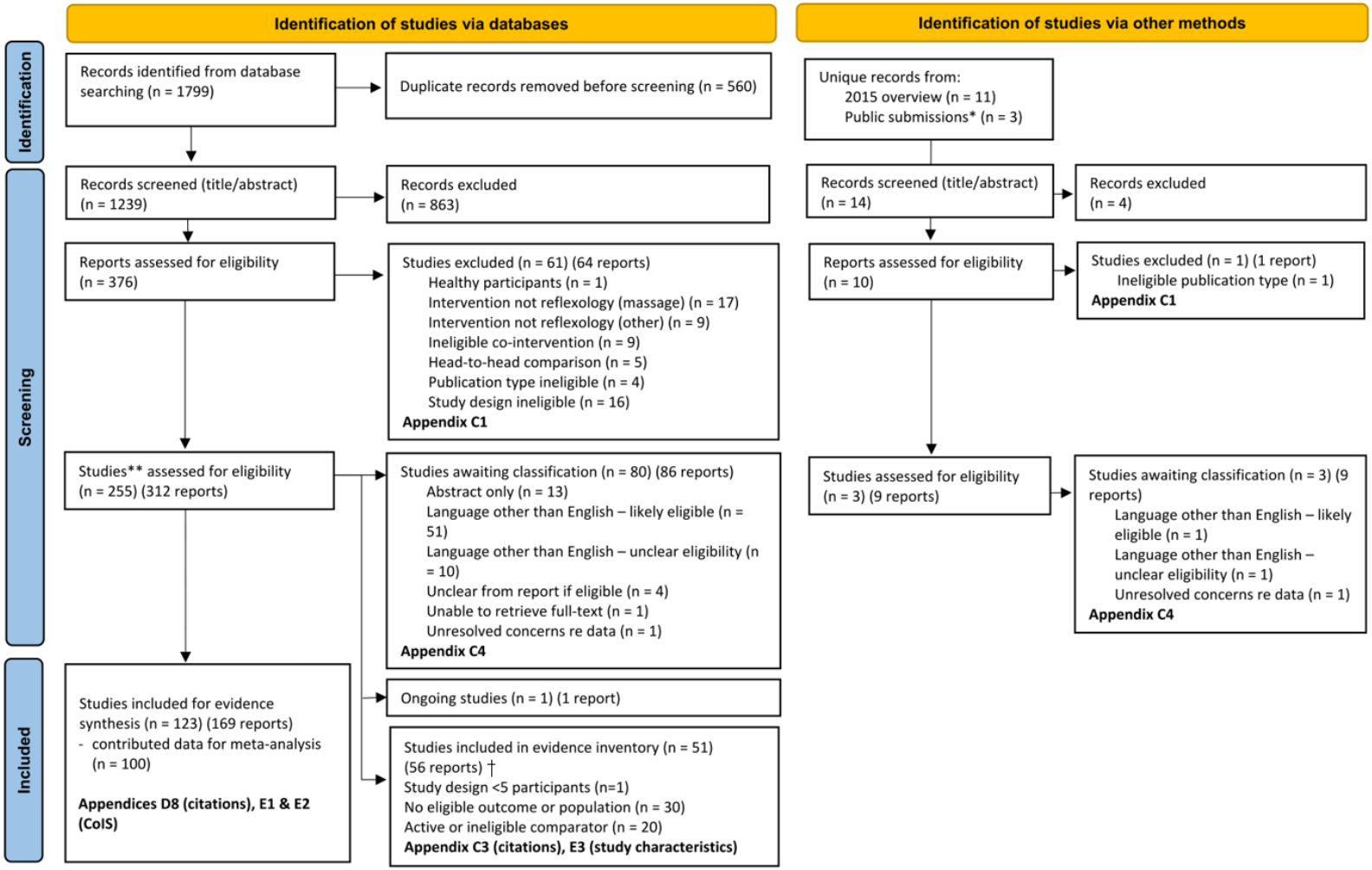

Fig. 4.1.1 | PRISMA diagram showing the flow of studies through the review. * See Appendix C2. ** Studies are the unit of interest in the review. For each study there may be multiple reports. † Exclusion of these studies from synthesis was agreed through the prioritisation process (Fig 3.5.1; Methods appendix A5, A6). CoIS: characteristics of included studies.

After removing duplicates, 1239 records from the database search were screened at title/abstract, of which 863 were excluded. Of the 376 reports assessed at full-text review, 61 studies (64 reports) were excluded, 80 (86 reports) were categorised as studies awaiting classification, 1 study (1 report) was ongoing, 51 (56 reports) were included in the evidence inventory, and 123 studies (169 reports) were included in the evidence synthesis. A further 14 unique records were assessed from other sources (11 from the 2015 evidence review and 3 from the public submissions). After screening and full-text review, 3 studies (9 reports) were categorised as studies awaiting classification, none was eligible for the evidence inventory or the evidence synthesis. Details follow.

Included studies

We included 174 studies in the review (after screening 1239 records and 386 full-text reports). Of these, 123 studies (169 reports) were included in the evidence synthesis and 51 studies (56 reports) were included in the evidence inventory (mainly studies with active comparators or ineligible outcomes). No unique studies (i.e. not found at search) were included from the public submissions.

Studies included for the evidence synthesis

Of the 123 studies included for the evidence synthesis (i.e., studies that measured at least one outcome of interest):

- 100 contributed data to at least one meta-analysis; of these, 91 reported data required for inclusion in all of the meta-analyses for which the study was eligible and 9 reported required data for a subset only,
- 23 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 included and missing from each analysis is reported in the results section for each outcome (Sections 4.2 to 4.8). In total, 44 out of 210 eligible results (22%) could not be included in the synthesis.

The reasons why studies did not contribute to any of the meta-analyses for which they were eligible were as follows:

(1) The trialists did not report any quantitative data for an outcome that was measured.
(2) The trialists did not report the statistics required for inclusion of results in meta-analysis, and these statistics could not be calculated or imputed from available data (despite application of all recommended methods in the Cochrane handbook).
(3) The reported results were uninterpretable (e.g. no information in the paper or other sources from which to interpret an outcome measure or direction of effect; ambiguous reporting such that we could not interpret the effects reported; combining of multiple arms that include treatments that were ineligible for the review).
(4) Identification of major errors or multiple minor errors.

Where problematic reporting raised concerns about selective non-reporting of unfavourable results (which may change the estimates of effect), we considered this in the GRADE assessment for the meta-analytic result.

Studies included in the evidence inventory

Of the 51 studies (56 reports) included in the evidence inventory (see Appendix C3 and E3)

- Population not prioritised. 2 studies examined effects of reflexology on hypertension (a decision endorsed in the prioritisation process; see Appendix A5)
- Active comparators. 20 studies had a comparator that was active
- Outcomes not prioritised. 28 studies only reported ineligible outcomes (e.g. physiological signs and symptoms, biomarkers of stress, biomechanical outcomes)
- Methodological concerns. one study had 5 participants per group or less. While this study was described as randomised, it was deemed to have too few participants for randomisation to be successful.

Excluded studies

After full-text screening (databases and other sources), 62 studies (65 reports) were excluded from the review (Figure 4.1.1, Appendix C1 for list of excluded studies). Among the reasons for exclusion, 26 were deemed not to be reflexology, 16 did not have an eligible study design and 9 had an ineligible co-intervention.

Studies awaiting classification

Following screening (databases and other sources), 83 studies (95 reports) were categorised as awaiting classification. Most were published in a language other than English and judged likely to be eligible for the review (52 studies) or the eligibility was unclear (11 studies). Thirteen (13) studies were reported in an abstract only, the eligibility was unclear for 4 studies, 2 studies had unresolved concerns regarding data, and for one study the full-text was unable to be retrieved (Appendix C4 for studies awaiting classification).

Ongoing and unpublished studies

From our database searches one protocol published in 2022 for a randomised trial was identified. From the 502 trial registry entries retrieved from CENTRAL, 74 were linked to included studies. We did not screen the remaining 428 records. While it may have been feasible to screen these records, a scan of the records suggests that a high proportion may be for eligible trials. Given this, we consider that it would not be feasible to perform the more detailed analyses required to determine which meta-analysis each ongoing or missing trial would contribute to and whether any missing results might change the findings of a given synthesis. 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).

Public submissions

Nine (9) submissions were received via the Department's public call for evidence. Three (3) were not retrieved by the database searches and were assessed separately – 2 had ineligible study designs and one was a systematic review (Appendix C2 for public submissions).

4.2 Pain

Fifty-four (54) trials examined the effect of reflexology on pain, of which 46 were included for meta-analysis and 8 (668 participants) could not be included (see below).

Of the 8 trials that were eligible for this analysis but could not be included:

- 2 trials (276 participants) reported results that were *unsuitable for meta-analysis* (i.e. the required statistics were unavailable and could not be calculated or imputed).
- 6 trials (392 participants) had results that we counted as 'missing' - either the outcome was measured but results were not reported, or the results were uninterpretable (e.g. information required to interpret the measure or results was missing, or there were errors in reported data).

Characteristics of included studies

Types of populations

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

Included studies examined the effect of reflexology as follows

Acute pain

- Surgery (acute postoperative) (8 trials, 603 participants; any surgery, back/spinal, Coronary Artery Bypass Grafting (CABG), appendectomy, kidney transplant, abdominal, hysterectomy)
- Procedures (during or after) (11 trials, 805 participants; chemotherapy, electroconvulsive therapy (ECT), angiography, neonatal needles/heel lancing, endovenous thermal ablation, burns dressing, haemodialysis/fistula needle insertion, angiography)
- Labour and childbirth (7 trials, 490 participants)
- *Acute musculoskeletal conditions (no studies)*
- Other acute pain (3 trials, 178 participants; critical care unit (CCU) – unspecified conditions, dysmenorrhea, pain after childbirth).

Chronic or longer-term pain

- Cancer or advanced disease (8 trials, 630 participants; any, gynaecological, metastatic, lymphoma, breast, lung)
- Chronic musculoskeletal conditions (6 trials, 334 participants; low back pain, rheumatoid arthritis)
- *Migraine or headache (no studies)*
- Other chronic pain (3 trials, 147 participants; multiple sclerosis).

Types of interventions

Of the 46 trials included in the meta-analysis, the majority evaluated foot reflexology (40 trials of foot only, 1 trial of foot and ear, 1 trial of foot and hand). In three trials, reflexology was applied to the hand and one did not report the location.

Session duration, frequency and intervention period. Treatment sessions were generally 30 to 60 minutes duration.

In all 19 trials that delivered a single session of reflexology, the treatment was administered for an acute indication:

- procedural pain or chemotherapy in 7 trials,
- acute postoperative pain or recent surgery in 6 trials, and
- labour and childbirth in 6 trials.

Of the 27 trials in which multiple sessions of reflexology were delivered, the treatment period and frequency varied.

- One day (3 trials; 2 labour and childbirth, 1 procedures)
- One week or less (10 trials; 4 procedures/dialysis/hospitalisation, 4 surgery, 2 cancer)
- One month or less (2 trials; 1 chronic musculoskeletal conditions, 1 cancer)
- More than a month (12 trials; 5 chronic musculoskeletal conditions, 3 chronic conditions, 2 cancer, 1 surgery, 1 acute pain).

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 were pain among neonates (e.g. N-PASS, Oucher, or FLACC were used) and older people (geriatric pain measure). Two studies used the Brief pain inventory (BPI), and one used the Wong-Baker FACES scale.

All but one study report results as a score on the ordinal scale (e.g. pain intensity on a VAS). One study reported ordinal data (categorising scores using cut-offs for severity). 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. For the single trial that reported dichotomous data, an effect estimate was calculated (odds ratio) and transformed to a standardised mean difference (Appendix B).

Effects of reflexology compared to an inactive control on pain

The effects of reflexology compared to an inactive control on pain are presented in Table 4.2.1 the GRADE summary of findings table. The certainty of evidence, and factors that influenced our certainty, are presented and explained in these tables. Study level and meta-analytic results are presented in the forest plot (Figure 4.2.1).

Across conditions, the evidence about the effect of reflexology on pain is of very low certainty due to very serious study design limitations and bias due to missing results (46 trials, 3187 participants). Effects in the following groups are also very uncertain

- Surgery (acute postoperative) (8 trials, 603 participants)
- Procedures (during or after) (11 trials, 805 participants)
- Labour and childbirth (7 trials, 490 participants)
- Other acute pain (3 trials, 178 participants)
- Cancer or advanced disease (8 trials, 630 participants)
- Chronic musculoskeletal conditions (6 trials, 334 participants)
- Other chronic pain (3 trials, 147 participants)

No studies examined the effect of reflexology on pain for people with acute musculoskeletal conditions or migraine / headache.

Table 4.2.1 | Summary of findings for the effect of reflexology versus inactive control for pain.

Outcomes	**Anticipated absolute effects*** (95% CI)		Relative effect (95% CI)	№ of participants (studies) contributing to meta-analysis	Certainty of the evidence (GRADE)	Interpretation (evidence statement)
	Risk with inactive control	**Risk with reflexology**				
Pain: all population groups	-	SMD **1.02 SD lower** (1.26 lower to 0.78 lower)	-	3187 (46 RCTs)	⨁◯◯◯ Very low[a,b,c]	The evidence is very uncertain about the effect of reflexology on pain in general (across all population groups).
Pain after surgery (acute postoperative period)	-	SMD **0.97 SD lower** (1.3 lower to 0.64 lower)	-	603 (8 RCTs)	⨁◯◯◯ Very low[c,d,e]	The evidence is very uncertain about the effect of reflexology on pain after surgery (acute postoperative period).
Pain during or after a procedure (acute procedural period)	-	SMD **1.31 SD lower** (1.92 lower to 0.7 lower)	-	805 (11 RCTs)	⨁◯◯◯ Very low[a,c,f]	The evidence is very uncertain about the effect of reflexology on pain during or after a procedure (acute procedural period).

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies) contributing to meta-analysis	Certainty of the evidence (GRADE)	Interpretation (evidence statement)
	Risk with inactive control	**Risk with reflexology**				
Pain during labour & childbirth	-	SMD **1.27 SD lower** (2.05 lower to 0.49 lower)	-	490 (7 RCTs)	⨁◯◯◯ Very low[a,c,f]	The evidence is very uncertain about the effect of reflexology on pain during labour & childbirth.
Other acute pain	-	SMD **0.78 SD lower** (3.28 lower to 1.71 higher)	-	178 (3 RCTs)	⨁◯◯◯ Very low[c,g,h,i]	The evidence is very uncertain about the effect of reflexology on other acute pain.
Pain: cancer and advanced disease	-	SMD **0.84 SD lower** (1.41 lower to 0.28 lower)	-	630 (8 RCTs)	⨁◯◯◯ Very low[b,c,j]	The evidence is very uncertain about the effect of reflexology on pain for people with cancer and advanced disease.
Pain: chronic musculoskeletal conditions	-	SMD **0.95 SD lower** (2.15 lower to 0.25 higher)	-	334 (6 RCTs)	⨁◯◯◯ Very low[c,k,l,m]	The evidence is very uncertain about the effect of reflexology on pain for people chronic musculoskeletal conditions.
Other chronic pain	-	SMD **0.37 SD lower** (1.85 lower to 1.11 higher)	-	147 (3 RCTs)	⨁◯◯◯ Very low[n,o]	The evidence is very uncertain about the effect of reflexology on other chronic pain.
Other conditions - not reported				-	-	No studies reported on the effects of reflexology on pain from acute musculoskeletal conditions or migraine/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). The resulting interpretation is as follows.

- For pain: < -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 are provided for domains for which there is a downgrade or a borderline judgment. In line with GRADE guidance, we not explain that there are no limitations unless the judgment was challenging (https://pubmed.ncbi.nlm.nih.gov/26796947/)

Explanations

a. Very serious risk of bias (-2). High proportion of studies are at high risk of bias and show large effects, such that the observed benefit may be overestimated.

b. No serious inconsistency. Heterogeneity statistics indicate inconsistent results and the confidence intervals do not overlap for many studies. However, the point estimate for a majority of studies indicate important benefit (SMD of -0.2 or lower) and other effects favour reflexology. For this reason, we have not downgraded for 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). For population groups, publication bias was strongly suspected because of the evidence from the funnel plots in combination with a high proportion of small studies showing large, statistically significant effects favouring reflexology (combined effect estimate is moderate to large).

d. Very serious risk of bias (-2). All studies in subgroup are at high risk of bias and show large effects, such that the observed benefit may be overestimated.

e. No serious inconsistency. Heterogeneity statistics indicate inconsistent results. However, the confidence intervals overlap for many studies and the point estimates for all studies indicate important benefit (SMD of -0.2 or lower).

f. No serious inconsistency. Heterogeneity statistics indicate inconsistent results and the confidence intervals do not overlap for many studies. However, the point estimate for all studies indicate important benefit (SMD of -0.2 or lower).

g. Serious risk of bias (-1). Two studies at high risk of bias show large effects, attenuated by one study at some concerns showing an effect favouring control. Serious concern that the observed benefit may be overestimated.

h. Serious inconsistency (-1). Non-overlapping confidence intervals; two studies showing large effects, one with an effect favouring control.

i. Very serious imprecision (-2). The 95% confidence interval crosses two thresholds for a small but important effect (SMD of 0.2 and -0.2), so the result is compatible with important benefit (SMD 3.28 lower) and important harm (SMD 1.71 higher). In part, this is due to the inconsistent effects, therefore we have rated down two levels (not three).

j. Very serious risk of bias (-2). All studies in subgroup are at high risk of bias and most show moderate to large effects, raising concerns that the observed benefit may be overestimated.

k. Serious risk of bias (-1). Majority of studies at high risk of bias show large effects, attenuated by two studies showing little to no effect Serious concern that the observed benefit may be overestimated.

l. Serious inconsistency (-1). Non-overlapping confidence intervals; four studies showing large effects, two showing little to no effect.

m. Serious imprecision (-1). The 95% confidence interval crosses two thresholds for a small but important effect (SMD of 0.2 and -0.2), so the result is compatible with important benefit (SMD 2.15 lower) and important harm (SMD 0.25 higher). In part, this is due to the inconsistent effects, and the extent to which the upper threshold is crossed is modest, therefore we have rated down one level (not two).

n. Serious risk of bias (-1). Two studies at with some concerns show little to no effect, whereas one study at high risk of bias show large benefit. Serious concern that the observed benefit may be overestimated, influenced by the effect observed in the single study.

o. Extremely serious imprecision (-3). The 95% confidence interval crosses two thresholds for a small but important effect (SMD of 0.2 and -0.2), so the result is compatible with important benefit (SMD 1.85 lower) and important harm (SMD 1.11 higher). The result is too imprecise to interpret.

Forest plots and analysis results

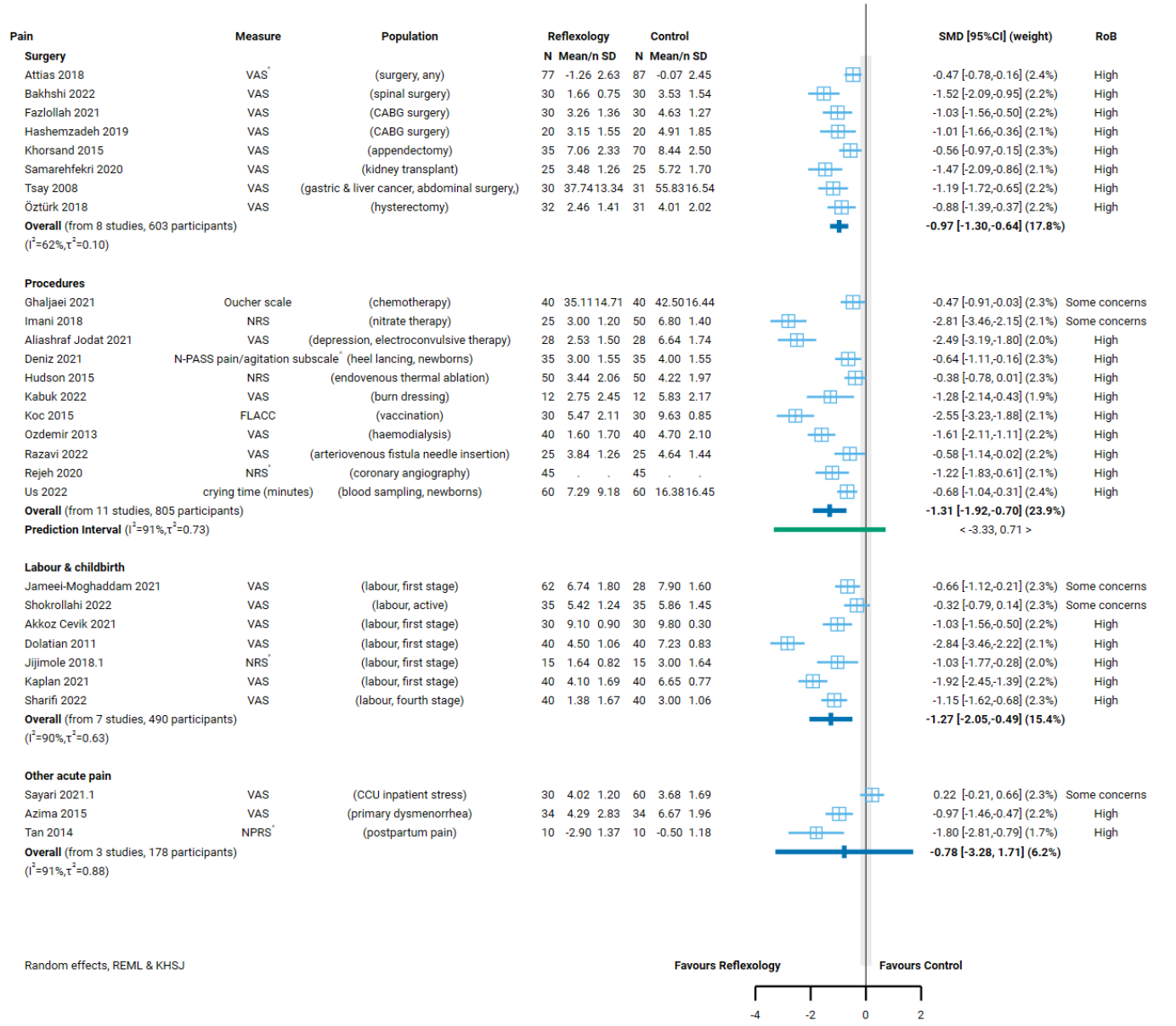

Fig 4.2.1 | Forest plot for main comparison. The effect of reflexology versus inactive control on pain. See next page for continuation of plot and figure caption. This section shows results for population groups with acute pain. Population groups chronic or longer-term pain are shown on the second part of the plot together with the combined estimate across groups and tests for subgroup differences. ^ indicates studies for which data transformation or imputation was required to include the result in the meta-analysis.

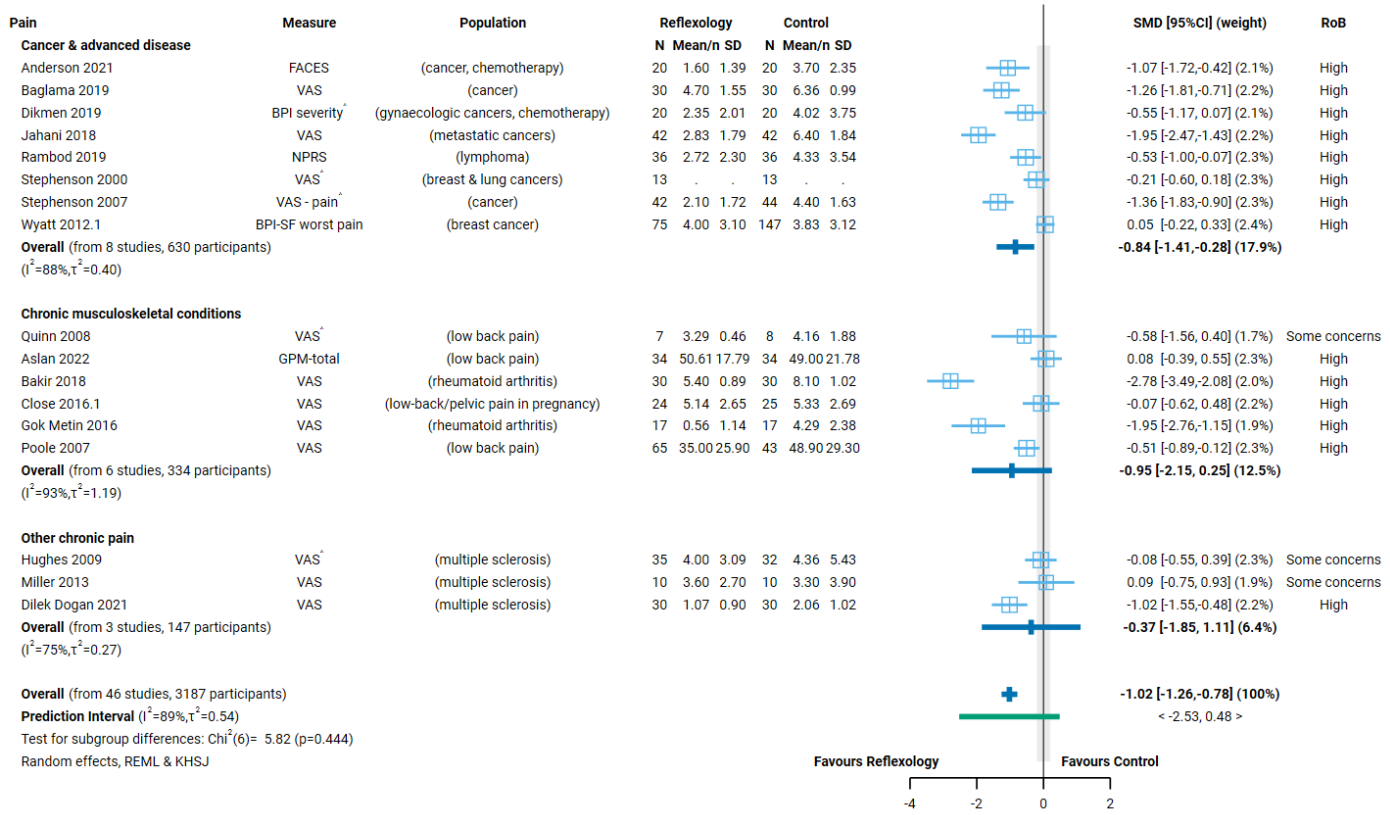

Fig 4.2.1 | Forest plot for main comparison. The effect of reflexology versus inactive control on pain. SMD=standardised mean difference. Blue lines show 95% confidence intervals (CI). The shaded grey area indicates the pre-specified range where the effect of reflexology is considered to be no different 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.

4.3 Sleep quality

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).

Sixteen (16) trials examined the effect of reflexology on sleep, of which 12 (782 participants) were included for meta-analysis and 4 (282 participants) could not be included (see below).

Of the 4 trials that were eligible for this analysis but could not be included.

- 2 trials (142 participants) reported results that were *unsuitable for meta-analysis* (i.e. the required statistics were unavailable and could not be calculated or imputed)
- 2 trials (140 participants) had results that we counted as 'missing' - either the outcome was measured but results were not reported, or the results were uninterpretable (e.g. information required to interpret the measure or results was missing, or there were errors in reported data).

Characteristics of included studies

Types of populations

The specific condition addressed in each trial 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.

Included trials examined the effect of reflexology as follows.

Acute or shorter-term sleep disruption

- Surgery (acute postoperative) (2 trials, 110 participants; CABG, kidney transplant)
- Hospitalisation (not for surgery) (2 trials, 164 participants; burns, cardiovascular disease (CVD) inpatient)
- Sleep disruption (6 trials, 376 participants; sleep disruption (primary diagnosis); rheumatoid arthritis, pregnancy, multiple sclerosis, haemodialysis)

Chronic or longer-term conditions

- Cancer and advanced disease (2 trials, 132 participants; colorectal, cancer lymphoma)

No studies were found for other prioritised conditions of chronic insomnia and dementia.

Types of interventions

All 12 trials included in the meta-analysis evaluated foot reflexology.

Session duration, frequency and intervention period. Treatment sessions were generally 20 to 60 minutes duration.

All 12 trials delivered multiple sessions of reflexology, the treatment period and frequency varied.

- One week or less (6 trials; 2 surgery, 1 periprocedural, 1 hospitalisation, 1 postnatal, 1 cancer)
- One month or less (6 trials; sleep disturbance or insomnia as primary diagnosis, sleep disturbance among people with chronic musculoskeletal conditions, chronic conditions, or during pregnancy).

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 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 studies measured sleep quality on a scale, the majority using the Pittsburgh Sleep Quality Index (PSQI; 7/12 trials). Other scales used were the Richards-Campbell Sleep Questionnaire (RCSQ; 2/12 trials), Verran and Snyder-Halpern Sleep Scale (VSH 2/12 trials) and the Sleep Condition Indicator (SCI, 1 trial).

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

Effects of reflexology compared to an inactive control on sleep quality

The effects of reflexology compared to an inactive control on sleep are presented in Table 4.3.1 the GRADE summary of findings table. The certainty of evidence, and factors that influenced our certainty, are presented and explained in these tables. Study level and meta-analytic results are presented in the forest plot (Figure 4.3.1).

Across conditions there is low certainty evidence, due to study design limitations and bias due to missing results, that reflexology may improve sleep quality (12 trials, 782 participants).

There is low certainty evidence, due to study design limitations and bias due to missing results, that reflexology may improve sleep quality for people with symptoms of sleep disruption (6 trials, 376 participants).

Effects on sleep for the following groups are very uncertain

- Surgery (acute postoperative period) (2 trials, 110 participants)
- Hospitalisation (not for surgery) (2 trials, 164 participants)
- Cancer and advanced disease (2 trials, 132 participants)

No studies examined the effect of reflexology on sleep for people with chronic insomnia or dementia.

Table 4.3.1 | Summary of findings for the effect of reflexology versus inactive control for 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	**Risk with reflexology**				
Sleep quality: all population groups	-	SMD **1.37 SD higher** (0.88 higher to 1.87 higher)	-	782 (12 RCTs)	⨁⨁◯◯ Low[a,b,c]	Reflexology may improve sleep quality in general (across all population groups).
Sleep quality after surgery (acute postoperative period)	-	SMD **0.6 SD higher** (0.94 lower to 2.14 higher)	-	110 (2 RCTs)	⨁◯◯◯ Very low[a,c,d]	The evidence is very uncertain about the effect of reflexology on sleep quality after surgery (acute postoperative period).
Sleep quality during hospitalisation	-	SMD **1.55 SD higher** (8.19 lower to 11.29 higher)	-	164 (2 RCTs)	⨁◯◯◯ Very low[a,c,e,f]	The evidence is very uncertain about the effect of reflexology on sleep quality during hospitalisation.
Sleep quality among people with signs or symptoms of sleep disruption (primary symptoms or as a comorbidity)	-	SMD **1.49 SD higher** (0.65 higher to 2.33 higher)	-	376 (6 RCTs)	⨁⨁◯◯ Low[a,c,e]	Reflexology may improve sleep quality among people with signs or symptoms of sleep disruption (primary symptoms or as a comorbidity).

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Risk with inactive control	Risk with reflexology				
Sleep quality among people living with cancer or advanced disease	-	SMD **1.72 SD higher** (6.59 lower to 10.02 higher)	-	132 (2 RCTs)	⨁◯◯◯ Very low[a,c,e,g]	The evidence is very uncertain about the effect of reflexology on sleep quality among people living with cancer or advanced disease.
Other conditions - not reported				-	-	No studies reported on the effects of reflexology on sleep for people with dementia.

***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). The resulting interpretation is as follows.

- For sleep quality: > 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 are provided for domains for which there is a downgrade or a borderline judgment. In line with GRADE guidance, we not explain that there are no limitations unless the judgment was challenging (https://pubmed.ncbi.nlm.nih.gov/26796947/)

Explanations

a. Serious risk of bias (-1). All studies in analysis at high risk of bias and show large effects. Serious concerns that the observed benefit may be overestimated.

b. No serious inconsistency. Confidence intervals do not overlap and heterogeneity statistics indicate important inconsistency, not fully explained by populations groups (I squared is high in most groups). However, all effect estimates are clearly above the threshold for important benefit, and hence have a consistent interpretation.

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). For population groups, publication bias was strongly suspected because of the evidence from the funnel plots in combination with a high proportion of small studies showing large, statistically significant effects favouring reflexology (combined effect estimate is moderate to large).

d. 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 2.14 higher) and important harm (SMD 0.94 lower). However, neither study has a CI that indicates important harm and the CI for the combined estimate has been calculated using a conservative method, so we rated down so we rated down two levels (i.e. not 3 as might be indicated by interpreting the CI for the combined estimate alone).

e. No serious inconsistency. Confidence intervals do not overlap and heterogeneity statistics indicate potentially important inconsistency. However all effect estimates are clearly above the threshold for important benefit, and hence have a consistent interpretation.

f. 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 11.29 higher) and important harm (SMD 8.19 lower). However, neither study has a CI that indicates important harm and the CI for the combined estimate has been calculated using a conservative method, so we rated down two levels (i.e. not 3 as might be indicated by interpreting the CI for the combined estimate alone).

g. 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 10.02 higher) and important harm (SMD 6.59 lower). However, neither study has a CI that indicates important harm and the CI for the combined estimate has been calculated using a conservative method, so we rated down two levels (i.e. not 3 as might be indicated by interpreting the CI for the combined estimate alone).

Forest plots and analysis results

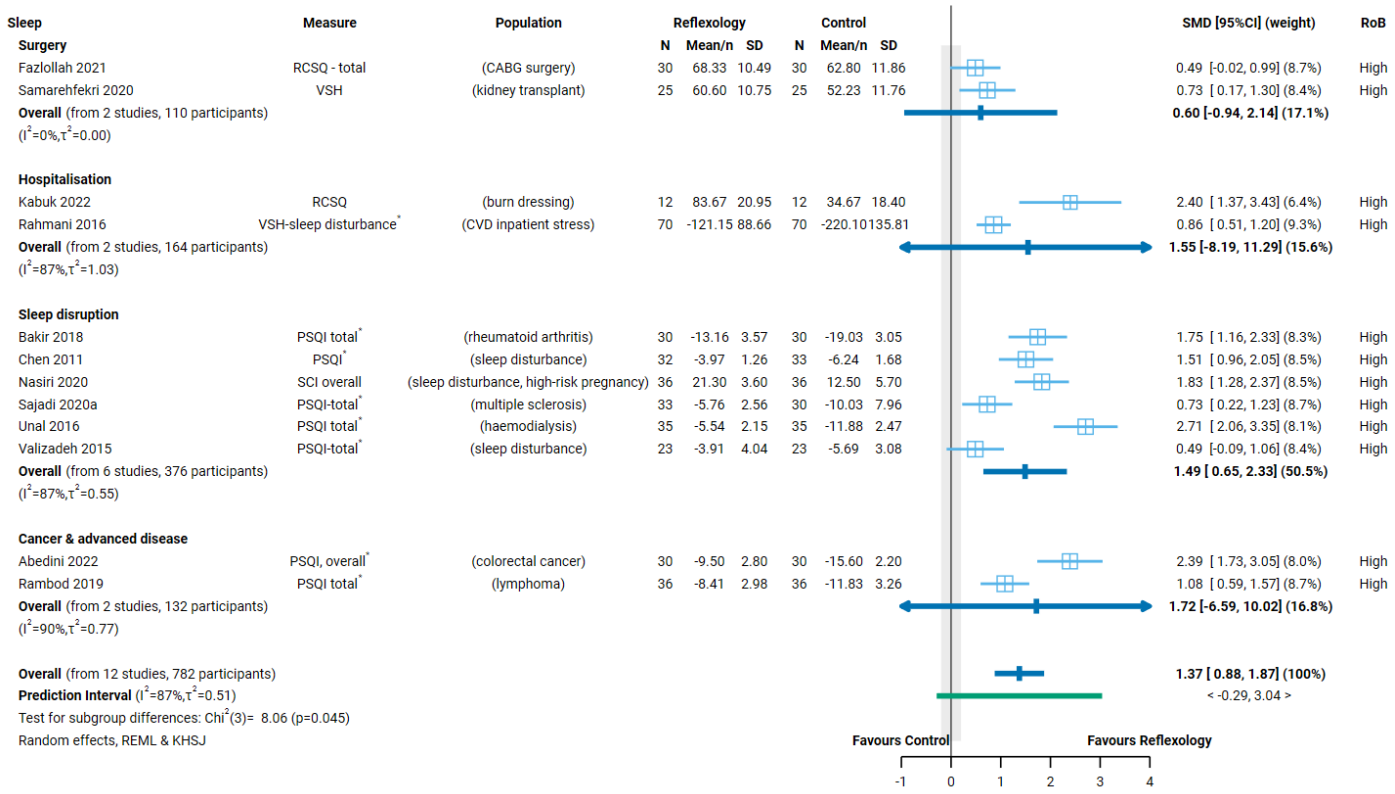

Fig 4.3.1 | Forest plot for main comparison. The effect of reflexology versus inactive control on sleep quality. SMD=standardised mean difference. Blue lines show 95% confidence intervals (CI). The shaded grey area indicates the pre-specified range where the effect of reflexology is considered to be no different from control (SMD -0.2 to 0.2 standard units). * Denotes studies for which the direction of effect was changed to match the overall plot (positive numbers are beneficial).

4.4 Fatigue

To be considered for the fatigue analysis, trials generally had to administer reflexology 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).

Twenty-one (21) trials examined the effect of reflexology on fatigue, of which 19 (1527 participants) were included for meta-analysis and two could not be included.

The 2 trials (108 participants) that were eligible for this analysis but could not be included were counted as 'missing' because the outcome was measured but results were not reported.

Characteristics of included studies

Types of populations

The specific condition addressed in each trial 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.

Included studies examined the effect of reflexology as follows.

Chronic or longer-term conditions

- Cancer and advanced disease (7 trials, 741 participants; any cancer (with/without chemotherapy), gynaecological, breast, lymphoma)
- Chronic musculoskeletal conditions (3 trials, 157 participants; low back pain, rheumatoid arthritis)
- Other chronic conditions (7 trials, 472 participants; menopause, multiple sclerosis, chronic kidney disease – haemodialysis, COPD)
- Pregnancy (2 trials, 157 participants).

Types of interventions

All 19 trials included in the meta-analysis evaluated foot reflexology.

Session duration, frequency and intervention period. Treatment sessions were generally 20 to 60 minutes duration.

In all 19 trials reflexology was delivered in multiple sessions. The treatment period and frequency varied.

- One week or less (4 trials; 2 cancer, 1 chronic conditions, 1 pregnancy)
- One month or less (1 trial; 1 cancer)
- More than a month (14 trials; 6 chronic conditions, 3 chronic musculoskeletal conditions, 4 cancer, 1 pregnancy).

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 4.4.1). 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 fatigue was measured, but the population was ineligible for inclusion in the analysis (i.e. not longer-term care, follow-up insufficient to detect meaningful improvement).

All studies measured fatigue on a scale, but different scales were used. The fatigue severity scale (FSS) was used in 6 trials and a visual analogue scale (VASF) in three trials. The fatigue symptoms checklist (FSC), brief fatigue inventory (BFI), and the short-form health survey (SF-36) – vitality scale were each used in two trials. Other scales used in single trials were the Piper fatigue scale (PFS), multidimensional fatigue inventory – 20 items (MFI-20), fatigue impact scale (FIS) and the MD Anderson symptom inventory (MDASDI).

All results were reported as a score on the original scale.

Effects of reflexology compared to an inactive control on fatigue

The effects of reflexology compared to an inactive control on fatigue are presented in Table 4.4.1 the GRADE summary of findings table. The certainty of evidence, and factors that influenced our certainty, are presented and explained in these tables. Study level and meta-analytic results are presented in the forest plot (Figure 4.4.1).

Across conditions, the evidence about the effect of reflexology on fatigue is of very low certainty due to study design limitations, inconsistency and bias due to missing results (19 trials, 1527 participants).

There is low certainty evidence, due to study design limitations and bias due to missing results, that reflexology may reduce fatigue for people with other chronic and longer-term conditions (7 trials, 472 participants).

Effects in the following groups are very uncertain

- Cancer and advanced disease (7 trials, 741 participants)
- Chronic musculoskeletal conditions (3 trials, 157 participants)
- Pregnancy (2 trials, 157 participants).

Table 4.4.1 | Summary of findings for the effect of reflexology versus inactive control for 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, no intervention, sham, co-intervention given in both groups)	**Risk with reflexology**				
Fatigue: All population groups	-	SMD **0.85 SD lower** (1.20 lower to 0.50 lower)	-	1527 (19 RCTs)	⨁◯◯◯ Very low[a,b,c]	The evidence is very uncertain about the effect of reflexology on fatigue in general (across all population groups).
Fatigue among people with cancer or advanced disease	-	SMD **0.73 SD lower** (1.56 lower to 0.1 higher)	-	741 (7 RCTs)	⨁◯◯◯ Very low[c,d,e,f]	The evidence is very uncertain about the effect of reflexology on fatigue among people living with cancer or advanced disease.
Fatigue among people with chronic musculoskeletal conditions	-	SMD **0.84 SD lower** (2.49 lower to 0.81 higher)	-	157 (3 RCTs)	⨁◯◯◯ Very low[c,d,g]	The evidence is very uncertain about the effect of reflexology on fatigue among people with chronic musculoskeletal conditions.
Fatigue among people with other chronic conditions	-	SMD **1.16 SD lower** (1.66 lower to 0.67 lower)	-	472 (7 RCTs)	⨁⨁◯◯ Low[c,d]	Reflexology may reduce fatigue among people with other chronic conditions.
Fatigue during pregnancy	-	SMD **0.24 SD lower** (5.94 lower to 5.45 higher)	-	157 (2 RCTs)	⨁◯◯◯ Very low[d,h,i]	The evidence is very uncertain about the effect of reflexology on fatigue 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). The resulting interpretation is as follows.

- For fatigue: < -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 are provided for domains for which there is a downgrade or a borderline judgment. In line with GRADE guidance, we not explain that there are no limitations unless the judgment was challenging (https://pubmed.ncbi.nlm.nih.gov/26796947/)

Explanations

a. Serious risk of bias (-1). High proportion of studies are at high risk of bias and show large effects, attenuated to some extent by study/studies showing little to no effect or favouring control. However, serious concerns that the observed benefit may be overestimated.

b. Serious inconsistency (-1). Non-overlapping confidence intervals; effect estimates vary importantly (some showing moderate to large benefit, others little to no effect). Differences are not explained by population subgroups (I squared high within most subgroups; test for subgroup differences not significant).

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). For population groups, publication bias was strongly suspected because of the evidence from the funnel plots in combination with a high proportion of small studies showing large, statistically significant effects favouring reflexology (combined effect estimate is moderate to large).

d. Serious risk of bias (-1). All studies at high risk of bias and multiple show large effects. Serious concerns that the observed benefit may be overestimated.

e. Serious inconsistency (-1). Non-overlapping confidence intervals; effect estimates vary importantly (some showing moderate to large benefit, others little to no effect).

f. Serious imprecision (-1). The 95% confidence interval crosses the threshold for a small but important improvement (SMD of -0.2), so the result is compatible with important benefit (SMD 1.56 lower) and little or no difference (SMD 0.10 higher).

g. Very serious imprecision (-2). The 95% confidence interval crosses two thresholds for a small but important effect (SMD of 0.2 and -0.2), so the result is compatible with important benefit (SMD 2.49 lower) and important harm (SMD 0.81 higher). However, we have rated down two levels (not three) because all effects favour reflexology and a conservative method was used to calculate the confidence interval (Hartung-Knapp-Sidik-Jonkman).

h. Serious inconsistency (-1). Non-overlapping confidence intervals; effect estimates vary importantly (one showing moderate to large benefit, the other harm).

i. Extremely serious imprecision (-3). 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 5.94 lower) and important harm (SMD 5.45 higher).

Forest plots and analysis results

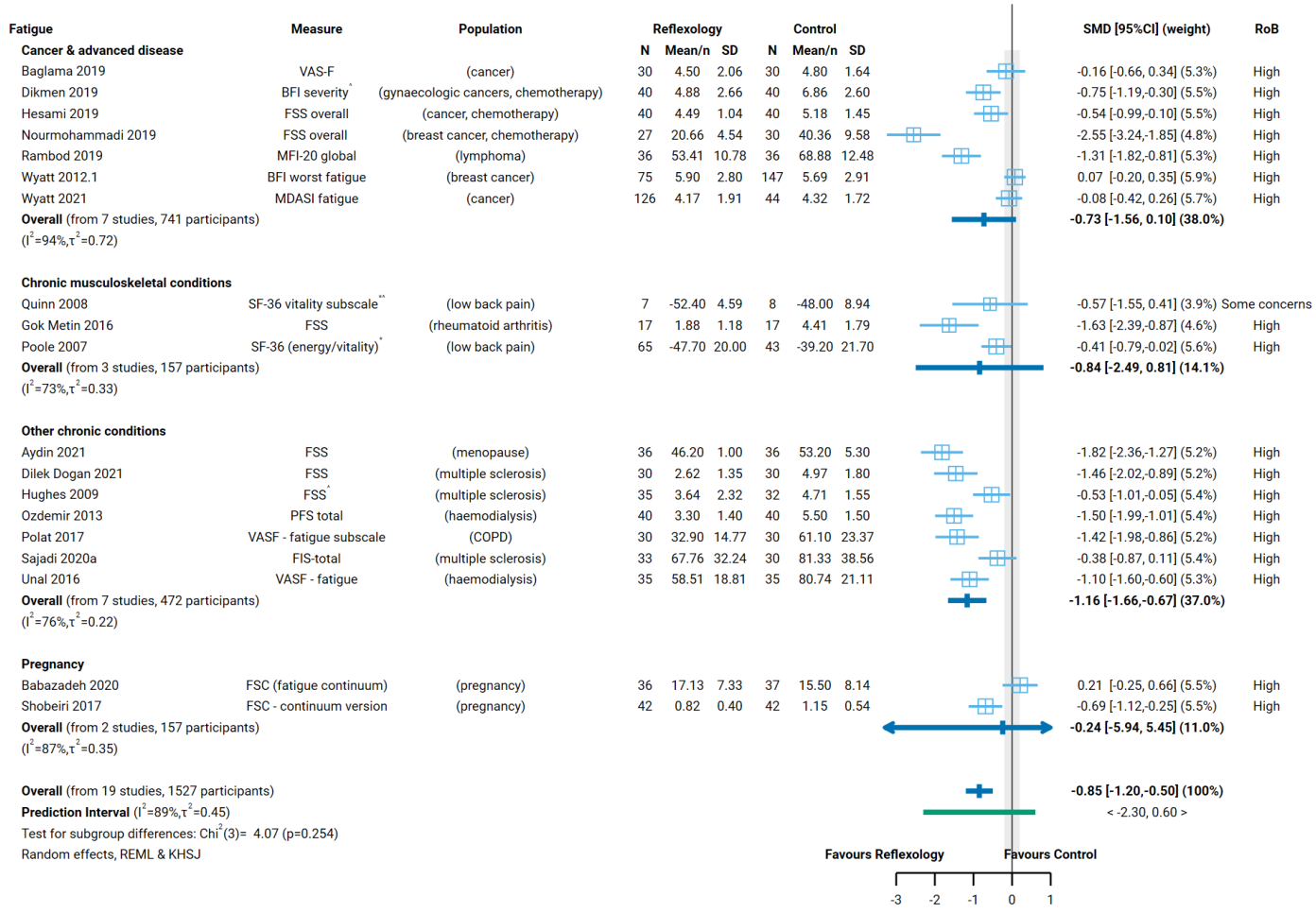

Fig 4.4.1 | Forest plot for main comparison. The effect of reflexology versus inactive control on fatigue. SMD=standardised mean difference. Blue lines show 95% confidence intervals (CI). The shaded grey area indicates the pre-specified range where the effect of reflexology is considered to be no different 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. * Denotes studies for which the direction of effect was changed to match the overall plot (negative numbers are beneficial).

4.5 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).

Fifty (50) trials examined the effect of reflexology on emotional functioning and mental health, of which 40 (3220 participants) were included for meta-analysis and 10 could not be included.

Of the 10 trials (899 participants) that were eligible for this analysis but could not be included.

- 3 trials (272 participants) reported results that were *unsuitable for meta-analysis* (i.e. the required statistics were unavailable and could not be calculated or imputed)
- 7 trials (627 participants) had results that we counted as 'missing' - either the outcome was measured but results were not reported, or the results were uninterpretable (e.g. information required to interpret the measure or results was missing, or there were errors in reported data).

Characteristics of included studies

Types of populations

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

Included studies examined the effect of reflexology as follows.

Acute or shorter-term conditions

- Surgery (perioperative anxiety) (4 trials, 473 participants; elective/acute, CABG, caesarean, hysterectomy)
- Procedures (periprocedural anxiety) (7 trials, 551 participants; angiography, endovenous thermal ablation, burns dressing, endoscopy)
- Hospitalisation (3 trials, 270 participants; critical care unit)
- Labour and childbirth (5 trials, 503 participants)
- Mental distress (8 trials, 447 participants; anxiety (primary symptoms), multiple sclerosis, pregnancy).

Chronic or longer-term conditions

- Cancer and advanced disease (10 trials, 826 participants; any, breast, gynaecological, metastatic, lung, digestive, advanced cancer)
- Mental disorders (2 trials, 130 participants; depression in menopause, anxiety as comorbidity of chronic kidney disease - haemodialysis)
- Dementia (1 trial, 20 participants, anxiety symptoms)
- Cancer and advanced disease (2 trials, 132 participants; colorectal, cancer lymphoma)
- *Chronic insomnia (no studies)*
- *Dementia (no studies).*

Types of interventions

Of the 40 trials included in the meta-analysis, the majority evaluated foot reflexology (36 trials of foot only). In three trials, reflexology was applied to the hand and one did not report the location.

Session duration, frequency and intervention period. Treatment sessions were generally 30 to 60 minutes duration.

In 17 trials a single session of reflexology was delivered, all for an acute indication.

- procedural pain in 6 trials,
- acute postoperative pain or recent surgery for cancer in 4 trials,
- labour and childbirth in 5 trials, and
- during hospitalisation (both in critical care unit) in 2 trials.

Of the 23 trials in which multiple sessions of reflexology were delivered, the treatment period and frequency varied. Longer treatment periods tended to be used for people with chronic or longer-term conditions.

- One day (1 trial after surgery for cancer in which 2 sessions were given on one day)
- One week or less (6 trials; 2 surgery, 1 procedures, 1 hospitalisation, 1 cancer, 1 mental distress - anxiety)
- One month or less (3 trials; 2 cancer, 1 mental distress - anxiety)
- More than a month (13 trials; 6 mental distress (4 among people with multiple sclerosis), 4 cancer, 2 mental disorders – depression, 1 dementia).

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). 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 40 trials measured emotional functioning and mental health with a scale, and all reported a scale score. Twenty trials used the State-Trait Anxiety Inventory (STAI). Other measures used in more than one trial were a Visual Analogue Scale for Anxiety (VAS-A, 7 trials), the Hospital Anxiety and Depression Scale (HADS, 5 trials), the Depression Anxiety and Stress Scale 21 (DASS-21, 3 trials), and the Beck Depression Inventory (BDI, 2 trials). A Numerical Rating Scale (NRS), the Apparent affect rating scale (AARS), and the Beck Anxiety Inventory (BAI) were each used in a single trial.

Effects of reflexology compared to an inactive control (usual care, no intervention, sham, co-intervention given in both groups) on emotional functioning and mental health

The effects of reflexology compared to an inactive control on emotional functioning and mental health are presented in Table 4.5.1 the GRADE summary of findings table. The certainty of evidence, and factors that influenced our certainty, are presented and explained in these tables. Study level and meta-analytic results are presented in the forest plot (Figure 4.5.1).

Across conditions, the evidence about the effect of reflexology on emotional functioning and mental health is of very low certainty due to study design limitations, inconsistent effects and bias due to missing results (40 trials, 3220 participants; Table 4.5.1, Figure 4.5.1). Effects in the following groups are also very uncertain

- Surgery (perioperative anxiety) (4 trials, 473 participants)
- Procedures (periprocedural anxiety) (7 trials, 551 participants)
- Hospitalisation (3 trials, 270 participants)
- Labour and childbirth (5 trials, 503 participants)
- Mental distress (8 trials, 447 participants)
- Cancer and advanced disease (10 trials, 826 participants)
- Mental disorders (2 trials, 130 participants)
- Dementia (1 trial, 20 participants).

Table 4.5.1 | Summary of findings for the effect of reflexology versus inactive control for 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	Risk with reflexology				
Emotional functioning and mental health: all population groups	-	SMD **0.69 SD lower** (0.93 lower to 0.44 lower)	-	3220 (40 RCTs)	⨁◯◯◯ Very low[a,b,c]	The evidence is very uncertain about the effect of reflexology on emotional functioning and mental health in general (across all population groups).
Emotional functioning and mental health: perioperative anxiety (surgery)	-	SMD **1.03 SD lower** (2.13 lower to 0.07 higher)	-	473 (4 RCTs)	⨁◯◯◯ Very low[c,d,e]	The evidence is very uncertain about the effect of reflexology on perioperative anxiety (surgery).

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Risk with inactive control	Risk with reflexology				
Emotional functioning and mental health: periprocedural anxiety (procedures)	-	SMD **0.95 SD lower** (1.85 lower to 0.05 lower)	-	551 (7 RCTs)	⨁◯◯◯ Very low[c,f,g]	The evidence is very uncertain about the effect of reflexology on emotional functioning and mental health: periprocedural anxiety (procedures) .
Emotional functioning and mental health: anxiety during hospitalisation (critical care unit)	-	SMD **0.51 SD lower** (1.7 lower to 0.69 higher)	-	270 (3 RCTs)	⨁◯◯◯ Very low[a,c,h,i]	The evidence is very uncertain about the effect of reflexology on anxiety during hospitalisation (critical care unit).
Emotional functioning and mental health: anxiety during labour and childbirth	-	SMD **0.92 SD lower** (2.02 lower to 0.17 higher)	-	503 (5 RCTs)	⨁◯◯◯ Very low[c,f,j]	The evidence is very uncertain about the effect of reflexology on anxiety during labour and childbirth.
Emotional functioning and mental health among people with signs or symptoms of mental distress	-	SMD **0.62 SD lower** (1.39 lower to 0.14 higher)	-	447 (8 RCTs)	⨁◯◯◯ Very low[c,f,k,l]	The evidence is very uncertain about the effect of reflexology on emotional functioning and mental health among people with signs or symptoms of mental distress.
Emotional functioning and mental health among people living with cancer or advanced disease	-	SMD **0.38 SD lower** (0.86 lower to 0.09 higher)	-	826 (10 RCTs)	⨁◯◯◯ Very low[c,f,m,n]	The evidence is very uncertain about the effect of reflexology on emotional functioning and mental health among people living with cancer or advanced disease.
Emotional functioning and mental health among people with a diagnosed mental disorder (depression or anxiety)	-	SMD **0.85 SD lower** (3.52 lower to 1.81 higher)	-	130 (2 RCTs)	⨁◯◯◯ Very low[c,o,p]	The evidence is very uncertain about the effect of reflexology on emotional functioning and mental health among people with a diagnosed mental disorder (depression or anxiety).
Emotional functioning and mental health among people with dementia (anxiety)	-	SMD **0.05 SD lower** (0.49 lower to 0.39 higher)	-	20 (1 RCT)	⨁◯◯◯ Very low[c,q,r,s]	The evidence is very uncertain about the effect of reflexology on anxiety among people with dementia. No studies measured other behavioural or psychological symptoms of dementia.

***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). The resulting interpretation is as follows.

- For emotional functioning and mental health (EFMH): < -0.2 is beneficial, -0.2 to 0.2 is trivial or unimportant ("little or no difference" between treatments), > 0.2 is harmful. Most studies in this analysis report measures of anxiety (or another mental health outcome where a lower score is beneficial). For this reason, a lower score is interpreted as favouring reflexology (e.g. a reduction in perioperative anxiety with reflexology is a beneficial effect).

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 are provided for domains for which there is a downgrade or a borderline judgment. In line with GRADE guidance, we not explain that there are no limitations unless the judgment was challenging (https://pubmed.ncbi.nlm.nih.gov/26796947/)

Explanations

a. Serious risk of bias (-1). High proportion of studies are at high risk of bias and show large effects, attenuated to some extent by study/studies showing little to no effect or favouring control. However, serious concerns that the observed benefit may be overestimated.

b. Serious inconsistency (-1). Non-overlapping confidence intervals; effect estimates vary importantly (majority showing moderate to large benefit, 14/40 show little to no effect or favour control). Differences are not explained by population subgroups (I squared high within most subgroups; test for subgroup differences not significant).
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). For population groups, publication bias was strongly suspected because of the evidence from the funnel plots in combination with a high proportion of small studies showing large, statistically significant effects favouring reflexology (combined effect estimate is moderate to large).
d. Very serious risk of bias (-2). All studies in subgroup are at high risk of bias and most show moderate to large effects, raising concerns that the observed benefit may be overestimated.
e. Serious imprecision (-1). The 95% confidence interval crosses the threshold for a small but important improvement (SMD of -0.2), so the result is compatible with important benefit (SMD 2.13 lower) and little or no difference (SMD 0.07 higher).
f. Serious risk of bias (-1). High proportion of studies are at high risk of bias and show large effects. Serious concerns that the observed benefit may be overestimated.
g. Serious imprecision (-1). The 95% confidence interval crosses the threshold for a small but important improvement (SMD of -0.2), so the result is compatible with important benefit (SMD 1.85 lower) and little or no difference (SMD 0.05 lower).
h. No serious inconsistency. Confidence intervals overlap, so results are compatible despite effect estimates that suggest different effects (2 studies show moderate to large benefit, one little difference).
i. Very serious imprecision (-2). The 95% confidence interval crosses two thresholds for a small but important effect (SMD of 0.2 and -0.2), so the result is compatible with important benefit (SMD 1.7 lower) and important harm (SMD 0.69 higher). In part, this is due to the inconsistent effects, therefore we have rated down two levels (not three).
j. Serious imprecision (-1). The 95% confidence interval crosses the threshold for a small but important improvement (SMD of -0.2), so the result is compatible with important benefit (SMD 2.02 lower) and little or no difference (SMD 0.17 higher).
k. No serious inconsistency. Confidence intervals overlap, so results are compatible despite variable effect estimates (some suggesting important benefit and others a trivial effect).
l. Serious imprecision (-1). The 95% confidence interval crosses the threshold for a small but important improvement (SMD of -0.2), so the result is compatible with important benefit (SMD 1.39 lower) and little or no difference (SMD 0.14 higher).
m. Serious inconsistency (-1). Non-overlapping confidence intervals; effect estimates vary importantly (some showing moderate to large benefit, some little to no effect or favouring control).
n. Serious imprecision (-1). The 95% confidence interval crosses the threshold for a small but important improvement (SMD of -0.2), so the result is compatible with important benefit (SMD 0.86 lower) and little or no difference (SMD 0.09 higher).
o. Serious risk of bias (-1). Both studies at high risk of bias and show large effects. Serious concerns that the observed benefit may be overestimated.
p. 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 3.52 lower) and important harm (SMD 1.81 higher). However, neither study has a CI that indicates important harm and the CI for the combined estimate has been calculated using a conservative method, so we rated down one level (i.e. not 3 as might be indicated by interpreting the CI for the combined estimate alone).
q. No serious risk of bias. Single study at some concerns, however there is little to no effect. While it plausible that bias could lead to underestimation of harm, this seems unlikely.
r. Serious indirectness (-1): Evidence from one small study among people with dementia. Uncertain whether results are generalisable to other people with dementia.
s. Very serious imprecision (-2). The 95% confidence interval crosses two thresholds for a small but important effect (SMD of 0.2 and -0.2), so the result is compatible with important benefit (SMD 0.49 lower) and important harm (SMD 0.39 higher).

Forest plots and analysis results

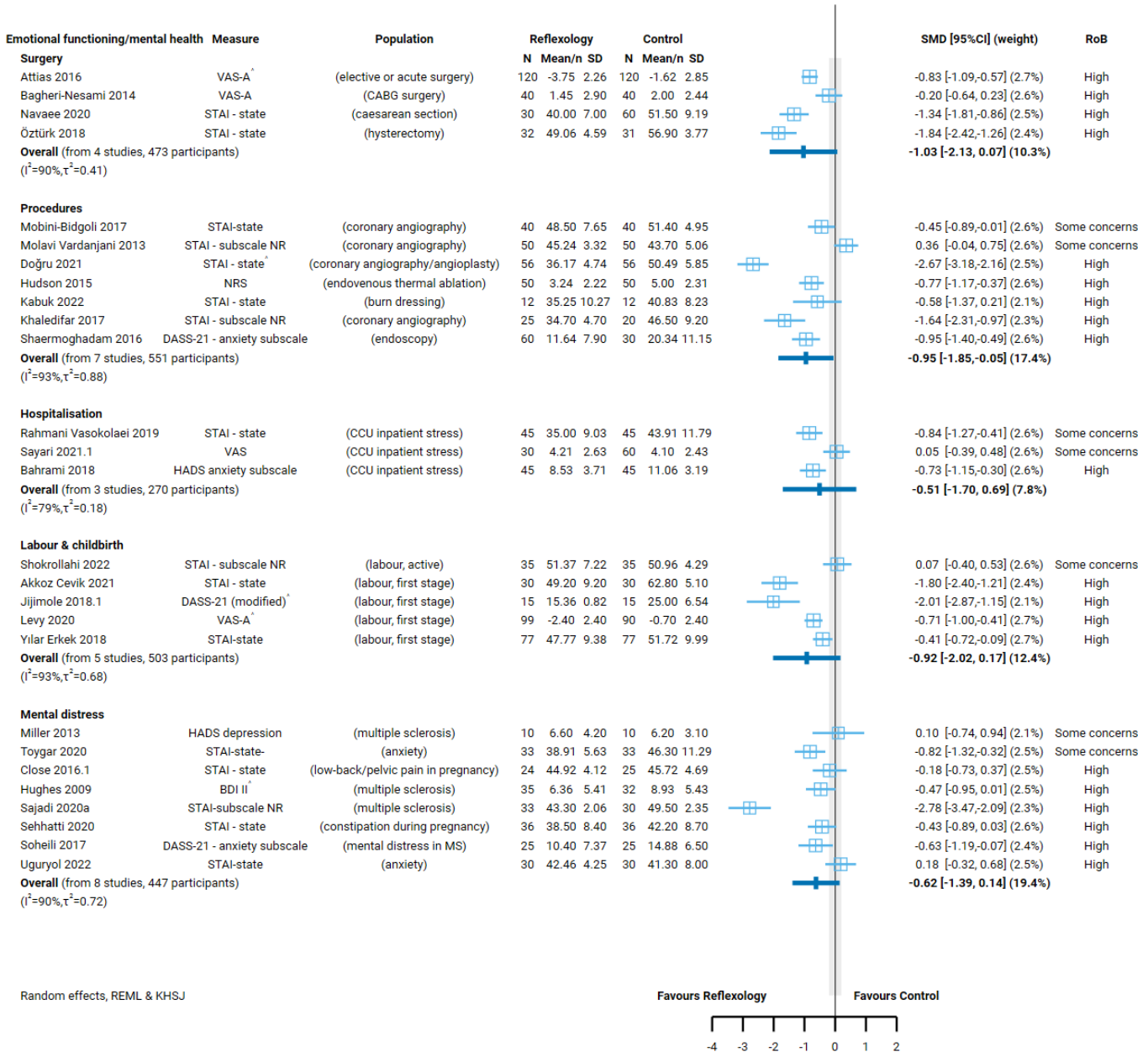

Fig 4.5.1 | Forest plot for main comparison. The effect of reflexology versus inactive control on emotional functioning and mental health. Most studies in this analysis reported measures of anxiety where a reduction in score is beneficial, so the direction of effect is standardised so that negative values indicate an effect favouring reflexology. See next page for continuation of plot and figure caption. This section shows results for population groups with acute or shorter-term conditions. Population groups with chronic or longer-term conditions are shown on the second part of the plot together with the combined estimate across groups and tests for subgroup differences. ^ indicates studies for which data transformation or imputation was required to include the result in the meta-analysis.

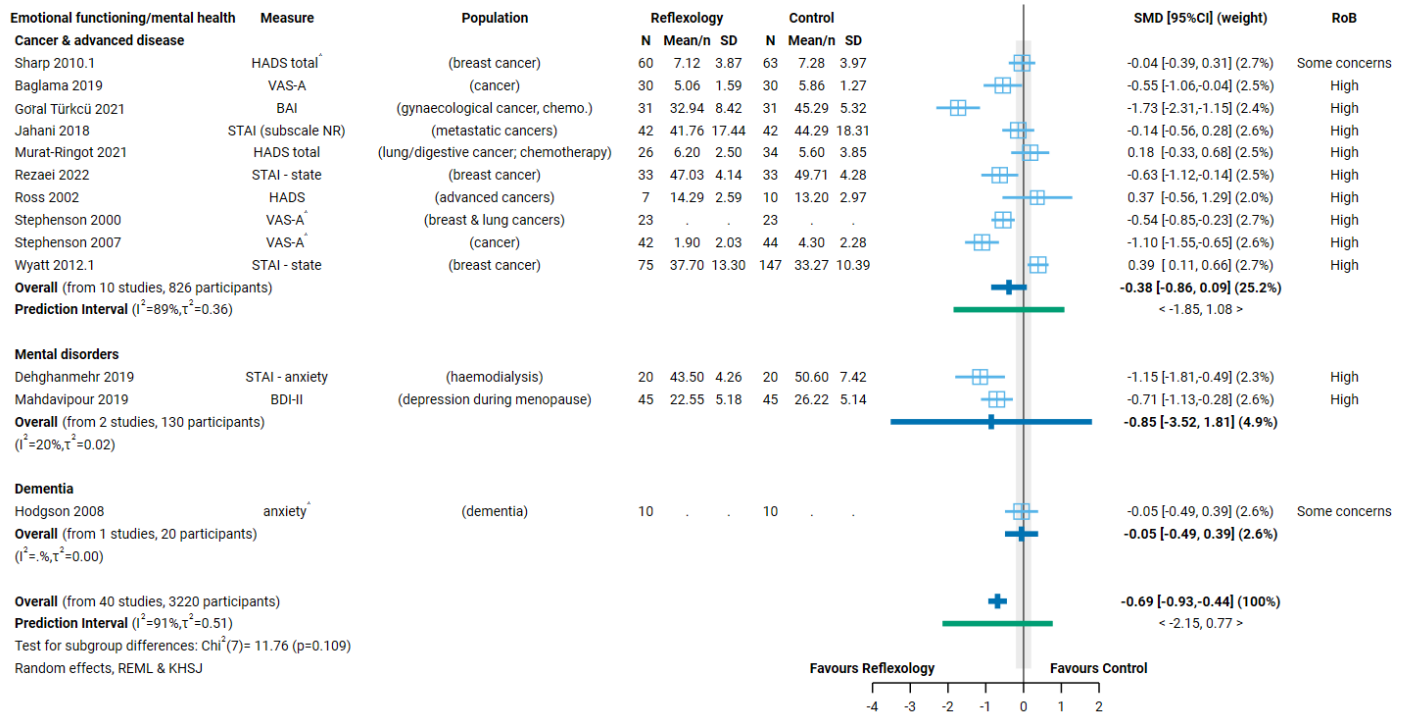

Fig 4.5.1 | Forest plot for main comparison. The effect of reflexology versus inactive control on emotional functioning and mental health. SMD=standardised mean difference. Blue lines show 95% confidence intervals (CI). The shaded grey area indicates the pre-specified range where the effect of reflexology 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.

4.6 Health-related quality of life (HR-QoL)

To be considered for the HR-QoL analysis, trials had to administer reflexology 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).

Twenty-seven (27) trials examined the effect of reflexology on HR-QoL, of which 20 (1575 participants) were included for meta-analysis and 7 could not be included.

Of the 7 trials (279 participants) that were eligible for this analysis but could not be included.

- 3 trials (132 participants) reported results that were *unsuitable for meta-analysis* (i.e. the required statistics were unavailable and could not be calculated or imputed)
- 4 trials (147 participants) had results that we counted as 'missing' - either the outcome was measured but results were not reported, or the results were uninterpretable (e.g. information required to interpret the measure or results was missing, or there were errors in reported data).

Characteristics of included studies

Types of populations

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

Included studies examined the effect of reflexology as follows.

Chronic or longer-term conditions

- Cancer and advanced disease (8 trials, 798 participants; breast, gynaecological, any, lung, digestive, colorectal)
- Other chronic and longer-term conditions (12 trials, 77 participants; multiple sclerosis, constipation, overactive bladder, low back pain, menopause, asthma).

Types of interventions

Of the 20 trials included in the meta-analysis, 19 evaluated foot reflexology and one did not report the location.

Session duration, frequency and intervention period. Treatment sessions were generally 30 to 60 minutes duration.

In all 20 trials, reflexology was delivered in multiple sessions. The treatment period and frequency varied.

- One month or less (3 trials; 1 chronic musculoskeletal condition, 1 chronic condition, 1 cancer)
- More than a month (17 trials; 2 chronic musculoskeletal conditions, 8 chronic conditions, 7 cancer).

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 studies measured health-related quality of life on a scale, and all results were reported as a score on the original scale. Most scales were used in a single study (except as noted below).

Generic measures used were the Short Form Health Survey (SF-36, 3 trials), the World Health Organisation quality of life measure (WHOQOL; older people module), and the Quality of Life Index (QLI).

Condition specific measures in studies of cancer were the European organisation for research and treatment of cancer quality of life questionnaire (EORTC QLQ-C30, 3 trials) and the module for chemotherapy induced peripheral neuropathy (EORTC-CIPN-20), the Functional Assessment of Cancer Therapy-Breast (FACT-B, 2 trials) scale and the Multidimensional quality of life scale – cancer (MQOLS-CA). Other condition specific scales were the Asthma Quality of Life Questionnaire (AQLQ), the Multiple Sclerosis Impact Scale (MSIS-29, 2 trials), the Constipation Quality of Life Scale

(CQLS), the Menopause-specific Quality of Life (MENQOL), the Women's health questionnaire and the King's Health Questionnaire (KHQ) (overactive bladder).

Effects of reflexology compared to an inactive control on health-related quality of life

The effects of reflexology compared to an inactive control on health-related quality of life are presented in Table 4.6.1 the GRADE summary of findings table. The certainty of evidence, and factors that influenced our certainty, are presented and explained in these tables. Study level and meta-analytic results are presented in the forest plot (Figure 4.6.1).

Across conditions, the evidence about the effect of reflexology on health-related quality of life is of very low certainty due to study design limitations, inconsistent effects and bias due to missing results (20 trials, 1575 participants; Table 4.6.1, Figure 4.6.1).

There is low certainty evidence, due to study design limitations and imprecision, that reflexology may improve health-related quality of life for people with other chronic and longer-term conditions (12 trials, 777 participants).

Effects in the following group are very uncertain

- People with cancer and advanced disease (8 trials, 798 participants).

Table 4.6.1 | Summary of findings for the effect of reflexology versus inactive control for 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	Risk with reflexology				
Health-related quality of life: all populations	-	SMD **0.53 SD higher** (0.19 higher to 0.86 higher)	-	1575 (20 RCTs)	⨁◯◯◯ Very low[a,b,c,d]	The evidence is very uncertain about the effect of reflexology on health-related quality of life in general (across all population groups).
Health-related quality of life: cancer & advanced disease	-	SMD **0.54 SD higher** (0.06 lower to 1.13 higher)	-	798 (8 RCTs)	⨁◯◯◯ Very low[a,b,d,e]	The evidence is very uncertain about the effect of reflexology on health-related quality of life for people cancer & advanced disease.
Health-related quality of life: other chronic conditions	-	SMD **0.52 SD higher** (0.04 higher to 0.99 higher)	-	777 (12 RCTs)	⨁⨁◯◯ Low[a,f,g,h]	Reflexology may increase health-related quality of life for people with other 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). The resulting interpretation is as follows.

- For HR-QoL: > 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 are provided for domains for which there is a downgrade or a borderline judgment. In line with GRADE guidance, we not explain that there are no limitations unless the judgment was challenging (https://pubmed.ncbi.nlm.nih.gov/26796947/)

Explanations

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

b. Serious inconsistency (-1). Non-overlapping confidence intervals; effect estimates vary importantly (some studies show moderate to large benefit, while others show little to no effect).

c. No serious imprecision. The 95% confidence interval crosses the threshold for a small but important improvement (SMD of 0.2), so the result is compatible with important benefit (SMD 0.86 higher) and important harm (SMD 0.19 higher), however the extent to which the threshold is crossed is modest and the likely due to inconsistent results for which the results has already been downgraded.

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). For population groups, publication bias was strongly suspected because of the evidence from the funnel plots in combination with a high proportion of small studies showing large, statistically significant effects favouring reflexology (combined effect estimate is moderate to large).

e. Serious imprecision (-1). The 95% confidence interval crosses the threshold for a small but important improvement (SMD of 0.2), so the result is compatible with important benefit (SMD 1.13 higher) and little or no difference (SMD 0.06 lower).

f. No serious inconsistency. Most confidence intervals overlap, so results are compatible despite variable effect estimates (some suggesting important benefit and others a trivial effect).

g. Serious imprecision (-1). The 95% confidence interval crosses the threshold for a small but important improvement (SMD of 0.2), so the result is compatible with important benefit (SMD 0.99 higher) and little or no difference (SMD 0.04 higher).

h. Publication bias not detected. Although there is evidence of selective non-reporting of unfavourable results from a contour enhanced funnel plot for the overall analysis, for this population group more than half the studies show a trivial effect and the majority have statistically non-significant results. This lessens concerns about publication bias for this result.

Forest plots and analysis results

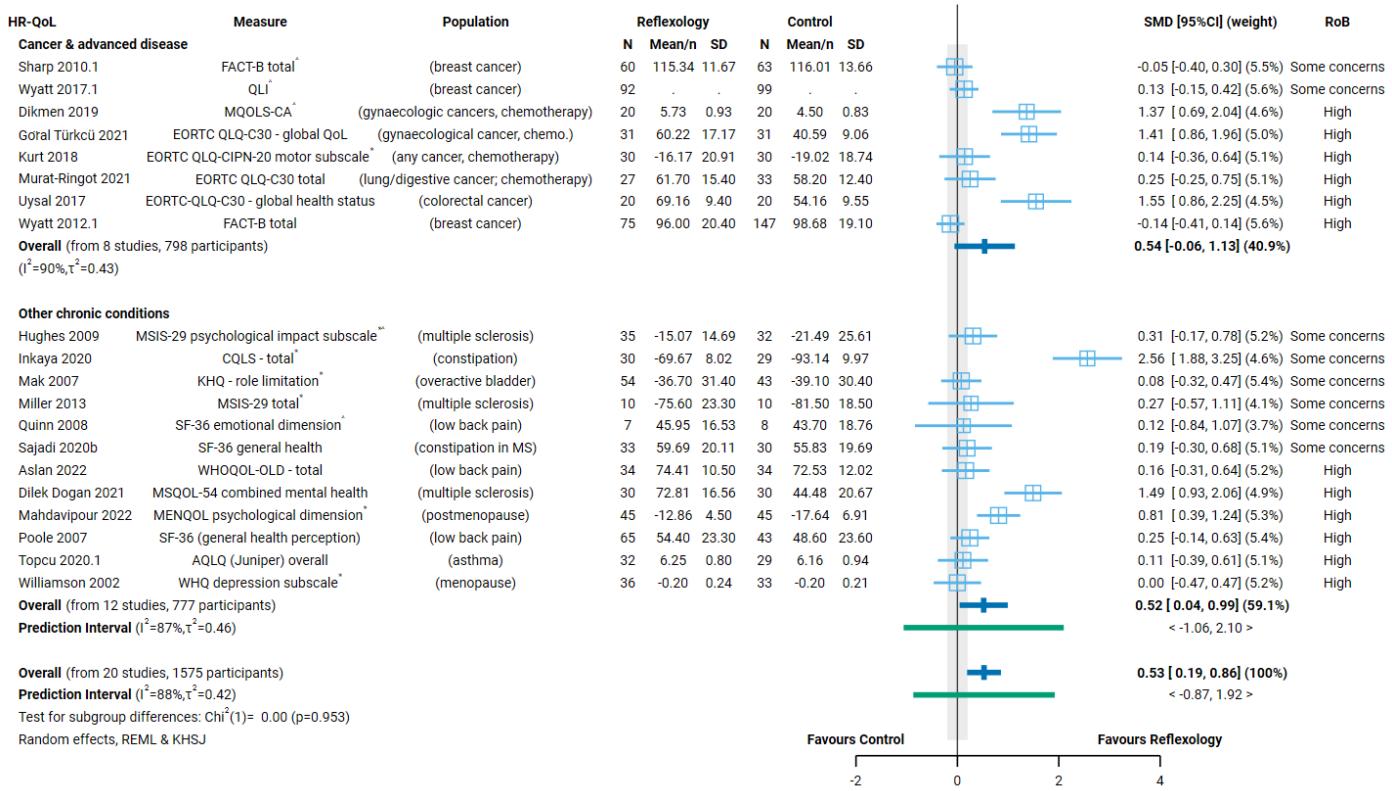

Fig 4.6.1 | Forest plot for main comparison. The effect of reflexology versus inactive control on health-related quality of life. SMD=standardised mean difference. Blue lines show 95% confidence intervals (CI). The shaded grey area indicates the pre-specified range where the effect of reflexology is 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.* Denotes studies for which the direction of effect was changed to match the overall plot (positive numbers are beneficial).

4.7 Physical function

To be considered for the physical function analysis, trials had to administer reflexology to an eligible population for longer-term care (i.e. delivering treatment over weeks or longer, not days) and measure physical function in a timeframe likely to detect meaningful improvement (i.e. not immediately after a single treatment).

Fourteen (14) trials examined the effect of reflexology on physical function, of which 10 (877 participants) were included for meta-analysis and 4 could not be included.

Of the 4 trials (180 participants) that were eligible for this analysis but could not be included.

- One trial (60 participants) reported results that were *unsuitable for meta-analysis* (i.e. the required statistics were unavailable and could not be calculated or imputed)
- 3 trials (120 participants) had results that we counted as 'missing' - either the outcome was measured but results were not reported, or the results were uninterpretable (e.g. information required to interpret the measure or results was missing, or there were errors in reported data).

Characteristics of included studies

Types of populations

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

Included studies examined the effect of reflexology as follows.

Chronic or longer-term conditions

- Cancer and advanced disease (3 trials, 475 participants; breast, gynaecological)
- Chronic musculoskeletal conditions (3 trials, 172 participants; low back pain)
- Other chronic conditions (4 trials, 230 participants; multiple sclerosis, cerebral palsy)

No studies were found for the prioritised condition migraine or headache (chronic or episodic).

Types of interventions

Of the 10 trials included in the meta-analysis, all 10 evaluated foot reflexology.

Session duration, frequency and intervention period. Treatment sessions were 20 to 60 minutes duration.

All 10 trials delivered multiple sessions of reflexology, the treatment period and frequency varied.

- One month or less (1 trial involving people with cancer who received 3 sessions per week over 2 weeks)
- More than a month (9 trials; 3 chronic musculoskeletal conditions, 4 chronic conditions, 2 cancer). In most of these trials, participants received a single session of reflexology per week over 4 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). 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 physical function on a scale, and all results were reported as a score on the original scale.

For five trials, physical function was measured with a subscale of a health-related quality of life measure (SF-36 in 3 trials, EORTC QLQ-C30 in 1 trial, MSQOL-54 in 1 trial). The other five trials each reported results from different measures of physical function, namely the Pregnancy mobility index (PMI), the Gross motor function measure (GMFM), the Modified fatigue impact scale (MFIS), the Oswestry low back pain disability questionnaire (ODQ), and the item bank for measuring physical function from the Patient-reported outcomes measurement information system (PROMIS).

Effects of reflexology compared to an inactive control (usual care, no intervention, sham, co-intervention given in both groups) on physical function

The effects of reflexology compared to an inactive control on physical function are presented in Table 4.7.1 the GRADE summary of findings table. The certainty of evidence, and factors that influenced our certainty, are presented and explained in these tables. Study level and meta-analytic results are presented in the forest plot (Figure 4.7.1).

Across conditions, the evidence about the effect of reflexology on physical function is of very low certainty due to study design limitations, imprecise effects (compatible with both little to no effect and benefit), and bias due to missing results (10 trials, 877 participants; Table 4.7.1, Figure 4.7.1). Effects in the following groups are also very uncertain

- Cancer and advanced disease (3 trials, 475 participants)
- Chronic musculoskeletal conditions (3 trials, 172 participants)
- Other chronic conditions (4 trials, 230 participants)

No studies examined the effect of reflexology on physical function for people with migraine or headache (chronic or episodic).

Table 4.7.1 | Summary of findings for the effect of reflexology versus inactive control for physical function.

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies) contributing to meta-analysis	Certainty of the evidence (GRADE)	Interpretation (evidence statement)
	Risk with inactive control	**Risk with reflexology**				
Physical function: all populations	-	SMD **0.6 SD higher** (0.02 lower to 1.22 higher)	-	877 (10 RCTs)	⨁◯◯◯ Very low[a,b,c,d]	The evidence is very uncertain about the effect of reflexology on physical function in general (across all population groups).
Physical function among people living with cancer and advanced disease	-	SMD **0.93 SD higher** (3.09 lower to 4.95 higher)	-	475 (3 RCTs)	⨁◯◯◯ Very low[e,f,g]	The evidence is very uncertain about the effect of reflexology on physical function among people living with cancer and advanced disease.
Physical function among people living with chronic musculoskeletal conditions	-	SMD **0.34 SD higher** (0.12 higher to 0.57 higher)	-	172 (3 RCTs)	⨁◯◯◯ Very low[h,i,j]	The evidence is very uncertain about the effect of reflexology on physical function among people living with chronic musculoskeletal conditions.
Physical function among people living with chronic conditions	-	SMD **0.56 SD higher** (0.38 lower to 1.49 higher)	-	230 (4 RCTs)	⨁◯◯◯ Very low[h,j,k]	The evidence is very uncertain about the effect of reflexology on physical function among people living 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). The resulting interpretation is as follows.

- For physical function (disability): > 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 are provided for domains for which there is a downgrade or a borderline judgment. In line with GRADE guidance, we not explain that there are no limitations unless the judgment was challenging (https://pubmed.ncbi.nlm.nih.gov/26796947/)

Explanations

a. Serious risk of bias (-1). High proportion of studies are at high risk of bias and show large effects, attenuated to some extent by study/studies showing little to no effect or favouring control. However, serious concerns that the observed benefit may be overestimated.

b. No serious inconsistency. Most confidence intervals overlap, so results are compatible despite effect estimates that suggest different effects (3 studies show little to no difference, the rest show benefit).

c. Serious imprecision (-1). The 95% confidence interval crosses the threshold for a small but important improvement (SMD of 0.2), so the result is compatible with important benefit (SMD 1.22 higher) and little or no difference (SMD 0.02 lower).

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). For population groups, publication bias was strongly suspected because of the evidence from the funnel plots in combination with a high proportion of small studies showing large, statistically significant effects favouring reflexology (combined effect estimate is moderate to large).

e. Serious risk of bias (-1). Two of three studies at high risk of bias, one showing large effects. Serious concerns that the observed benefit may be overestimated.

f. Serious inconsistency (-1). Non-overlapping confidence intervals; effect estimates vary importantly (two showing little to no difference, one showing very large benefit).

g. Very serious imprecision (-2). The 95% confidence interval crosses two thresholds for a small but important effect (SMD of 0.2 and -0.2), so the result is compatible with important benefit (SMD 4.95 higher) and important harm (SMD 3.09 lower).

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

i. Serious imprecision (-1). The 95% confidence interval crosses the threshold for a small but important improvement (SMD of 0.2), so the result is compatible with important benefit (SMD 0.57 higher) and little or no difference (SMD 0.12 lower).

j. 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). Selective non-reporting of unfavourable results (null or favouring control) could importantly change this combined estimate.

k. Very serious imprecision (-2). The 95% confidence interval crosses two thresholds for a small but important effect (SMD of 0.2 and -0.2), so the result is compatible with important benefit (SMD 1.49 higher) and important harm (SMD 0.38 lower).

Forest plots and analysis results

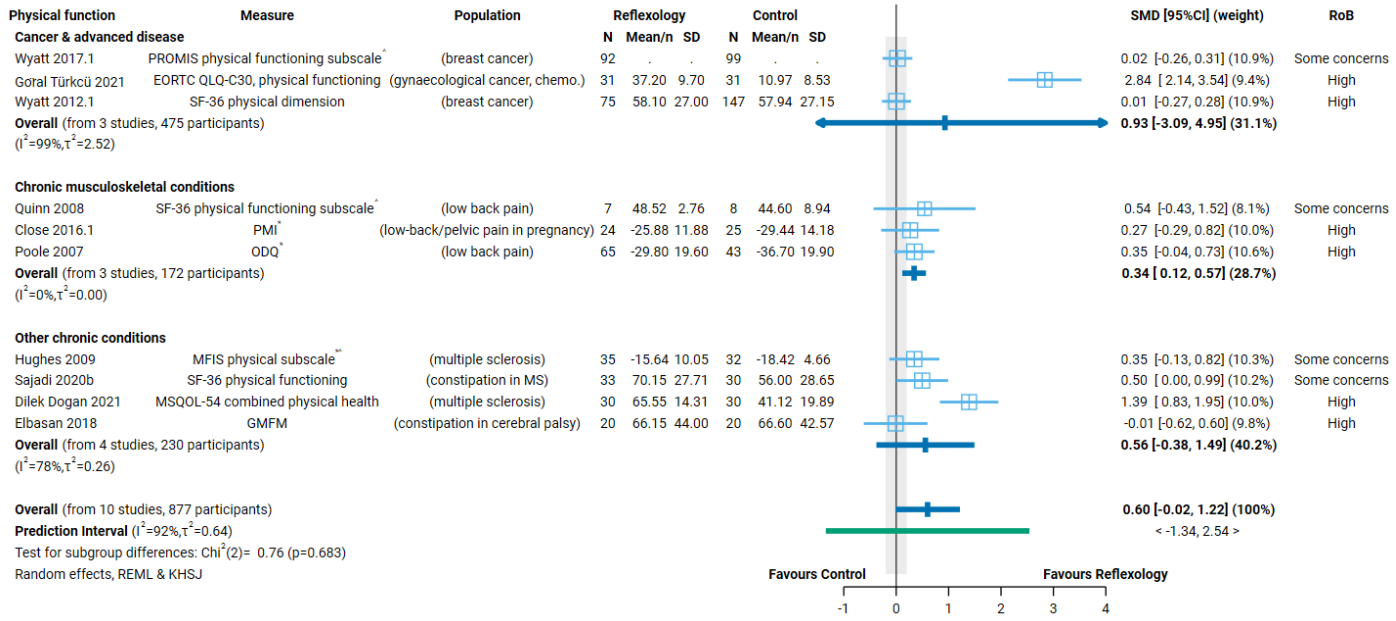

Fig 4.7.1 | Forest plot for main comparison. The effect of reflexology versus inactive control on physical function. SMD=standardised mean difference. Blue lines show 95% confidence intervals (CI). The shaded grey area indicates the pre-specified range where the effect of reflexology is considered to 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.* Denotes studies for which the direction of effect was changed to match the overall plot (positive numbers are beneficial).

4.8 Global symptoms

Twenty-seven (27) trials examined the effect of reflexology on symptoms (global, overall), of which 18 (1284 participants) were included for meta-analysis and 9 could not be included.

Of the 9 trials (475 participants) that were eligible for this analysis but could not be included.

- 3 trials (120 participants) reported results that were *unsuitable for meta-analysis* (i.e. the required statistics were unavailable and could not be calculated or imputed)
- 6 trials (355 participants) had results that we counted as 'missing' - either the outcome was measured but results were not reported, or the results were uninterpretable (e.g. information required to interpret the measure or results was missing, or there were errors in reported data).

Characteristics of included studies

Types of populations

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

Included studies examined the effect of reflexology as follows.

Chronic or longer-term conditions

- Cancer and advanced disease (6 trials, 591 studies; breast, gynaecological, any, lung/digestive, colorectal)
- Other chronic conditions (10 trials, 603 participants; restless leg syndrome – haemodialysis, constipation [multiple underlying conditions], peripheral neuropathy, infantile colic, menopause, premenstrual syndrome)
- Chronic respiratory conditions (2 trials, 90 participants; asthma)

No studies were found for the prioritised condition chronic musculoskeletal conditions.

Types of interventions

Of the 18 trials included in the meta-analysis, all 18 evaluated foot reflexology (17 trials of foot only, 1 trial of foot, hand and ear).

Session duration, frequency and intervention period. Treatment sessions were generally 20 to 30 minutes duration, but in three trials shorter sessions were used (one involving infants with colic, one during pregnancy, one people with asthma).

In all 18 trials, multiple sessions of reflexology were delivered once, twice or three times per week. The treatment period varied.

- One month or less (3 trials; 2 chronic conditions, 1 cancer). In the trial of infant colic, professional reflexology sessions were supplemented by parent-delivered reflexology.
- More than a month (15 trials; 10 chronic conditions, 5 cancer).

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 4.8.1). 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.

Most studies measured symptoms using condition specific scales. The only measures used in more than one study were the M.D. Anderson Symptom Inventory (MDASI, 2 trials), the Constipation Assessment Scale (CAS, 2 trials), International Restless Legs Scale (IRLS, 2 trials), and the EORTC QLQ-C30 (symptoms dimension, 2 trials).

Effects of reflexology compared to an inactive control (usual care, no intervention, sham, co-intervention given in both groups) on global symptoms

The effects of reflexology compared to an inactive control on global symptoms are presented in Table 4.8.1 the GRADE summary of findings table. The certainty of evidence, and factors that influenced our certainty, are presented and explained in these tables. Study level and meta-analytic results are presented in the forest plot (Figure 4.8.1).

Across conditions, the evidence about the effect of reflexology on global symptoms is of very low certainty due to very serious study design limitations and bias due to missing results (18 trials, 1284 participants; Table 4.8.1, Figure 4.8.1). Effects in the following groups are also very uncertain

- Cancer and advanced disease (6 trials, 591 studies)
- Other chronic conditions (10 trials, 603 participants)
- Chronic respiratory conditions (2 trials, 90 participants)

No studies examined the effect of reflexology on global symptoms for people with chronic musculoskeletal conditions.

Table 4.8.1 | Summary of findings for the effect of reflexology versus inactive control for global symptoms.

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies) contributing to meta-analysis	Certainty of the evidence (GRADE)	Interpretation (evidence statement)
	Risk with inactive control	**Risk with reflexology**				
Global symptoms: any population	-	SMD **0.96 SD lower** (1.37 lower to 0.55 lower)	-	1284 (18 RCTs)	⨁◯◯◯ Very low[a,b,c]	The evidence is very uncertain about the effect of reflexology on global symptoms in general (across all population groups).
Global symptoms: people living with cancer or advanced disease	-	SMD **0.42 SD lower** (0.97 lower to 0.13 higher)	-	591 (6 RCTs)	⨁◯◯◯ Very low[d,e,f,g]	The evidence is very uncertain about the effect of reflexology on global symptoms for people living with cancer or advanced disease.
Global symptoms: other chronic conditions	-	SMD **1.38 SD lower** (1.97 lower to 0.8 lower)	-	603 (10 RCTs)	⨁◯◯◯ Very low[h,i,j]	The evidence is very uncertain about the effect of reflexology on global symptoms for people with other chronic conditions.
Global symptoms: chronic respiratory conditions	-	SMD **0.54 SD lower** (6.41 lower to 5.32 higher)	-	90 (2 RCTs)	⨁◯◯◯ Very low[k,l]	The evidence is very uncertain about the effect of reflexology on global symptoms for people with chronic respiratory 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). The resulting interpretation is as follows.

- For global symptoms: < -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 are provided for domains for which there is a downgrade or a borderline judgment. In line with GRADE guidance, we not explain that there are no limitations unless the judgment was challenging (https://pubmed.ncbi.nlm.nih.gov/26796947/)

Explanations

a. Very serious risk of bias (-2). All studies are at high risk of bias or some concerns, with many showing large effects. Serious concerns that the observed benefit may be overestimated.
b. Serious inconsistency (-1). Non-overlapping confidence intervals; effect estimates vary importantly (some showing moderate to large benefit, some little to no effect or favouring control).
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). Studies in this analysis are almost all small studies many show large benefit raising concerns about selective non-reporting of unfavourable results.
d. Serious risk of bias (-1). All studies at high risk of bias or some concerns, such that the observed benefit may be overestimated.
e. No serious inconsistency. Most confidence intervals overlap, so results are compatible despite effect estimates that suggest different effects (2 studies show little to no difference, 3 benefit and 1 favours control).
f. Serious imprecision (-1). The 95% confidence interval crosses the threshold for a small but important improvement (SMD of -0.2), so the result is compatible with important benefit (SMD 0.97 lower) and little or no difference (SMD 0.13 higher).
g. 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). Selective non-reporting of unfavourable results (null or favouring control) could importantly change this combined estimate.
h. Very serious risk of bias (-2). High proportion of studies are at high risk of bias and show large effects. Very serious concerns that the observed benefit may be overestimated.
i. No serious inconsistency. Non-overlapping confidence intervals overlap, but all effect estimates indicate important benefit and only two studies have a CI that is also compatible with little to no difference.
j. 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). Studies in this analysis are all small studies, and almost all show statistically significant benefit raising concerns about selective non-reporting of unfavourable results.
k. Serious risk of bias (-1). Both studies at high risk of bias, one showing a very large effect. Serious concerns that the observed benefit may be overestimated.
l. Extremely serious imprecision (-3). The 95% confidence interval crosses two thresholds for a small but important effect (SMD of 0.2 and -0.2), so the result is compatible with important benefit (SMD 6.41 lower) and important harm (SMD 5.32 higher). The CI is too wide to interpret this result.

Forest plots and analysis results

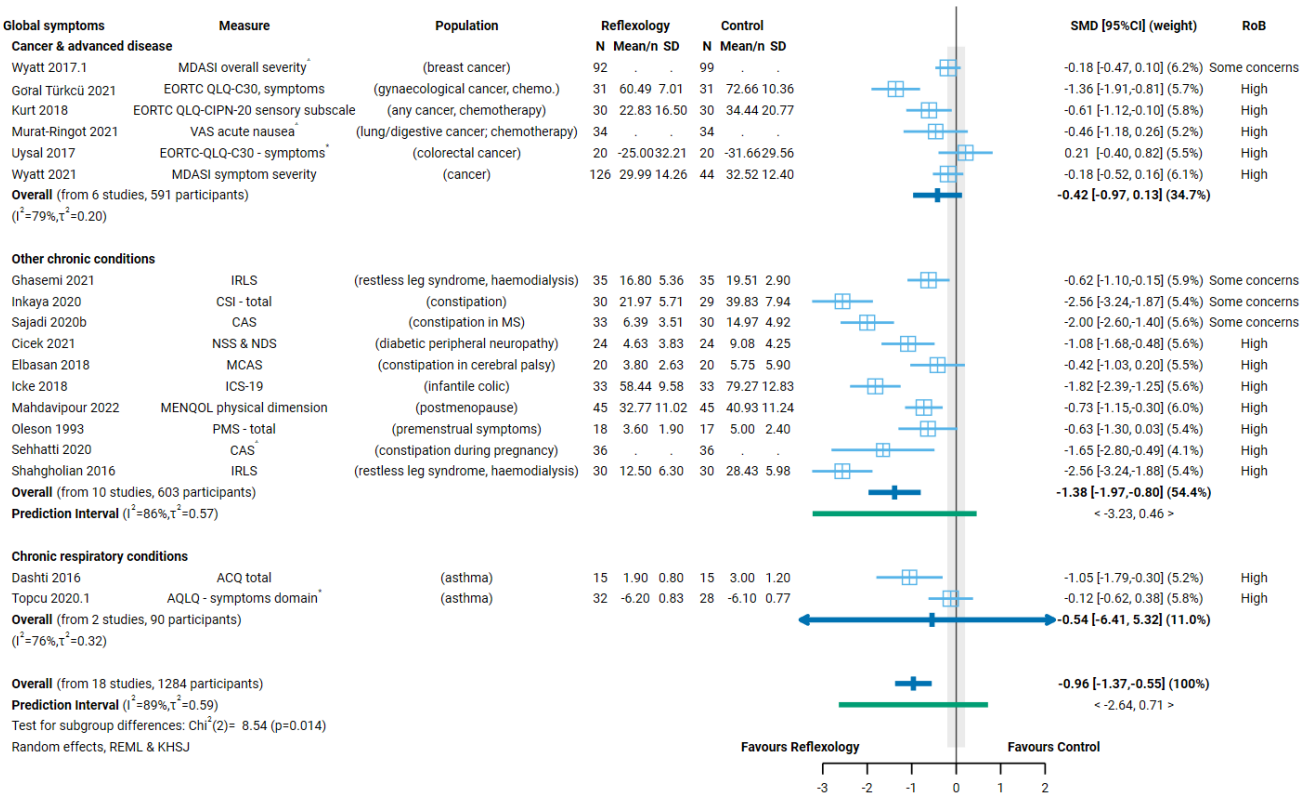

Fig 4.8.1 | Forest plot for main comparison. The effect of reflexology versus inactive control on global symptoms. SMD=standardised mean difference. Blue lines show 95% confidence intervals (CI). The shaded grey area indicates the pre-specified range where the effect of reflexology is considered to be 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. * Denotes studies for which the direction of effect was changed to match the overall plot (negative 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 reflexology, or the reasons practitioners prescribe reflexology and was not intended to inform individual choices about using reflexology.

This systematic review included a large body of research from trials of reflexology. For each outcome, we examined the effects of reflexology 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 reflexology works across multiple population groups or whether any effects might be limited to specific population groups.

Of the 174 studies included in the review, 123 were eligible for at least one synthesis (the remaining 51 studies were included in the evidence inventory). Across all syntheses, we were able to include data from 100 studies. The largest syntheses examined the effect of reflexology on emotional functioning and mental health (3220 participants in 40 trials) and pain (3187 participants in 46 trials).

We found that compared to an inactive control, reflexology:

- may improve sleep quality for people with symptoms of sleep disruption, fatigue for people with chronic conditions, and health-related quality of life for people with chronic conditions, but the evidence is of low certainty meaning that the true effects may be substantially different from the estimated effects.
- has very uncertain effects overall (across conditions) on pain, fatigue, emotional functioning and mental health, health related quality of life, physical function, and global symptoms.
- has very uncertain effects for specific populations
    - pain from surgery, procedures, labour and childbirth, other acute pain, cancer or advanced disease, chronic musculoskeletal conditions or other chronic pain,
    - sleep quality for surgery, hospitalisation (not procedures) or cancer or advanced disease,
    - fatigue for cancer or advanced disease, chronic musculoskeletal conditions or pregnancy,
    - emotional functioning and mental health for surgery, procedures, hospitalisation (not procedures), labour or childbirth, mental distress, cancer or advanced disease, mental disorders, dementia,
    - health related quality of life for cancer or advanced disease,
    - physical function for cancer or advanced disease, chronic musculoskeletal conditions or other chronic conditions.
    - global symptoms for cancer or advanced disease, other chronic conditions or chronic respiratory conditions.

No studies were found comparing reflexology to inactive controls for pain for people with acute musculoskeletal conditions or migraine / headache, sleep for people with chronic insomnia or dementia, or physical function for people with migraine or headache (chronic or episodic).

Comparability of these findings with other systematic reviews

We identified 17 systematic reviews examining the effects of reflexology, all focussed on specific populations or outcomes. Most included fewer studies per outcome than the current review, and a majority used a synthesis method that involved counting statistically significant results (i.e. a method with serious limitations that can lead to wrong conclusions). Given the high proportion of small studies contributing to most syntheses, it is perhaps unsurprising that many reviews that interpreted results based on statistical significance typically described effects on most outcomes as ‘mixed’ (since non-statistically significant results are erroneously counted as not showing benefit or harm).

Of the five reviews among people with cancer, one reported effect estimates for mental health outcomes. The authors found very low certainty evidence for effects on anxiety (2 trials), low certainty evidence of 'no difference' on depression (1 trial) and incomplete reporting in trials of effects on HR-QoL (2 trials) [50]. The other four reviews did not report effect estimates or their synthesis methods (most appear to have used some form of vote counting based on statistical significance), so results are difficult to compare to the current review [33, 34, 51, 52].

Three reviews examined effects during pregnancy, labour and childbirth. A 2018 Cochrane review found no trials on reflexology for labour and childbirth [53]. Two more recent reviews in pregnancy examined effects on pain and anxiety (alongside other outcomes not prioritised in the current review) [54, 55]. In the most recent of these (Yang 2023 [54]), effects on pain and anxiety were similar to our analysis for the labour and childbirth subgroup (noting pain is reported as a mean difference, so effect size is not directly comparable). Their analysis of anxiety included studies at other stages of pregnancy and effects were more consistent. Effects on pain and anxiety were rated as more certain than our results. In part, this is explained by differences in included studies (Yang et al's inclusion of studies at other stages of pregnancy), not rating down for publication bias, and overly precise estimates because the authors appear not to have accounted for including multiple treatment arms from the same study in their meta-analysis, so some studies received too much weight.

One review examined effects on sleep, including 42 randomised trials in English or Chinese [30]. Their meta-analysis of effects on sleep quality included 25 trials (some with multiple arms, unclear if adjusted to ensure appropriate weight). Effects were similar to our overall analysis. The authors did not use GRADE, but the evidence appears to be of very low certainty. All 42 studies were judged to be at high risk of bias and results were inconsistent. Reasons for inconsistency were explored (using meta-regression and subgroup analyses), without finding a credible explanation except that age may modify effects (a finding which needs to be interpreted with caution due to multiple analyses).

Other reviews were not directly comparable, examining effects of reflexology on outcomes for infants and children (1 SR, examined similar outcomes, concluded heterogeneous data and high risk of bias) [56], constipation (1 SR, mixed effects) [38], multiple sclerosis (3 SRs) [57-59], hypertension [60], and pre-menstrual syndrome [61].

Overall completeness and applicability of evidence

There is an extensive body of evidence examining the effects of reflexology 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, in that they are often treated but were not addressed by eligible studies. The evidence includes large numbers of studies evaluating reflexology for acute indications such as control of pain and anxiety perioperatively (21 of 123 trials), periprocedurally (24 trials), and during labour and childbirth (9 trials). There are also large numbers of studies that have examined the effects of reflexology among people living with cancer (24 trials) for whom the use of reflexology for supportive care is of interest [33, 34, 50]. Of the 123 trials included in analyses, 34 were among people with chronic conditions but very few address the same condition and there is considerable diversity in the conditions covered.

The vast majority of studies included in the analysis were conducted in Iran (55 of 123 trials), followed by Türkiye (34), the United Kingdom (9 trials) and the United States of America (7 trials). Other countries in which multiple trials were conducted included Israel (3 trials), and Taiwan, Denmark, and Northern Ireland (2 trials each).

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 reflexology, 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 101 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). Methodological limitations of the included studies included a risk of bias arising from the randomisation 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; 29 of 123 trials 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 to ensure consistency across reviewers and training in the assessment of design features relevant to this review.

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. We did not plan to contact trialists for additional information, and the large number of studies in the review meant it was not feasible to do so, nor was it feasible to review trial registry entries to conduct a comprehensive 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 body of evidence examining the effects of reflexology on health. Despite this, it is not possible to draw conclusions about the effects of reflexology with confidence for any condition or outcome. The uncertainty reflects significant methodological problems with the evidence base. Although an interpretation is made for some results from meta-analyses, the evidence for these results is of low certainty, meaning that the true effects of reflexology 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 reflexology (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 study design limitations and the likelihood that results that show large beneficial effects from reflexology may have been selectively published by trialists, should be considered when deciding whether there is any credible evidence to support the use of reflexology. This review listed, but did not assess studies that compared reflexology to other interventions, so no conclusions can be drawn on whether reflexology is as effective as other interventions. Studies published in a language other than English were listed, but not included in the assessment. The period over which reflexology was delivered varied across studies, although this generally reflected the reasons why reflexology was used (e.g. single treatments for relief of acute effects of surgery and procedures; treatment over a month or longer for chronic conditions). The timing of follow up also varied. In most studies, effects were measured immediately after the end of the reflexology treatment period. Longer-term effects were generally not reported and, as such, were not examined in the review so it is unknown whether any effects are sustained.

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 reflexology 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 reflexology (including, but not limited to, bias arising from the randomisation process, the method of outcome assessment; and the reporting of results). Ensuring trials are registered and reported in accordance with relevant reporting guidelines (most importantly the CONSORT statement [9], but also TIDIER) for reporting interventions [10]) is essential, as is reporting all measured outcomes and avoiding common errors that rendered results unusable. The value of conducting more trials on reflexology would need to be carefully assessed to avoid further research waste.

7. Author contributions and declaration of interest

Sue Brennan[1*] sue.brennan@monash.edu *(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.
Phoebe Nguyen[2]	Appraised studies for risk of bias and extracted data from studies.
Elizabeth Korevaar[2]	Appraised studies for risk of bias and extracted data from studies.
Kimberley Jones[3]	Appraised studies for risk of bias 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. Wrote code for reanalysis of data from crossover 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 2021-22C011.

8. References

1. 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.
2. 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.
3. 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.
4. 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.
5. **American reflexology certification board (ARCB). Legislative and other resources.** [https://arcb.net/other-resources/]
6. Brennan SE, Johnston RV: **Research Note: Interpreting findings of a systematic review using GRADE methods.** *Journal of Physiotherapy* 2023.
7. 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
8. 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
9. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, Elbourne D, Egger M, Altman DG: **CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials.** *Int J Surg* 2012, **10:**28-55.
10. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, Altman DG, Barbour V, Macdonald H, Johnston M, et al: **Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide.** *Bmj* 2014, **348:**g1687.
11. National Health and Medical Research Council: **Statement of requirement: Evidence evaluations for review of natural therapies (tranche two).** 2020.
12. 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.
13. 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.
14. 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.
15. **Association of reflexologists United Kingdom. About reflexology** [https://www.aor.org.uk/what-is-reflexology/]
16. **Reflexology Association of America. Defining reflexology.** [https://reflexology-usa.org/reflexology-is-not-massage/]
17. **Earl E. Bakken Centre for spirituality & healing, University of Minnesota. Taking charge of your health and wellbeing - reflexology** [https://www.takingcharge.csh.umn.edu/reflexology]
18. **Reflexology Association of America. Reflexology is not massage** [https://reflexology-usa.org/reflexology-is-not-massage/]
19. **The Reflexology Association of Australia. Definition** [https://www.reflexology.org.au/]
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. **Australian Traditional Medicine Society (ATMS): About Us** [https://www.atms.com.au/about-us]
23. **The Reflexology Association of Australia. Education** [https://www.reflexology.org.au/home/ethics-and-complaints]
24. **American reflexology certification board. Code of ethics.** [https://arcb.net/about/]
25. **Association of reflexologists United Kingdom. Code of practice & ethics** [https://www.aor.org.uk/wp-content/uploads/2022/02/Code-of-Practice-Jan-2022-accessibility-changes.pdf]
26. **The Reflexology Association of Australia. Code of ethics** [https://www.reflexology.org.au/home/ethics-and-complaints]
27. National Health and Medical Research Council: **Reflexology 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.
28. **National Centre for Complementary and Integrative Health. Reiki** [https://www.nccih.nih.gov/health/reiki]
29. Wang WL, Hung HY, Chen YR, Chen KH, Yang SN, Chu CM, Chan YY: **Effect of Foot Reflexology Intervention on Depression, Anxiety, and Sleep Quality in Adults: A Meta-Analysis and Metaregression of Randomized Controlled Trials.** *Evid Based Complement Alternat Med* 2020, **2020:**2654353.
30. Huang HC, Chen KH, Kuo SF, Chen IH: **Can foot reflexology be a complementary therapy for sleep disturbances? Evidence appraisal through a meta-analysis of randomized controlled trials.** *J Adv Nurs* 2021, **77:**1683-1697.
31. Gok Metin Z, Ozdemir L: **The Effects of Aromatherapy Massage and Reflexology on Pain and Fatigue in Patients with Rheumatoid Arthritis: A Randomized Controlled Trial.** *Pain Manag Nurs* 2016, **17:**140-149.
32. Steel A: **PRACI Study: unpublished data.** 2021.
33. 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.
34. 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.
35. 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.
36. Grant SJ, Hunter J, Seely D, Balneaves LG, Rossi E, Bao T: **Integrative Oncology: International Perspectives.** *Integr Cancer Ther* 2019, **18:**1534735418823266.
37. **NIH National Center for Complementary and Integrative Health. Reflexology** [https://www.nccih.nih.gov/health/reflexology]
38. Yaqi H, Nan J, Ying C, Xiaojun Z, Lijuan Z, Yulu W, Siqi W, Shixiang C, Yue Z: **Foot reflexology in the management of functional constipation: A systematic review and meta-analysis.** *Complement Ther Clin Pract* 2020, **40:**101198.
39. **Association of Reflexologists: Information for Health Professionals** [https://www.aor.org.uk/healthcare-professionals/]
40. 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.
41. **The Reflexology Association of Australia. Definition. Available from: https://www.reflexology.org.au/ Access date: 03 Feb 2022** [https://www.reflexology.org.au/]
42. 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.
43. 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.
44. 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.

45. 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.
46. 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.
47. 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.
48. 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.
49. 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.
50. 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.