Cochrane Australia

Systematic review of evidence on the clinical effectiveness of Buteyko

Technical report prepared by Cochrane Australia

17 December 2024

Trusted evidence.
Informed decisions.
Better health.

Contents

Scope of the technical report

This Technical Report includes a complete description of the methods for the review (Appendices A, B and G), results of the search and prioritisation process (Appendix A), citations for studies included in the evidence synthesis (Appendix D), risk of bias assessments for studies contributing to meta-analyses (Appendix F) and abbreviations used in the report (Appendix I).

It also includes an overview of Appendices C and E which are listed below but presented in separate files.

Appendices contained in this file are in light grey rows. Those in separate files are in blue rows.

Appendix A. Study eligibility criteria, identification and selection
Appendix B. Data collection, analysis and interpretation of findings
Appendix C. Lists of excluded studies, public submissions, studies awaiting classification, ongoing studies (1 file)
Appendix D. Citations for studies included in the evidence synthesis
Appendix E. Characteristics of studies included in the evidence synthesis (2 files)
Appendix F. Risk of bias assessments for studies contributing to meta-analyses (1 file)
Appendix G. Differences between the protocol and the review and methods not used
Appendix H. Response to methodological review
Appendix I. Abbreviations and list of measures

Appendix A. Study eligibility criteria, identification and selection

Overview of Appendix A

Appendix A is comprised of Appendices A1-A7 (below). These Appendices report the methods (grey rows) and results (blue rows) from the first four stages of the review (Figure A, 1-4). These stages encompass the initial specification of questions to be addressed in the synthesis and criteria for including studies in the review, the specification and implementation of search methods, and the selection of studies. From this set of studies, we compiled information about the populations and outcomes addressed in studies eligible for the review. This information was reviewed by the NHMRC, NTWC and NTREAP in order to confirm populations and outcomes for inclusion in the evidence synthesis.

Appendix A1. Review questions and criteria for considering studies for this review
Appendix A2. Search methods for identification of studies
Appendix A3. Methods for selecting studies
Appendix A4. Results of the search
Appendix A5. Prioritisation process: methods used to refine the questions addressed in the synthesis
Appendix A6. Final framework: synthesis questions and criteria for including studies in each synthesis
Appendix A7. Summary of inclusion decisions based on the final framework

Appendices A1-A3 and A5 report the pre-specified methods from the protocol endorsed by NTWC, prospectively registered on the International prospective register of systematic reviews (PROSPERO ID CRD42023467144). Appendix A6 reports the framework that resulted from the prioritisation process shown in Figure A and described in Appendix A5. The framework was finalised prior to commencing data extraction (Figure A, panel 5). 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.

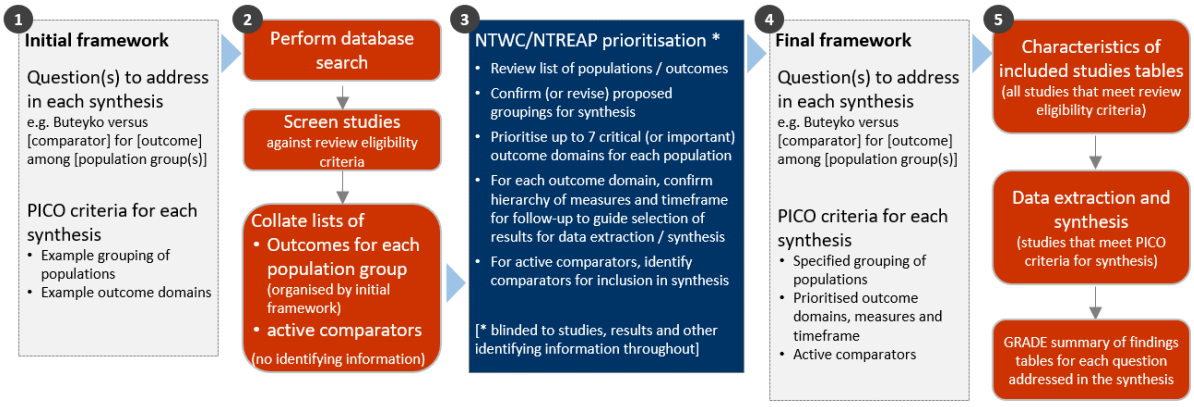

Fig A | Staged approach for developing the questions and analytic framework for this review.

Appendix A1. Review questions and criteria for considering studies

The overall objective of this systematic review is to examine the evidence for the clinical effectiveness of the Buteyko Method in preventing and/or treating injury, disease, medical conditions or preclinical conditions [1]. The questions for the review follow (framed as primary and secondary objectives). An initial analytic framework for the review was presented in the protocol to illustrate the breadth of questions and a possible structure for the synthesis, with indicative populations and outcome domains (Figure A1.1). The framework was refined through the prioritisation process (described in Appendix A5) leading to the final framework and criteria for including studies in the synthesis (Appendix A6). Outcomes listed in the objectives were agreed through the prioritisation process.

Primary objective was to answer the following question

1. What is the effect of *the Buteyko Method* compared to inactive control (no intervention, sham, placebo, wait list control, or a co-intervention offered to both groups, or continuation of usual care) on outcomes for each underlying condition, pre-condition, injury or risk factor?

Secondary objective

Secondary objectives related to the following questions:

2. What is the effect of the Buteyko Method compared to evidence-based treatments (active comparators) on outcomes for each underlying condition, pre-condition, injury or risk factor?
3. What evidence exists examining the effects of the Buteyko Method compared to other active comparators? (for inclusion in evidence inventory only, not the synthesis)

As per protocol, for objective 2, pre-specified criteria needed to be met to proceed with synthesis. That is, at least two low risk of bias studies with comparable population, evidence-based comparator and outcomes. Where the criteria were not met, studies that only contributed an active comparator were included in the inventory.

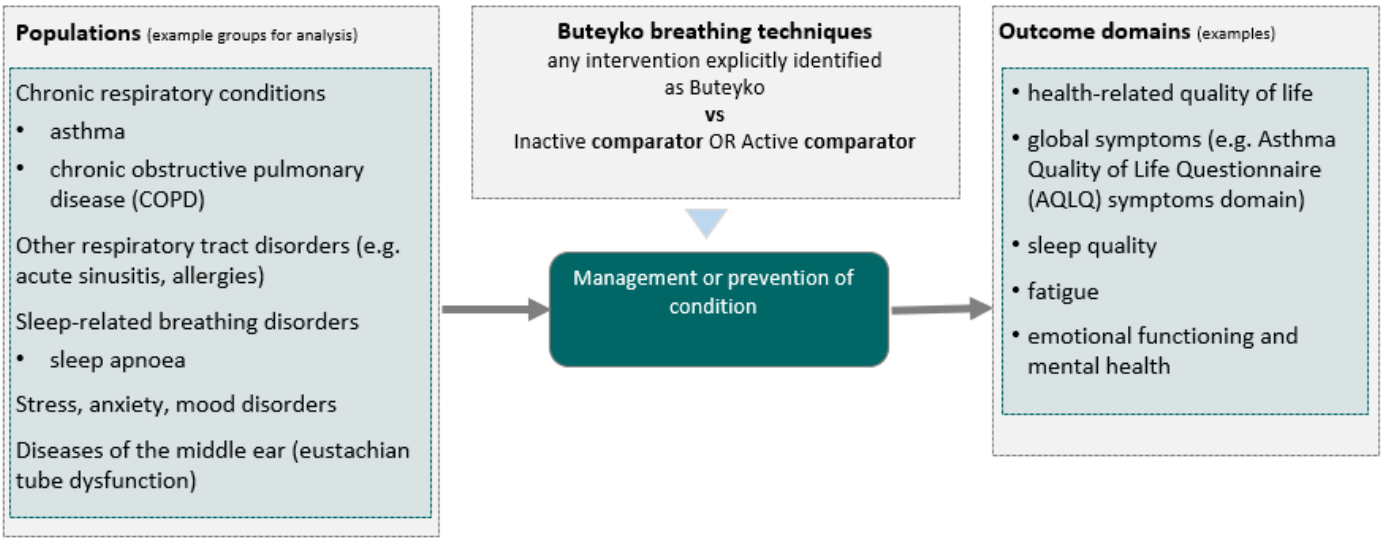

Fig A1.1 | Initial analytic framework for the review showing example population groups and outcome domains for the Evidence Synthesis. The framework was informed by research on the outcomes (and underlying conditions) for which the Buteyko Method is commonly sought or prescribed, the wider literature on the Buteyko Method, and consideration of frameworks for classifying disease and outcomes [2, 3].

A1.1 Criteria for considering studies for this review

A1.1.1 Types of studies

We included randomised controlled trials (RCTs) (including individually and cluster randomised, and cross-over trials).

Controlled trials in which the allocation sequence did not include a truly random element, was predictable, or was not adequately concealed from investigators were eligible as long as there was an attempt to have some kind of 'randomisation' to groups. Examples included studies that used methods for sequence generation based on alternation, dates (of birth or attendance at a clinic) and patient record numbers [4].

Non-randomised studies of interventions (NRSIs) with specific design features that are suitable for estimating a causal effect were eligible for inclusion in the review, in line with current Cochrane guidance. While study design labels were used as an aid to communicating about eligible designs and for use in the review, eligibility decisions were based on assessment of the specific design features of each study rather than the label used by the study authors (see checklist Appendix 2 in protocol published on PROSPERO CRD) [5, 6].

Eligible non-randomised study designs were those in which the following features are present.

- The intervention may be allocated to individuals or clusters. We anticipated that the Buteyko Method (or the control) would be allocated to individuals in most studies, although clustering was likely in these studies given the way in which the Buteyko Method lessons are delivered (i.e. the same teacher may deliver the intervention to multiple participants) [7].
- Treatment groups may be formed by some action of the researchers or in the course of usual treatment decisions (including healthcare decision makers, practitioners or participants/patients/peoples' choices).
- Studies must include a contemporaneous control.
- There must be an attempt to control for confounding (either by using methods that control in principle for confounding or that control for observed covariates)
- The design must be suitable for estimating a causal effect.

We excluded:

- 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 not eligible for synthesis. In accordance with the protocol, these studies were to be included in the list of studies 'Awaiting classification' and coded according to whether they were likely to be eligible or whether eligibility could not be determined.

A1.1.2 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 (e.g. biomedical, health behaviours, or other). There were no restrictions on age or other demographic factors.

For trials in which the Buteyko Method was used for primary or secondary prevention, participants must have had a clearly-identified factor that put them at heightened risk of the condition or injury that the intervention is intended to prevent compared to the population at large. Where possible, decisions about whether a population was at risk was informed by evidence from a systematic review of risk factors.

We operationalised the criteria for risk as follows:

- The risk factor(s) for the condition that the Buteyko Method was used to prevent was part of the eligibility criteria for the trial or reported in the baseline data (e.g. older age in a trial aimed at preventing falls; work that involves demanding posture or repetitive movement in a trial aiming to prevent workplace-related musculoskeletal conditions), and
- There was a direct link between the risk factor and the trial outcomes (i.e. an outcome that demonstrates progression to a diagnosable condition or pre-condition; musculoskeletal pain or injury in a trial that aims to prevent injury)

We expected that studies would include participants within broad population groups, such as those shown in **Figure 2**. These were indicative groups, included to illustrate the breadth of populations eligible for the review and possible groupings for synthesis. Decisions about which groups to include in the final analytic framework were made through the prioritisation process (**Figure 1**).

Exclusions. Healthy populations seeking health improvement.

Studies that included both healthy participants and participants eligible for the review, were to be included if separate data were available or a majority of participants met the review eligibility criteria as per guidance in the Cochrane handbook [8]. No such studies were identified.

While studies involving any population were to be included in the review (except for the specific exclusions above), if the number of eligible studies for synthesis was unmanageable, the synthesis could be limited to populations (conditions) most relevant to the use of the Buteyko Method in Australia. Population prioritisation was not needed for this review.

A1.1.3 Types of interventions

For the purpose of this review, the Buteyko Method was defined as a " breathing retraining technique that may include a range of specific breathing techniques taught by a therapist…with the aim of returning breathing to normal physiological levels [and providing] relief and prevention of symptoms" [9].

Because of the potential challenge of distinguishing components of the Buteyko Method from related modalities (especially other systematised breathing interventions that use similar techniques to the Buteyko Method), and the likelihood of identifying studies where the defining techniques and principles of the Buteyko Method are incompletely reported, studies were included if:

- the therapy was described as 'Buteyko', or
- it was implicit that the therapy was the Buteyko Method (e.g. a Buteyko therapist teaches the breathing techniques).

It was expected that the majority of studies would involve participants undertaking education in the Buteyko Method techniques. Except for the specific exclusions below, the Buteyko Method interventions were eligible irrespective of:

- whether the study examines the effects of undertaking a series of educational sessions or the routine use of the Buteyko Method,
- the specific breathing techniques used by the therapist,
- mode of delivery (individual or group; face-to-face or virtual),
- whether the intervention was guided by a teacher or self-directed (the latter possible when trained individuals use the Buteyko Method in daily life),
- the training or qualifications of the teacher or practitioner,
- the setting in which the Buteyko Method is taught or used,
- the dose and duration of treatment, or
- whether or not the therapy includes posture and lifestyle interventions (if identified in the trial as 'usual practice').

Comparisons

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

2. The Buteyko Method *versus versus* evidence-based treatment(s) (see below).

3. The Buteyko Method *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).

Comparisons 1 and 2 were to be addressed in separate syntheses (meta-analyses). Where a study included multiple arms, with at least one eligible comparator (e.g. a placebo control arm), we included the eligible comparison(s).

For comparison 2, active comparators were listed but did not contribute to the synthesis because the criteria for synthesis were not met (at least two low risk of bias studies with the same comparator, population and outcome).

Characteristics of studies involving active comparators are briefly described in the evidence inventory of available evidence (Appendix E3).

Exclusions. In line with the main review objective, which was to examine the effects of the Buteyko Method rather than the comparative effects of different implementations of the Buteyko Method, we excluded head-to-head comparisons of the Buteyko Method from the review (see exceptions below). For example, we excluded studies where the only comparator is:

- a different dose (frequency, duration, schedule or combination thereof) of the Buteyko Method (e.g. different numbers of lessons)
- a different mode of delivery of the Buteyko Method (e.g. individual versus group),
- where the person teaching the Buteyko Method has a different qualification, or level of experience (e.g. specialist teacher versus other health professional with teacher training),
- or combinations of the above.

A1.1.4 Types of outcomes

We considered for inclusion in the review any outcome that aligned with the reasons why the Buteyko Method is sought by patients and prescribed by practitioners. In principle, this could include any patient-important outcome that helps elucidate the effects of the Buteyko Method on an underlying condition or its symptoms, recovery, rehabilitation, or prevention of disease among people with specific risk factors or pre-conditions. Example outcome domains were shown in the initial analytic framework to illustrate the breadth of outcomes likely to be relevant across a wide range of conditions (Figure A1.1). The outcome domains were based on ICD11 codes and the COMET outcome taxonomy [2, 3]. These systems provide a widely agreed and understood structure for categorising different outcomes.

Studies were included in the review irrespective of the outcome(s) measured, but the summary and synthesis was limited to outcomes considered to be critical or important for each population group. Outcomes for inclusion in the synthesis were determined through the prioritisation process described in Appendix A5.

The outcome domains determined to be critical or important for the synthesis were as follows (see Appendix A6 and Figure A6.1 for details).

- health-related quality of life
- global symptoms / overall disease status
- physical function (activity limitations)
- lung function (where relevant for the condition)
- emotional functioning and mental health
- breathing patterns, respiration and physiological signs and symptoms (e.g. blood pressure)
- healthcare resource use (including exacerbations requiring an emergency department visit, medication use)
- pain (where relevant for the condition)

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, timing of outcome measurement and data reported as follows.

Outcome measures. For each of these outcome domains, we considered for inclusion any measure of the outcome. Where studies reported multiple outcomes within an outcome domain, we used a population-specific hierarchy of outcomes measures to select the most relevant and valid outcome. The hierarchy of measures was proposed by the review team and agreed through the prioritisation process.

Outcome timing. Where trials reported outcomes measured at multiple timepoints, we selected the first measurement taken after the end of the Buteyko Method intervention period (i.e. if administered three times over a week, we took the first measure after the third administration).

Data reported

- When authors reported results for both change scores (change from baseline) and post-intervention (final) values, we selected results for final values.
- If data for the preferred measure was incompletely reported or uninterpretable, we selected another measure.

Excluded outcomes. experience of care (e.g. satisfaction), safety, quality, and economic outcomes.

Appendix A2. Search methods for identification of studies

A2.1 Electronic searches

Studies were sought from the following databases: Cochrane Central Register of Controlled Trials (Cochrane Library, Issue 10, 2023), MEDLINE (Ovid), Embase (Ovid), Emcare (Ovid), AMED (Ovid), CINAHL (EBSCOhost) and Europe PMC. In addition, we searched two clinical trial registers for reports of ongoing or unpublished studies (ClinicalTrials.gov and WHO International Clinical Trials Registry Platform).

The search strategy comprised the text word 'buteyko' and, where available, the relevant subject heading term. No study design filter was applied. Searches were run on 6 October 2023 and were not limited by language, year of publication or publication status (see Appendix A4).

A2.2 Searching other resources

The 7 randomised trials included in the 2015 evidence evaluation for Buteyko were cross-checked against records retrieved by the search and considered for inclusion. Since there is no Medical Subject Heading (MeSH) term for the Buteyko Method, we screened the included studies of systematic reviews indexed in MEDLINE with Breathing Exercises as a major MeSH term published from 2020 onwards. Only two of the 25 systematic reviews included studies of the Buteyko Method. One of these reviews was the 2020 update of the Cochrane review of breathing exercises for adults with asthma. For this review, we checked studies listed as included, excluded, awaiting assessment and ongoing. The reference lists of 14 other systematic reviews retrieved by our search were also checked.

We searched the first 10 pages (100 entries) of Google Scholar using the phrases “Buteyko breathing” or “Buteyko method”. We also checked references to research listed on the Buteyko Health and Breathing website (buteykoairways.com.au).

We examined the reference lists of included studies and systematic reviews retrieved by our search to identify additional trials (i.e. backward citation searching), but we did not conduct forward citation searching (i.e. looking for studies that have cited included studies).

Finally, we searched PubMed for retracted publications, expressions of concern and published errata, as well as the Retraction Watch database but identified nothing of relevance.

A2.3 Public submissions

No additional citations were received through the public call for submissions, or from NTREAP or NTWC.

Appendix A3. Methods for selecting studies

A3.1 Selection of studies

Records from CENTRAL, PubMed, AMED and Emcare were imported into EndNote and duplicates removed. All remaining records were imported into Covidence for screening.

Two reviewers (MM, SM) piloted guidance for title and abstract screening on a sample of 50 records to ensure the eligibility criteria were applied consistently. All records were reviewed independently by two reviewers at both the title and abstract screening and full-text review stages in Covidence. Disagreements at either stage were resolved by consensus among members of the review team. Advice from NTWC regarding inclusion was sought if required (one study on mouth taping only).

While screening full-text study reports in Covidence, we extracted the trial register and registry record number (if reported) into notes in Covidence. On completion of study report screening, we matched any registry record details in the included study notes (e.g. registry record number) with the registry records search results.

Unmatched registry records were then screened to identify potentially eligible trials for which there was no published report to include in a list of 'ongoing studies' (Appendix C4) and for assessment of bias due to missing results (B1.6).

Published protocols for studies confirmed as meeting the eligibility criteria, but for which results were not available in a published report, were checked against potentially eligible trials identified from registry records and included in the list of 'ongoing studies' (Appendix C4). These were also considered in the assessment of bias due to missing results (B1.6)

The following categories of studies were to be included in a list of 'studies awaiting classification', if identified:

- Studies that were only published as abstracts or for which a full report was not available (i.e. we did not seek further information from study authors to confirm eligibility).
- Studies for which a full report was available but the report was incomplete or ambiguous such that eligibility based on one or more PICO criteria or study design could not be confirmed.
- Studies confirmed as likely to be eligible, but for which no English language translation of the full-text publication was available.
- Studies for which eligibility could not be confirmed following translation of the title and abstract using Google translate (Figure A3.1.1)
- Studies for which there were concerns about data that could not be resolved from full report(s) (e.g. where there were important discrepancies in study characteristics or data reported across multiple publications from the same study).

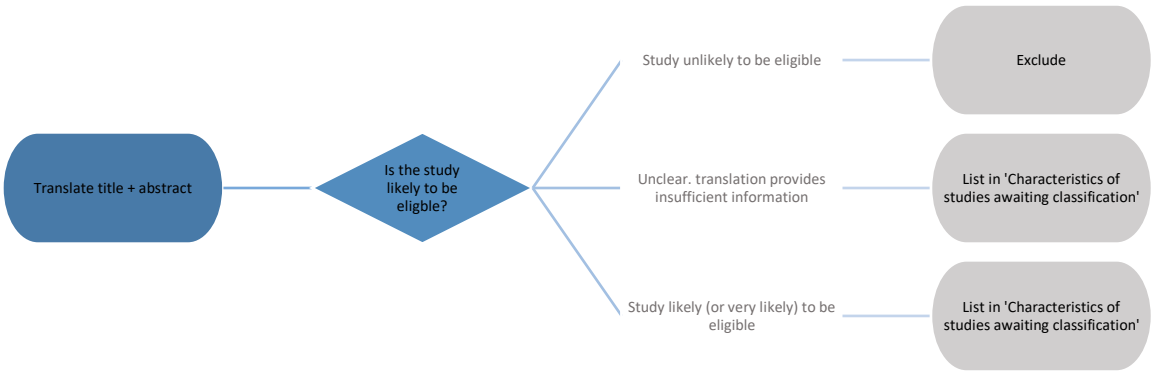

Fig. A3.1.1 | Flowchart showing handling of studies in languages other than English (reproduced from NHMRC framework for natural therapies systematic reviews [10]).

Studies that did not meet the eligibility criteria were excluded and the reason for exclusion was recorded at full-text screening. These studies are included in a list of excluded studies in which the reason for exclusion is reported (Appendix C1).

The search and study selection steps are summarised in the PRISMA flow diagram in Appendix A7.

For studies that originated from the call for evidence, we planned to record and report exclusion decisions irrespective of whether the study was excluded during title and abstract screening or full text review. However this did not arise as none were submitted.

Dealing with duplicate and companion publications

Multiple publications to the same study (e.g. protocols, trial registry entries, trial reports) were identified and linked at the study selection stage in Covidence. Identification and linking of multiple reports were also checked at data extraction in REDCap [11, 12]. Each study was given a unique identifier and all linked records are cited in the final report. Records were matched using trial registry numbers.

Dealing with multiple study IDs

If multiple study reports resulted in the same study ID (Author Surname, Year) and were reporting the same study, the study ID for index report was given the suffix '.1' after the Year (e.g. Ziyaeifard 2017.1), and the study ID for the secondary report was given the suffix '.2.' (e.g. Ziyaeifard 2017.2).

If multiple study reports resulted in the same study ID (Author Surname, Year) and were reporting different studies, the study IDs for each study were given the suffix 'a', 'b', etc after the Year (e.g. Ebrahimi 2021a, Ebrahimi 2021b) to differentiate them.

Appendix A4. Results of the search

Bibliographic databases

The search of bibliographic databases retrieved 296 records. After 154 duplicates were removed in EndNote and Covidence, 142 records were screened at title/abstract. The search strategies for each database are given below. The PRISMA flow diagram in Appendix A7 summarises inclusion decisions following title/abstract screening.

Other sources

The 'other sources' we searched resulted in identifying 30 potentially eligible studies that were not retrieved by the database searches: Google Scholar (n=16); published systematic reviews (n=3); reference checks of included studies (n=9) and Buteyko Health and Breathing website (n=2).

Trial register records

From trial registry entries (CENTRAL, ClinicalTrials.gov and WHO ICTRP), after removing duplicates we identified 21 potentially eligible trials for the review. Of these:

- 2 records were linked to trials included in the review (Jain 2023, Vagedes 2021 [13, 14])
- One was a trial published after the end date of the search (https://clinicaltrials.gov/study/NCT03098849)
- 13 were trials that we judged likely to be ongoing or in the publication process (registered 2020 or later)
- 5 were registered prior to 2020 and were assessed as potentially missing studies.

2015 evidence evaluation for Buteyko

The 2015 overview of Buteyko identified 2 systematic reviews (Burgess 2011, O'Connor 2012 [15, 16]) that included 7 randomised trials. Six (6) trials were retrieved by our search; one (Slader 2006) was added to Covidence for screening.

Public submissions

No submissions were received via the Department's public call for evidence.

Retractions and published errata

No errata were retrieved from PubMed or records from the Retraction Watch database.

Search strategies

Cochrane Central Register of Controlled Trials (Issue 10, 2023)

#	Search strategy	Records
1	(buteyko*):ti,ab,kw	62

MEDLINE ALL (Ovid) 1946 to October 4

#	Search strategy	Records
1	buteyko*.af. [af=all fields]	44

Embase Classic+Embase (Ovid) 1947 to October 4

#	Search strategy	Records

1	Buteyko Breathing/ or buteyko*.af.	81

Emcare (Ovid) 1995 to 2023 week 39

#	Search strategy	Records
1	Buteyko Breathing/ or buteyko*.af.	31

AMED (Ovid) 1985 to September 2023

#	Search strategy	Records
1	buteyko*.af.	19

CINAHL Plus (EBSCOhost)

#	Search strategy	Records
1	SU Buteyko Method OR TI buteyko* OR AB buteyko* OR MW buteyko*	52

Europe PMC

#	Search strategy	Records
1	(TITLE:"buteyko*") OR (ABSTRACT:"buteyko*")	46

ClinicalTrials.gov and WHO ICTRP

Buteyko (limited to Intervention/treatment in ClinicalTrials.gov (n=18) and limited to Intervention in WHO ICTRP (n=28))

Google Scholar

Advanced search options (phrase in title): “Buteyko breathing” and “Buteyko method” (first 10 pages)

Appendix A5. Prioritisation process: methods used to refine the questions addressed in the synthesis

The final synthesis questions, and criteria for including studies in each synthesis, were decided through the pre-specified prioritisation process (Figure A). The process was designed to minimise bias in the selection of results for inclusion in the synthesis and ensure coverage of populations and outcomes relevant to the Australian context. All information provided to NTREAP, NTWC and the NHMRC was de-identified and presented in aggregate form so that it was not possible to identify the studies (no bibliographic information, titles etc). No information was provided about the number of studies, number of participants, methodological quality of studies or results.

Prioritisation of populations and grouping of conditions for the summary and synthesis

There was no need to limit populations in this review, so the provision in the protocol to prioritise populations (conditions) for inclusion in the synthesis was not implemented. NTWC endorsed the proposal to structure the synthesis by the population groups outlined in the analytic framework.

Prioritisation and selection of outcomes for the synthesis

For each population, we collated information about the outcomes addressed in all eligible studies. The purpose was twofold: (1) to enable prioritisation of the most important **outcome domains** for each population (irrespective of whether studies measured these domains), and (2) to facilitate selection of the **most relevant results** from each study.

Prioritisation of outcome domains

- All outcomes and outcome measure were listed under an **outcome domain** from the initial analytic framework for the review (Figure A1.1). For outcomes not covered by the initial framework, additional outcome domains were specified allowing categorisation of all outcomes and measures.
- For each condition, NTWC, with input from NTREAP, rated **outcome domains** as critical, important or of limited importance for understanding the effects of the Buteyko Method on each population group. The intent was to identify up to seven outcome domains for which results would be reported.
- Only critical and important outcome domains were considered in the summary and synthesis.

Outcome selection. From each study, we selected one result per outcome domain for data extraction, risk of bias assessment and reporting of results in the summary and synthesis (using the standardised mean difference to combined effects measured on different scales see B1.2 and B2.1). Selecting one result per study for inclusion in each analysis ensures that individual studies do not receive too much weight. In addition, we aimed to ensure that all studies that should contribute to each synthesis were included.

Overall, the approach deals with multiplicity of results that arises when

(1) the outcomes and measures of outcome domain vary across studies;
(2) individual studies report results for multiple outcomes, measures and timepoints within an outcome domain (e.g. for HR-QoL, reporting an overall score and subscale scores for specific domains of HR-QoL).

To determine which results to select the following was done.

- For each outcome domain, we presented an initial ranking of all outcomes and measures. Where available, the ranking was informed by recommendations in core outcome sets, outcome hierarchies in published Cochrane reviews, and systematic reviews of outcome measures (i.e. to establish relevance, validity, and reliability).
- The NTWC considered the ranking and either confirmed or reranked the outcomes and measures.
- The highest ranked outcome/measure was selected from each study for each outcome domain.
- If data for the highest ranked outcome/measure could not be included in the analysis (e.g. due to incomplete reporting of data), this was reported and the next highest ranked outcome was selected (and so forth).
- Where an outcome measure was potentially eligible for more than one outcome domain, we selected the measure that enabled us to include a study in the largest number of syntheses (e.g. if a study reported scores for the psychological and physical domains of a HR-QoL measure, but no measure of emotional functioning and mental health (EFMH), we chose the physical domain for HR-QoL and the psychological domain for EFMH).

Appendix A6. Final framework for summary and synthesis

Figure A6.1, 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). We included all eligible studies in the summary and synthesis (i.e. no limitations by population or condition).

Prioritised outcomes and comparisons

The outcome domains specified in the initial analytic framework were endorsed. Other symptoms were not included in the list of domains rated by NTWC, but are retained in the framework because of their relevance to populations/conditions for which no studies were found. Where trials measured outcomes at multiple timepoints, we selected the first measurement after the end of the intervention period or closest to.

Because there were few studies, we broadened criteria for inclusion of outcomes to include HR-QoL and physical function outcomes irrespective of population (i.e. not limited to chronic or longer-term conditions), duration of intervention period (i.e. not limited to weeks or longer) and length of follow-up (not limited to time-frames likely to detect meaningful improvement).

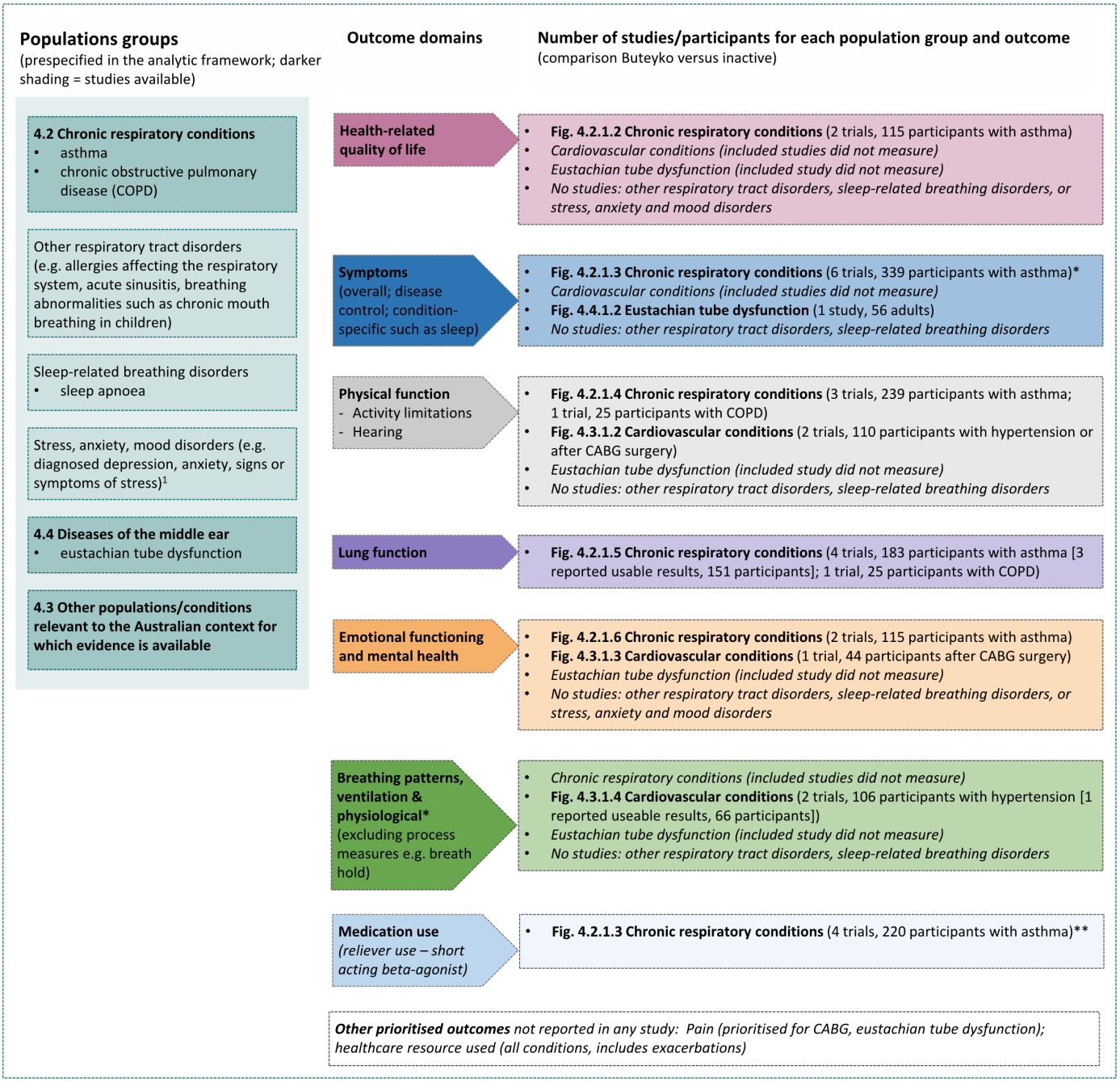

Fig A6.1 | Final analytic framework for the review as agreed through the prioritisation process (Appendix A5). Panel A, columns 1 to 3 show the populations[1], and outcome domains eligible for each comparison in the evidence synthesis. Column 3 also shows the number of studies and participants available to address each synthesis question (as defined by the PICO). Results are reported for each population group in the section indicated in column 1. Study-level data and meta-analyses are presented in the forest plot indicated in column 3. Population groups are those identified from various sources as treated with the Buteyko Method (as per Background); no PRACI data for this therapy. *Outcome domains combined for presentation in framework (not analysis) because similar measures were listed within (e.g. respiration rate). ** Medication use (reliever use – short acting beta-agonist) was included in the symptoms domain to avoid double counting data from 3 of the 4 trials that reported medication use based on a scale in the same measure used for overall asthma symptoms.

[1] Following completion of the review, a registered trial of the Buteyko Method for anxiety was identified as having been published. The trial did not include the word Buteyko in the title, abstract or keywords. The trial has not been integrated in the results, but results are presented in a summary of findings table in Appendix J.

Appendix A7. Summary of inclusion decisions based on the final framework

The flow of studies through the review is in Figure A7.1, the PRISMA flowchart. Inclusions for each synthesis are reported in Figure A6.1 and described in the main report.

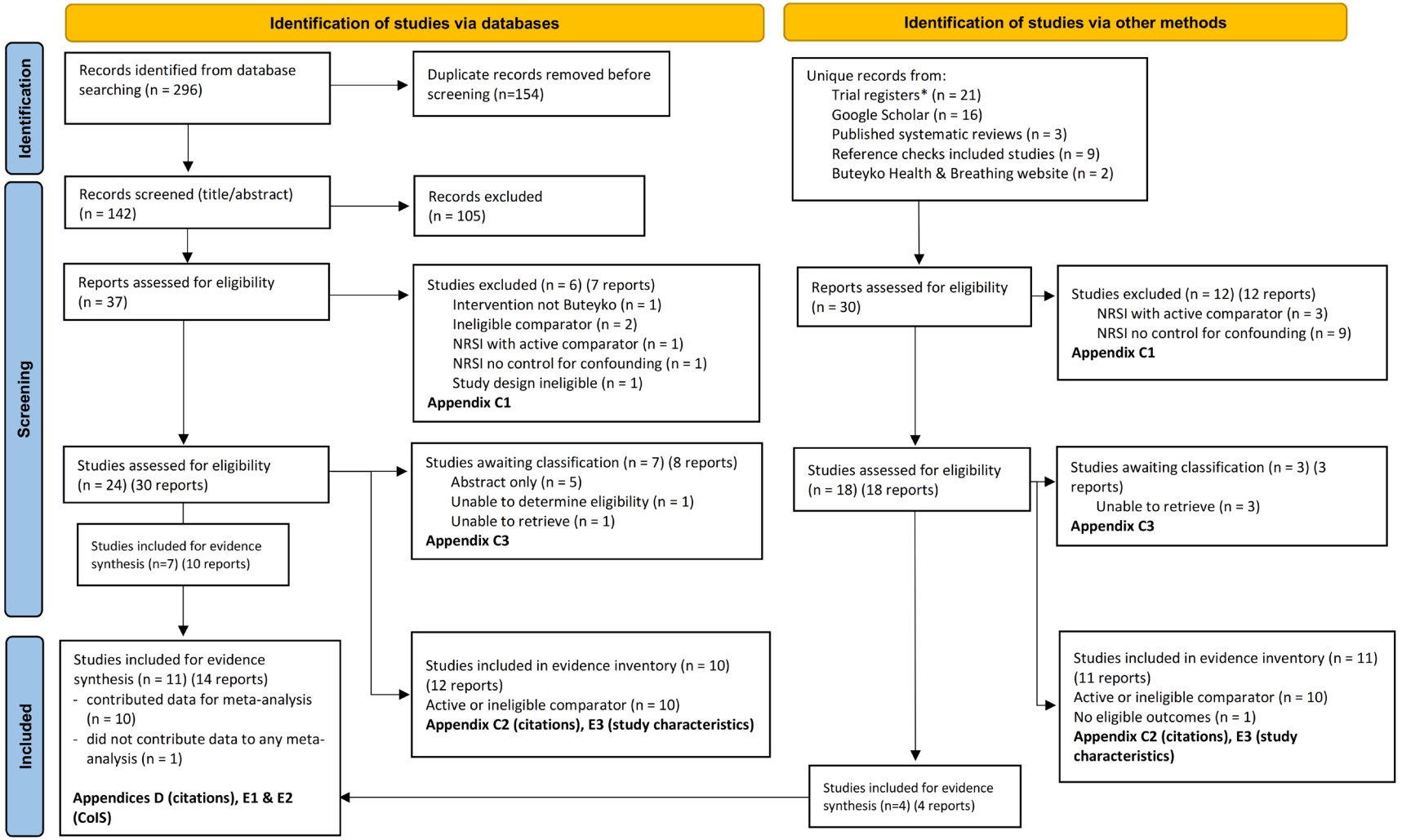

Fig. A7.1 | PRISMA diagram showing the flow of studies through the review. Studies are the unit of interest in the review. Each study could have multiple reports. CoIS: characteristics of included studies. *see results section 'Ongoing and unpublished studies'. No submissions were received via the Department's public call for evidence.

Appendix B. Data collection, analysis and interpretation of findings

B1 Data extraction and management

Study data were collected and managed using REDCap electronic data capture tools hosted at Monash University [11, 12]. The form for extracting results data was developed by the review biostatistician (JM). The form was developed for use by our team for the natural therapies reviews and had been applied to over 200 trials in the first review we conducted. Extensive pre-testing of the data extraction and coding form had been done on studies in earlier reviews with revisions made to the data extraction form to maximise the accuracy, quality and consistency of data collection.

A two-step data extraction process was implemented wherein a senior author (MM) coded the study PICO to allocate studies to the evidence inventory according to the analytic framework or include for the synthesis. Any queries from this stage were sent to the second senior author (SB) to review, with any disagreement resolved through consensus discussion. A second author (SB) then coded each study eligible for the synthesis by population group(s), outcome domains and comparisons, and allocated the study to analyses according to the analytic framework for the review. We listed all outcomes measured and selected the outcomes for inclusion in the synthesis according to our pre-specified decision rules. During this step study eligibility was confirmed and basic checks of methodology were done (e.g. confirming that a trial met the minimum requirements for randomisation). Questions about coding, allocation to analyses and outcome selection were recorded for checking by a second author (MM) or the biostatistician (ST, JM).

For each included study, one review author (SB) then extracted study characteristics and quantitative data using the data extraction and coding form. A second author (MM) independently verified the allocation to analyses, outcome selection and data extraction. All queries related to the quantitative data were referred to a biostatistician (ST). Discrepancies were resolved through discussion.

Where available, we extracted information relating to the characteristics of included studies and results as follows.

1. Study identifiers and characteristics of the study design
    - Study references (multiple publications arising from the same study were matched to an index reference; code as index paper, protocol, registry entry, results paper 1, 2, …)
    - Study name, location (country), enrolment dates (not reported by most studies), and trial registration number
    - Study design (categorised as ‘individually randomised’, ‘cluster randomised’, ‘crossover’, or ‘NRSI’); whether clustering was likely to arise because of the way the Buteyko Method was delivered (e.g. with one or two practitioners at a clinic; this information was used to determine which risk of bias tool to use for assessment).
    - Funding sources and funder involvement in study, financial and non-financial interests declared by investigators, potential conflicts (reviewer judgment), ethics approval.

2. Characteristics of each intervention group (including comparator groups)
    - Characteristics of the intervention covering domains of the Template for Intervention Description and Replication (TIDieR) checklist [17]
    - The Buteyko Method intervention goal (coded, for example: relieve symptoms of a condition, prevent a condition among people with risk factors)
    - Coding of comparators (e.g. inactive – sham, inactive – no intervention, active - massage)
    - Number of participants: randomised to each group, at follow up for selected outcome, and included in analysis and reasons for loss to follow-up

3. Characteristics of participants
    - Participant eligibility criteria (verbatim; precis of key criteria to characterise population)
    - Participant characteristics: age (e.g. mean, median, range), sex
    - Population group: coded using categories specified in the final analytic framework for the review (e.g. chronic musculoskeletal pain, headache or migraine, other chronic conditions)
    - Condition: specific underlying condition as described in study (e.g. cervical spine pain; chronic primary pain), including information about severity (if relevant) and closest ICD-11 code.

- Treatment/procedure: applied to studies in which the Buteyko Method was administered for the relief of symptoms or side effects of a treatment or procedure for an underlying condition (e.g. CABG). In theory, this could include pharmacological treatment (e.g. chemotherapy), surgical, diagnostic or other procedures (as described in study).
- Other characteristics of importance within the context of each study

4. Outcomes assessed and results
   - Outcomes measured (list of all outcomes categorised as 'eligible' or 'ineligible' and categorised according to the final analytic framework; measures used for each)
   - For outcomes selected for inclusion in the summary and synthesis of results:
     - Outcome domain: categorised according to the outcome domains specified in the final analytic framework for the review (e.g. health-related quality of life, global symptoms, emotional functioning and mental health, physical function – activity limitations)
     - Outcome as described in the included study (verbatim or precis)
     - Measurement method (e.g. Asthma Quality of Life Questionnaire; 6-minute walk distance test), information required to interpret the measure (scale range and direction, minimally important difference) and timing of outcome measurement (exact timing; described in relation to timing of the Buteyko Method sessions (e.g. immediately after end of intervention period)
     - Results including: summary statistics by group (means and standard deviations, or number of events for outcomes that have been dichotomised, and sample size), estimates of intervention effect (e.g. mean differences (or adjusted mean differences), confidence intervals, t-values, p-values, or risk ratios/odds ratios for binary outcomes).
     - Data required to support risk of bias judgements (see Assessment of risk of bias of included studies) [18]

B1.1 Assessment of risk of bias of included studies

B1.1.1 Assessment of risk of bias in RCTs

We assessed the risk of bias in included studies using the revised Cochrane 'Risk of Bias' tool (RoB 2) for randomised trials [4, 18] for each outcome included in the synthesis.

RoB 2 addresses five domains:

- bias arising from the randomisation process;
- bias due to deviations from intended interventions;
- bias due to missing outcome data;
- bias in measurement of the outcome;
- bias in selection of the reported result.

We applied review-specific guidance developed for the suite of natural therapies reviews to ensure consistency across reviewers. 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 (JM or SB) then applied the tool to the selected results from each study following the RoB 2 guidance [4], and a second author (SB) checked a subset of assessments. Areas of uncertainty and frequently asked questions were shared with extractors to promote concordance. Advice was sought from the lead reviewer (SB) where there was uncertainty. 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 [4].

When multiple effects of the intervention using different approaches were presented in the trial report, we selected one effect for inclusion in the meta-analysis and for risk of bias assessment. The selected effect was chosen according to the following hierarchy, which orders the approaches from (likely) least to most biased for estimating the *effect of assignment to the intervention*: 1. the effect that corresponds to a full intention-to-treat analysis, where missing data have been multiply imputed, or a model-based approach has been used (e.g. likelihood-based analysis, inverse-probability weighting); 2. the effect corresponding to an analysis that adheres to intention-to-treat principles except

that the missing outcome data are excluded; 3. the effect that corresponds to a full intention-to-treat analysis, where missing data have been imputed using methods that treat the imputed data as if they were observed (e.g. last observation carried forward, mean imputation, regression imputation, stochastic imputation); or 4. the effect that corresponds to an 'as-treated' or 'per-protocol' analysis, or an analysis from which eligible trial participants were excluded [4, 18]. The effect used in the assessment was recorded in the data extraction form.

B1.1.2 Assessment of risk of bias in NRSIs

We had planned to use ROBINS-I [19, 20] to assess risk of bias in NRSIs, however there were no NRSIs in the included studies.

B1.2 Measures of treatment effect

We anticipated that many of the outcomes would be continuous (e.g. physical function, HR-QoL), 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 the Buteyko Method using the standardised mean difference (SMD) (implementing the Hedges' adjusted *g* version). In trials where a continuous measure had been dichotomised (e.g. a continuous pain scale is dichotomised into improvement or no improvement) and analysed as binary outcomes, we re-expressed reported, or calculated, odds ratios as SMDs [21]. We did not report any of our meta-analysis results as dichotomous outcomes.

B1.2.1 Interpretation of treatment effects

Given the wide range of conditions, outcomes and measurement methods reported in the studies eligible for this review, it was not possible to specify thresholds for interpreting the size of the effect for each outcome measure. We planned to use Cohen's guiding rules for interpreting SMDs where 0.2 represents a small effect, 0.5 a moderate effect, and 0.8 a large effect [22]. In practice, our interpretation was based on whether there was an important effect or not [23, 24], 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 the Buteyko Method 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 health-related quality of life) and harm for other outcomes (an increase in pain). Because we were concerned that bias may be leading to exaggerated effect sizes, we chose not to describe the size of effect (i.e. we did not interpret effects as small, moderate or large).

B1.3 Unit of analysis issues

There were no unit of analysis issues in studies included in this review (studies with more than two eligible groups (arms) for a comparison were included after combining data from both arms, and no cluster or cross over trials).

B1.4 Dealing with missing data

As planned in the protocol, we did not contact trial authors to obtain missing information (e.g. study characteristics, description of conduct of the trial) or aggregate level statistics (e.g. missing standard deviations). However, we attempted to calculate statistics necessary for meta-analysis using algebraic manipulation of reported statistics (e.g. computing the standard error for the treatment effect from a reported p-value). Studies for which we calculated or imputed statistics are annotated in forest plots. We planned to explore the impact of these decisions in sensitivity analyses but there were too few studies to do so. Studies for which we could not calculate or impute the statistics required for inclusion in the meta-analysis are listed in the forest plot and in Appendix E4 with reasons for why the results could not be included.

We planned to deal with missing outcome data within the primary trials through sensitivity analyses, where trials judged to be at a high risk of bias or some concerns would be excluded; however, this was not possible because there

were too few trials included in the review [2]. Risk of bias 'due to missing outcome data' was considered within the overall bias judgement for each trial.

B1.5 Assessment of heterogeneity

We assessed statistical heterogeneity of the intervention effects visually by inspecting the overlap of confidence intervals on the forest plots. While we report formal tests for heterogeneity using the χ^2 test (using a significance level of α=0.1), and quantified heterogeneity using the I^2 statistic [25], these statistics are unlikely to be informative with so few studies. When there was evidence of heterogeneity, we judged its importance by considering where the point estimates for studies lay in relation to the threshold for an important difference (all on one side, indicating similar interpretations across the studies, or not).

B1.6 Assessment of biases due to missing results

We used a framework for assessing risk of bias due to missing results in which an assessment is made for each meta-analysis regarding the risk and potential impact of missing results from studies in which we knew an outcome was measured but not reported (termed 'known-unknowns') and the risk of other missing studies or results (termed 'unknown-unknowns') [26]. The assessment of 'known-unknowns' involves assessment of whether trials meeting the inclusion criteria for a particular meta-analysis have missing results through examination of the publication's methods section, trial registry entry (if available), and trial protocol (if available). We also examine the potential impact of studies for which data could not be included in the meta-analysis (see A1.1.1 Types of studies; A3.1 Selection of studies. We made an assessment as to whether the missing result was potentially due to the result itself (e.g. 'not statistically significant'), and whether inclusion of the result could lead to a notable change in the meta-analysis (e.g. if the missing result is from a large trial). These assessments are reported in the results section and considered in the GRADE assessment of publication bias.

We also planned to consider whether there was evidence of selective non-reporting of results from the assessment of 'unknown unknowns'. In assessing 'unknown-unknowns', we planned to judge whether the trials not identified were likely to have results eligible for inclusion (i.e. for the outcome domain 'pain', is it likely that missing studies would have been eligible for inclusion in the overall analysis or for particular conditions). We were unable to use contour enhanced funnel plots to examine whether there was evidence of small study effects [27]. We also did not undertake sensitivity analyses to compare the combined effect estimated from the random-effects model (primary analysis) with that estimated from a fixed (common) effect model (together these analyses would inform a decision to downgrade for 'suspected' reporting (publication) bias) because there were too few studies for these analyses to be informative. In the absence of these analyses, we considered whether there was concern about selective non-reporting arising from small study effects (multiple small studies reporting large effects) and evidence of selective non-reporting in the natural therapies literature more generally.

B2 Data synthesis

B2.1 Meta-analysis

Separate comparisons were set up for each population group and outcome domains agreed in the final framework (see Figure 3.5.1). Some comparisons were stratified by more specific conditions (with an overall estimate and estimate for each condition presented where appropriate) as agreed in the prioritisation process (see Figure A6.1 Appendix A6). Subgroup analysis by population group was used to examine whether these population groups explained any observed statistical heterogeneity in the intervention effects (see Subgroup analysis).

[2] In the protocol we reported that we would conduct sensitivity analyses excluding trials judged at high or unclear risk of bias. The terminology "Unclear risk of bias" has been replaced in ROB2 with "some concerns". The approach described here is consistent with the protocol in that the sensitivity analyses were to be restricted to studies at low risk of bias.

We combined the effects using a random effects meta-analysis model, since we expected and found there to be clinical and methodological diversity across the trials that may contribute to statistical heterogeneity. These analyses used the restricted maximum likelihood estimator (REML) of between trial heterogeneity variance and the Hartung-Knapp-Sidik-Jonkman confidence interval method. Analyses were conducted in Stata Statistical Software [28]

Forest plots were used to visually depict the intervention effect estimates and their confidence intervals. Forest plots are stratified by condition and risk of bias (within population group). For completeness, results for all studies for which an effect estimate (SMD) could be calculated are presented on the forest plot, including where a single study contributed to the comparison. Studies that had missing or uninterpretable results, or for which an effect estimate (SMD) could not be calculated, are also depicted on the plot.

B2.2 Summary and synthesis when meta-analysis is not possible

Studies that were eligible for the evidence synthesis but could not be included in meta-analyses, are included in the characteristics of included studies table (Appendix E1). These studies are counted as 'missing results' rather than included in a summary or other synthesis (i.e. the result was judged to be uninterpretable or there were major concerns about the integrity of the data). Details for each of the missing studies are reported in the results for the synthesis for which the outcome would be eligible, together with the reason why data are missing. We did not assess risk of bias because bias (under- or over-estimating the effect) is only relevant if results are included in a meta-analysis or reported. The reasons why these studies were not included in the analysis do not relate to bias (i.e. incomplete reporting of effects and their variances, errors in reporting or analysis of data, no information to interpret), so a risk of bias assessment would not characterise the problems with these studies.

B2.3 Subgroup analysis and investigation of heterogeneity

We did not undertake any subgroup analyses. A provision was made to conduct these analyse to examine whether population group explains any observed statistical heterogeneity in the intervention effects, using the pre-defined groups specified in the final framework (see Figure A6.1 for population groups in each meta-analysis). However, these analyses provide limited additional information due to the small number of studies and no pooled analyses were conducted across population groups in this review.

B2.4 Sensitivity analyses

We planned to undertake and report sensitivity analyses examining if the meta-analysis estimates were robust to the meta-analysis mode, assumptions made to enable inclusion of results in the meta-analysis, and the impact of excluding studies at risk of bias. However, there were too few studies for these analyses.

B2.5 Summary of findings tables and assessment of certainty of the body of evidence

We prepared GRADE summary of findings tables for each of the main comparisons, reporting results for critical and important outcome domains (up to seven). For each result, one authors (SB) used the GRADE approach to assess our confidence in where the effect lies relative to our threshold for a small effect (the certainty of evidence) (see Measures of treatment effect). In accordance with detailed GRADE guidance [24, 29, 30], 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 [23].

1. **Risk of bias**. We assessed the overall risk of bias across all studies contributing to each synthesised result. There were too few studies to perform sensitivity analyses to examine whether removing studies at high risk of bias or some concerns changed the direction or size of effect estimate importantly (a reduction in benefit or an increase in

harm being most concerning) (see Sensitivity analyses) [3]. We therefore considered the weight that studies at risk of bias contributed to each result. Where the majority of studies were at high risk of bias, we rated down for very serious concerns.

2. **Imprecision**. We judged imprecision by examining where the 95% confidence interval for each pooled effect estimate lay in relation to our threshold for an important effect (an SMD of -0.2 or 0.2 or the threshold for rate of falls; see Measures of treatment effect). Where the confidence interval clearly crossed a threshold leading to different interpretations (e.g. interpretation of the upper bound of the interval was 'an important effect' and the lower bound 'little or no effect'), we considered rating down for imprecision. If the extent to which the confidence interval crossed the threshold was modest, and the direction was consistent with the point estimate, we did not rate down (e.g. if the upper bound of the confidence interval was an SMD of -0.15 and the point estimate -0.50). We rated down for serious imprecision if the confidence interval crossed one threshold (important benefit or important harm) and the interpretation of either the upper or lower bound of the interval was different from the point estimate (e.g. if the upper bound of the confidence interval was an SMD of 0.40 indicating an important increase in symptoms, and the point estimate was -0.15 indicating an unimportant reduction in symptoms). We rated down for very serious imprecision if the confidence interval crossed two thresholds (important benefit and important harm) and for extremely serious imprecision where the confidence interval was so wide that the result was considered uninterpretable. In line with GRADE guidance, we considered the likely impact of inconsistency when rating imprecision since inconsistency can contribute to imprecision [31, 32].
3. **Inconsistency**. We assessed whether there was important, unexplained inconsistency in results across studies considering the overlap of confidence intervals (non-overlap indicating potentially important differences in direction or size of effect). Where there were concerns about inconsistency based on non-overlapping confidence intervals, we considered where the point estimates lie in relation to the threshold for an important effect (if all to one side of a threshold, we were less concerned). While we calculated statistical measures to quantify and test for heterogeneity (I^2 statistic, χ^2 test), there were too few studies for these statistics to be informative. To enhance our interpretation of whether inconsistency is important, we planned to calculate and examine the prediction interval, considering whether it included values that lead to a different conclusion than an assessment based on the confidence interval [33]. However, this is only informative with more than 10 studies, so the method could not be used. Due to the small number of studies, we were unable to used results of subgroup analyses to explain the inconsistency (see Assessment of heterogeneity; specifically, the population subgroups). Where inconsistency was not explained, we rated down. Where a result was based on a single study, inconsistency was not rated [31].
4. **Indirectness.** We assessed 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). For example, differences between the interventions delivered and delivery of the Buteyko Method in Australia that are likely to influence the size of effect. Where results came from a single small study, we were concerned that similar effects might not be observed in the population of interest more generally, and rated down for serious indirectness. Where the included studies addressed only part of the population of interest (e.g. the only form of cardiovascular disease was hypertension), we did not rate down for indirectness. Instead, we specified the population from which data came when interpreting results and indicated uncertainty for the population group more generally.
5. **Publication bias**. Our judgement of publication bias was based on assessment of bias due to missing results, primarily from interpretation of know unknowns as per Cochrane guidance for reviews with a small number of studies, where methods for investigating unknown unknowns are less useful (see Assessment of biases due to missing results). We planned to consider the potential impact of excluding studies in languages other than English, but did not identify any studies in languages other than English.

[3] In the protocol we reported that we would conduct sensitivity analyses excluding trials judged at high or unclear risk of bias. The terminology "Unclear risk of bias" has been replaced in ROB2 with "some concerns". The approach described here is consistent with the protocol in that the sensitivity analyses were to be restricted to studies at low risk of bias.

6. **Upgrading domains** (large effect size, dose response gradient, opposing plausible residual confounding). While, in principle, these domains apply to randomised trials, there is no precedent for rating up the evidence from randomised trials, and we did not have reason to apply them in this review.

Using GRADE decision rules, we derived an overall GRADE for the certainty of evidence for each result included in the summary of findings table [30]. A result from a body of evidence comprised of randomised trials begins as 'high' certainty evidence (score=4), and can be rated down (-1 or -2) for serious or very concerns on any GRADE domain that reduces confidence that the Buteyko Method has an important effect (as determined by the pre-specified thresholds) [29, 30, 34]. As indicated in point 2, we applied the most recent GRADE guidance which has provision for rating down (-3) for extremely serious imprecision.

Summary of findings tables were prepared using the GRADEpro GDT software [35]. The tables include:

- estimates of the effects of the Buteyko Method reported as standardised mean differences
- the overall GRADE (rating of certainty) and an explanation of the reason(s) for rating down (or borderline decisions) [36].
- 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 B2.6 interpretation of findings) [37].

We present the certainty of evidence in summary of findings tables using one of four levels as explained below.

Certainty	GRADE interpretation	Implications
High (⊕⊕⊕⊕)	we are very confident that the true effect lies close to that of the estimate of the effect	further research is very unlikely to change the confidence in the estimate of effect
Moderate (⊕⊕⊕⊖)	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.	further research is likely to have an important impact in the confidence in the estimate of effect
Low (⊕⊕⊖⊖)	our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect.	further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate
Very low (⊕⊖⊖⊖)	we have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect.	any estimate of effect is very uncertain

B2.6 Interpretation of findings (evidence statements)

When interpreting results, we followed GRADE guidance for writing informative statements [37]. 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, 'the Buteyko Method may improve symptoms' 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 the Buteyko Method on outcome'. This is one of two options that GRADE provides for interpreting findings based on very low certainty of evidence: "one option gives the direction of the effect, the other does not" [37]. The decision not to interpret very low certainty results was made independently by the NTWC to ensure a consistent and clear interpretation of findings across Natural Therapy Review reports (see Appendix G).

References for Appendix A and B

1. National Health and Medical Research Council: **Statement of requirement: Evidence evaluations for review of natural therapies (tranche two).** 2020.
2. Dodd S, Clarke M, Becker L, Mavergames C, Fish R, Williamson PR: **A taxonomy has been developed for outcomes in medical research to help improve knowledge discovery.** *Journal of Clinical Epidemiology* 2018, **96:**84-92.
3. World Health Organisation: **International statistical classification of diseases and related health problems (11th ed.) Available from: https://www.who.int/classifications/classification-of-diseases Access date: 15 January 2021** 2019.
4. 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.
5. Reeves B, Deeks J, Higgins J, Shea B, Tugwell P, Wells G: **Chapter 24: Including non-randomized studies on intervention effects.** In *Cochrane Handbook for Systematic Reviews of Interventions version 63 (updated February 2022) Cochrane, 2022 Available from wwwtrainingcochraneorg/handbook.* Edited by Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, VA W; 2022
6. Reeves B, Wells G, Waddington H: **Quasi-experimental study designs series-paper 5: a checklist for classifying studies evaluating the effects on health interventions-a taxonomy without labels.** *J Clin Epidemiol* 2017, **89:**30-42.
7. Higgins J, Eldridge S, Li T: **Chapter 23: Including variants on randomized trials.** In *Cochrane Handbook for Systematic Reviews of Interventions version 63 (updated February 2022) Cochrane, 2022 Available from wwwtrainingcochraneorg/handbook.* Edited by Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page M, Welch V; 2022
8. McKenzie JE, Brennan SE, Ryan RE, Thomson HJ, Johnson RV, Thomas J: **Chapter 3: Defining the criteria for including studies and how they will be grouped for the synthesis.** In *Cochrane Handbook for Systematic Reviews of Interventions version 63 (updated February 2022) Cochrane, 2022 Available from wwwtrainingcochraneorg/handbook.* 2nd edition. Edited by Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Welch V. Chichester (UK): John Wiley & Sons; 2022
9. National Health and Medical Research Council: **Buteyko 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.
10. National Health and Medical Research Council: **Draft framework for protocols for systematic reviews of randomised controlled trials and non-randomised studies of interventions.** 2020.
11. 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.
12. 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.
13. Jain M, Vardhan V, Yadav V, Pallavi H: **Efficacy of Buteyko Breathing Technique on Anxiety, Depression and Self Efficacy in Coronary Artery Bypass Graft Patients: A Randomised Clinical Trial.** *Journal of Clinical and Diagnostic Research* 2023, **17:**YC11-YC14.
14. Vagedes J, Helmert E, Kuderer S, Vagedes K, Wildhaber J, Andrasik F: **The Buteyko breathing technique in children with asthma: a randomized controlled pilot study.** *Complement Ther Med* 2021, **56:**102582.
15. Burgess J, Ekanayake B, Lowe A, Dunt D, Thien F, Dharmage SC: **Systematic review of the effectiveness of breathing retraining in asthma management.** *Expert review of respiratory medicine* 2011, **5:**789-807.
16. O'Connor E, Patnode CD, Burda BU, Buckley DI, Whitlock EP: **AHRQ Comparative Effectiveness Reviews.** In *Breathing Exercises and/or Retraining Techniques in the Treatment of Asthma: Comparative Effectiveness.* Rockville (MD): Agency for Healthcare Research and Quality (US); 2012
17. Hoffmann T, Glasziou P, Barbour V, Macdonald H: **Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide.** *BMJ* 2014, **1687:**1 - 13.
18. 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.
19. Sterne J, Higgins J, Elbers R, Reeves B, and the development group for ROBINS-I: **Risk Of Bias In Non-randomized Studies of Interventions (ROBINS-I): detailed guidance, updated 12 October 2016. Available from http://www.riskofbias.info [accessed 7 September 2022].** 2016.
20. Sterne JA, Hernan MA, Reeves BC, Savovic J, Berkman ND, Viswanathan M, Henry D, Altman DG, Ansari MT, Boutron I, et al: **ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions.** *BMJ* 2016, **355:**i4919.

21. Chinn S: **A simple method for converting an odds ratio to effect size for use in meta-analysis.** *Stat Med* 2000, **19:**3127-3131.
22. 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
23. Brennan SE, Johnston RV: **Research Note: Interpreting findings of a systematic review using GRADE methods.** *Journal of Physiotherapy* 2023.
24. 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.
25. Higgins JP, Thompson SG: **Quantifying heterogeneity in a meta-analysis.** *Stat Med* 2002, **21:**1539-1558.
26. Page MJ, Higgins JPT, Sterne JA: **Chapter 13: Assessing risk of bias due to missing results in a synthesis.** In *Cochrane Handbook for Systematic Reviews of Interventions version 63 (updated February 2022) Cochrane, 2022 Available from wwwtrainingcochraneorg/handbook.* Edited by Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page M, Welch V; 2022
27. Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L: **Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry.** *J Clin Epidemiol* 2008, **61:**991-996.
28. StataCorp: **Stata Statistical Software: Release 17. College Station, TX: StataCorp LLC.** 2021.
29. 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.
30. 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
31. Guyatt G, Zhao Y, Mayer M, Briel M, Mustafa R, Izcovich A, Hultcrantz M, Iorio A, Alba AC, Foroutan F, et al: **GRADE Guidance 36: Updates to GRADE's approach to addressing inconsistency.** *J Clin Epidemiol* 2023.
32. Zeng L, Brignardello-Petersen R, Hultcrantz M, Mustafa RA, Murad MH, Iorio A, Traversy G, Akl EA, Mayer M, Schunemann HJ, Guyatt GH: **GRADE Guidance 34: update on rating imprecision using a minimally contextualized approach.** *J Clin Epidemiol* 2022, **150:**216-224.
33. Nikolakopoulou A, Higgins JPT, Papakonstantinou T, Chaimani A, Del Giovane C, Egger M, Salanti G: **CINeMA: An approach for assessing confidence in the results of a network meta-analysis.** *PLOS Medicine* 2020, **17:**e1003082.
34. Hultcrantz M, Rind D, Akl EA, Treweek S, Mustafa RA, Iorio A, Alper BS, Meerpohl JJ, Murad MH, Ansari MT, et al: **The GRADE Working Group clarifies the construct of certainty of evidence.** *J Clin Epidemiol* 2017, **87:**4-13.
35. GDT G: **GRADEpro Guideline Development Tool [Software]. McMaster University and Evidence Prime, 2024. Available from gradepro.org.**
36. 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.
37. 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.

Appendix C. Lists of studies considered for review

Overview of Appendix C – separate file

Appendix C is comprised of four parts (below).

These Appendices report the studies excluded at full text review with reason for exclusion, the public submissions and eligibility decision for each, the studies awaiting classification, and ongoing studies.

Appendix C1. Citation details of studies from search results excluded
Appendix C2. Citation details of studies on evidence inventory
Appendix C3. Citation details of studies awaiting classification
Appendix C4. Characteristics of ongoing and unpublished studies

Appendix D. Citations for studies included in the evidence synthesis

If multiple reports, the first citation is the index (marked *).

Arora 2019 Arora, Rachna D; Subramanian, Visalakshi H To study the effect of Buteyko breathing technique in patients with obstructive airway disease. 2019. Int J Health Sci Res; 9(3) 50-64. doi:

Arora 2022 Arora, R.; Dhokpatil, S.; Verma, C.; Rajadhyaksha, G. Effect of Buteyko Breathing Technique on Haemodynamic Parameters and Functional Capacity in Subjects with Primary Hypertension. 2022. Indian Journal of Physiotherapy and Occupational Therapy; 16(3) 73-80. doi: 10.37506/ijpot.v16i3.18400

Hassan 2012 Hassan, Zahra Mohamed; Riad, Nermine Mounir; Ahmed, Fatma Hassan Effect of Buteyko breathing technique on patients with bronchial asthma. 2012. Egyptian Journal of Chest Diseases and Tuberculosis; 61(4) 235-241. doi:

Jain 2023 *Jain, M.; Vardhan, V.; Yadav, V.; Harjpal, P. Efficacy of Buteyko Breathing Technique on Anxiety, Depression and Self Efficacy in Coronary Artery Bypass Graft Patients: A Randomised Clinical Trial. 2023. Journal of Clinical and Diagnostic Research; 17(1) YC11-YC14. doi: 10.7860/JCDR/2023/60797.17388

Jain, Moli; Vardhan, Vishnu; Yadav, Vaishnavi Efficacy of Buteyko breathing technique on Anxiety, Depression and Self-efficacy in Coronary Artery Bypass Graft Surgery patients: A Randomized Clinical Trial. 2021. ; () . doi: 10.21203/rs.3.pex-1496/v1

Mohamed 2019 Mohamed, Yosreah; Elderiny, Sabah; Ibrahim, Lobna The effect of Buteyko breathing technique among patients with bronchial asthma: Comparative study. 2019. International Journal of Midwifery and Nursing Practice; 2(2) 45566. doi:

Mohamed 2022 Mohamed, Nahed Khamies; Sayed, Etemad Hussin; Abd-Elhamed, Amany Gamal Comparison between The Effect of Buteyko and Pranayama Breathing Techniques on Children with Asthma. 2022. Assiut Scientific Nursing Journal; 10(29) 139-151. doi:

Opat 2000 *Opat, A. J.; Cohen, M. M.; Bailey, M. J.; Abramson, M. J. A clinical trial of the Buteyko Breathing Technique in asthma as taught by a video. 2000. The Journal of asthma : official journal of the Association for the Care of Asthma; 37(7) 557-64.

Opat, A.; Cohen, M.; Bailey, M.; Abramason, M. The Buteyko breathing techniques in asthma: a clinical trial. 1999. Proceedings of the thoracic society of australia & new zealand, annual scientific meeting, canberra, 26 feb-3 march 1999; () 36 Abstract number: P32

Prem 2013 Prem, Venkatesan; Sahoo, Ramesh Chandra; Adhikari, Prabha Comparison of the effects of Buteyko and pranayama breathing techniques on quality of life in patients with asthma - a randomized controlled trial. 2013. Clinical rehabilitation; 27(2) 133-41. doi: 10.1177/0269215512450521

Sathe 2020.1 Sathe, S. S.; Rajandekar, T.; Hadake, S.; Shegaonkar, V. Immediate effect of buteyko breathing in hypertensive patients: An experimental prospective study. 2020. Indian Journal of Forensic Medicine and Toxicology; 14(4) 7136-7141. doi: 10.37506/ijfmt.v14i4.12771

Sathe, S. S.; Rajandekar, T.; Thodge, K.; Bhawane, A.; Thatere, U. Immediate effect of buteyko breathing and bhramari pranayama on blood pressure, heart rate and oxygen saturation in hypertensive patients: A comparative study. 2020. Indian Journal of Forensic Medicine and Toxicology; 14(4) 7106-7111. doi: 10.37506/ijfmt.v14i4.12764

Vagedes 2021 Vagedes, Jan; Helmert, Eduard; Kuderer, Silja; Vagedes, Katrin; Wildhaber, Johannes; Andrasik, Frank The Buteyko breathing technique in children with asthma: a randomized controlled pilot study. 2021. Complementary therapies in medicine; 56() 102582. doi: 10.1016/j.ctim.2020.102582

Zeng 2019 Zeng, Haicang; Chen, Xiaoxin; Xu, Yaodong; Zheng, Yiqing; Xiong, Hao Buteyko breathing technique for obstructive Eustachian tube dysfunction: Preliminary results from a randomized controlled trial. 2019. American journal of otolaryngology; 40(5) 645-649. doi: 10.1016/j.amjoto.2019.05.018

Appendix E. Characteristics of studies included in the review

Overview of Appendix E – separate file

Appendix E is comprised of two parts, each in a separate file.

Appendix E1 provides information about the characteristics of each of the studies eligible for the evidence synthesis.

- study ID, location, setting, and study design
- the population eligibility criteria, number of participants randomised, participant characteristics, and ICD codes
- the Buteyko method treatment goal, and details about the Buteyko Method intervention(s) and comparator(s)
- a list of all reported outcome(s) categorised according to whether they were eligible or ineligible for the synthesis, the measurement method for each eligible outcome, the timing of outcome measurement, and the outcome(s) selected for inclusion in the synthesis for each outcome domain

Appendix E2 provides information about funding, declaration of interest and ethics approval for each study.

Studies were included in E1 and E2 irrespective of whether they provided data that could be included in the meta-analysis.

Appendices are as follows

E1. Characteristics of studies included in the evidence synthesis
E2. Funding sources, potential conflicts of interest and ethics approval for studies included in the evidence synthesis
E3. Characteristics of studies included in the evidence inventory

Appendix F. Risk of bias assessments

All studies in this review were individually randomised, hence all assessments use the ROB 2 tools for trials with a parallel design. Assessments are presented in alphabetical order by study ID.

For each study, an assessment was done for each outcome and comparison contributing to the meta-analysis (MA; or where results could not be included in the MA but were tabulated).

For each study we report (bold indicates the header)

- the **comparison** for the assessment,
- the **outcome domain** for the assessment,
- other **assessments** of outcomes included in MAs for the study (noting if the assessment was the same for these or other comparisons),
- the study **design** (parallel trial)

Where the RoB assessment was the same for all outcomes, comparisons or both, only one assessment is reported.

The assessment includes

- The overall risk of bias judgement
- The judgement for each domain, with an explanation provided for each signalling questions for which the response could lead to a judgement of high risk of bias or some concerns
- The response to each signalling question (numbers, the questions are reported in full below)

We did not assess studies that were counted as 'missing results' (i.e. those studies where the result was judged to be uninterpretable or where there were major concerns about the integrity of the data such that it would be misleading to report the results). In such cases, concerns about bias leading to an under- or over-estimate of effect are inconsequential compared to the impact of major errors in reported data or the interpretation of that data.

Box F1. Signalling questions from the revised Cochrane risk of bias (ROB 2) tool for randomised trials (parallel design)

Parallel (individually randomised)
Domain 1. Bias arising from the randomisation process
1.1 Was the allocation sequence random?
1.2 Was the allocation sequence concealed until participants were enrolled and assigned to interventions?
1.3 Did baseline differences between intervention groups suggest a problem with the randomization process?
Domain 2. Bias due to deviations from intended interventions
2.1 Were participants aware of their assigned intervention during the trial?
2.2 Were carers and people delivering the interventions aware of participants' assigned intervention during the trial?
2.3 If Y/PY/NI to 2.1 or 2.2: Were there deviations from the intended intervention that arose because of the trial context?
2.4 If Y/PY to 2.3 Were these deviations likely to have affected the outcome?
2.5 If Y/PY to 2/4: Were these deviations from intended intervention balanced between groups?
2.6 Was an appropriate analysis used to estimate the effect of assignment to intervention?
2.7 If N/PN/NI to 2.6: Was there potential for a substantial impact (on the result) of the failure to analyse participants in the group to which they were randomized?
Domain 3. Bias due to missing outcome data
3.1 Were data for this outcome available for all, or nearly all, participants randomized?
3.2 If N/PN/NI to 3.1a or 3.1b: Is there evidence that the result was not biased by missing data?
3.3 If N/PN to 3.2 Could missingness in the outcome depend on its true value?
3.4 If Y/PY/NI to 3.3: Is it likely that missingness in the outcome depended on its true value?
Domain 4. Bias in the measurement of the outcome
4.1 Was the method of measuring the outcome inappropriate?
4.2 Could measurement or ascertainment of the outcome have differed between intervention groups?
4.3 If N/PN/NI to 4.1 and 4.2: Were outcome assessors aware of the intervention received by study participants?
4.4 If Y/PY/NI to 4.3: Could assessment of the outcome have been influenced by knowledge of intervention received?
4.5 If Y/PY/NI to 4.4: Is it likely that assessment of the outcome was influenced by knowledge of intervention received?
Domain 5. Bias from selection of the reported result
5.1 Were the data that produced this result analysed in accordance with a pre-specified analysis plan that was finalized before unblinded outcome data were available for analysis?
5.2 ... multiple eligible outcome measurements (e.g. scales, definitions, time points) within the outcome domain?
5.3 ... multiple eligible analyses of the data?

Study ID. **Arora 2019**	**Outcome domain.** physical function (activity limitations) **Assessments**. physical function, lung function		**Comparison.** Buteyko versus inactive control **Design.** parallel (individually randomised)						
Domain	**Judgment**	**Explanation** (for concerns that lead to high or some concerns about RoB)	**Response to signalling questions**						
			SQ1	SQ2	SQ3	SQ4	SQ5	SQ6	SQ7
1. Bias arising from the randomisation process	High	Block randomisation used, equal small sized blocks. Unclear if the person allocating participants to groups could have predicted the allocation sequence, or if they had motivation to change the allocation (excluding participant or delaying enrolment).	Y	PN	N				
2. Bias due to deviations from the intended intervention	Low	Intervention group received physiotherapy and Buteyko, and through the consent process, were likely to know their assigned intervention. People delivering the intervention were likely aware of the participants' assigned intervention because the allocation sequence was not concealed. Given the control group were receiving physiotherapy, it may be reasonable to assume that they would not feel the need to seek a treatment.	PY	Y	PN	NA	NA	Y	NA
3. Bias due missing outcome data	Low	I: 13/14 (7%) C: 2/14 (14%)	Y	NA	NA	NA			
4. Bias in the measurement of the outcome	Low	Assessment undertaken by an independent assessor (although it is not clear whether they are blinded to treatment allocation). The outcome measure is observer-reported and involves little or no judgement.	N	PN	NI	PN	NA		
5. Bias in the selection of the reported results	Some concerns		NI	N	N				
OVERALL risk of bias	**High**								

Study ID. **Arora 2019**	**Outcome domain.** lung function **Assessments**. physical function, lung function		**Comparison.** Buteyko versus inactive control **Design.** parallel (individually randomised)						
Domain	**Judgment**	**Explanation** (for concerns that lead to high or some concerns about RoB)	**Response to signalling questions**						
			SQ1	SQ2	SQ3	SQ4	SQ5	SQ6	SQ7
1. Bias arising from the randomisation process	High	Block randomisation used, equal small sized blocks. Unclear if the person allocating participants to groups could have predicted the allocation sequence, or if they had motivation to change the allocation (excluding participant or delaying enrolment).	Y	PN	N				
2. Bias due to deviations from the intended intervention	Low	Intervention group received physiotherapy and Buteyko, and through the consent process, were likely to know their assigned intervention.	PY	Y	PN	NA	NA	Y	NA

Study ID. Arora 2019	**Outcome domain.** lung function **Assessments**. physical function, lung function		**Comparison.** Buteyko versus inactive control **Design.** parallel (individually randomised)						
Domain	**Judgment**	**Explanation** (for concerns that lead to high or some concerns about RoB)	**Response to signalling questions**						
			SQ1	SQ2	SQ3	SQ4	SQ5	SQ6	SQ7
		People delivering the intervention were likely aware of the participants' assigned intervention because the allocation sequence was not concealed. Given the control group were receiving physiotherapy, it may be reasonable to assume that they would not feel the need to seek a treatment.							
3. Bias due missing outcome data	Low	I: 13/14 (7%) C: 2/14 (14%)	Y	NA	NA	NA			
4. Bias in the measurement of the outcome	Low	Assessment undertaken by an independent assessor (although it is not clear whether they are blinded to treatment allocation). The outcome measure is observer-reported and involves little or no judgement.	N	PN	NI	PN	NA		
5. Bias in the selection of the reported results	Some concerns		NI	N	N				
OVERALL risk of bias	**High**								

Study ID. Arora 2022	**Outcome domain.** physical function (activity limitations) **Assessments**. physical function, physiological signs and symptoms		**Comparison.** Buteyko versus inactive control **Design.** parallel (individually randomised)						
Domain	**Judgment**	**Explanation** (for concerns that lead to high or some concerns about RoB)	**Response to signalling questions**						
			SQ1	SQ2	SQ3	SQ4	SQ5	SQ6	SQ7
1. Bias arising from the randomisation process	High	No information provided on the randomisation process beyond participants being "randomly assigned" in the abstract.	NI	NI	N				
2. Bias due to deviations from the intended intervention	Some concerns	Intervention group received Buteyko and comparator no intervention (i.e. no placebo, no sham or no 'active' standard care), so it is likely that participants were aware of their assigned intervention. People delivering the intervention were likely aware of the participants' assigned intervention because the allocation sequence was likely not concealed.	PY	PY	NI	NA	NA	Y	NA
3. Bias due missing outcome data	Low	I: 33 (calculated from summary statistics Table 3)/33 (0% missing) C: 33 (calculated from summary statistics Table 3)/33	PY	NA	NA	NA			
4. Bias in the measurement of the outcome	Low	The outcome measure is observer-reported and involves little or no judgement.	PN	PN	NI	PN	NA		
5. Bias in the selection of the reported results	Some concerns		NI	PN	N				
OVERALL risk of bias	**High**								

Study ID. Arora 2022	**Outcome domain.** physical function (activity limitations) **Assessments**. physical function, physiological signs and symptoms		**Comparison.** Buteyko versus inactive control **Design.** parallel (individually randomised)					
Domain	**Judgment**	**Explanation** (for concerns that lead to high or some concerns about RoB)	**Response to signalling questions**					
			SQ1	SQ2	SQ3	SQ4	SQ5	SQ6	SQ7

Study ID. Arora 2022	**Outcome domain.** physiological signs and symptoms **Assessments**. physical function, physiological signs and symptoms		**Comparison.** Buteyko versus inactive control **Design.** parallel (individually randomised)					
Domain	**Judgment**	**Explanation** (for concerns that lead to high or some concerns about RoB)	**Response to signalling questions**					
			SQ1	SQ2	SQ3	SQ4	SQ5	SQ6	SQ7
1. Bias arising from the randomisation process	High	No information provided on the randomisation process beyond participants being "randomly assigned" in the abstract.	NI	NI	N				
2. Bias due to deviations from the intended intervention	Some concerns	Intervention group received Buteyko and comparator no intervention (i.e. no placebo, no sham or no 'active' standard care), so it is likely that participants were aware of their assigned intervention. People delivering the intervention were likely aware of the participants' assigned intervention because the allocation sequence was likely not concealed.	PY	PY	NI	NA	NA	Y	NA
3. Bias due missing outcome data	Low	I: 33 (calculated from summary statistics Table 3)/33 (0% missing) C: 33 (calculated from summary statistics Table 3)/33	PY	NA	NA	NA			
4. Bias in the measurement of the outcome	Low	The outcome measure is observer-reported and involves little or no judgement.	PN	PN	NI	PN	NA		
5. Bias in the selection of the reported results	Some concerns		NI	PN	N				
OVERALL risk of bias	**High**								

Study ID. Hassan 2012	**Outcome domain.** lung function **Assessments**. lung function, global symptoms		**Comparison.** Buteyko versus inactive control **Design.** parallel (individually randomised)					
Domain	**Judgment**	**Explanation** (for concerns that lead to high or some concerns about RoB)	**Response to signalling questions**					
			SQ1	SQ2	SQ3	SQ4	SQ5	SQ6	SQ7
1. Bias arising from the randomisation process	High	No information provided on the randomisation process beyond participants being "randomly assigned" into two groups.	NI	NI	N				
2. Bias due to deviations from the intended intervention	Some concerns	Intervention group received Buteyko and comparator no intervention (i.e. no placebo, no sham or no 'active' standard care), so it is likely that participants were aware of their assigned intervention. People delivering the intervention were likely aware of the participants' assigned	PY	PY	NI	NA	NA	Y	NA

Study ID. Hassan 2012	**Outcome domain.** lung function **Assessments**. lung function, global symptoms		**Comparison.** Buteyko versus inactive control **Design.** parallel (individually randomised)						
Domain	**Judgment**	**Explanation** (for concerns that lead to high or some concerns about RoB)	**Response to signalling questions**						
			SQ1	SQ2	SQ3	SQ4	SQ5	SQ6	SQ7
		intervention because the allocation sequence was likely not concealed.							
3. Bias due missing outcome data	Low	I: 20 (evident from the degrees of freedom in tables with the pre-post summary statistics)/20 C: 20 (evident from the degrees of freedom in tables with the pre-post summary statistics)/20	Y	NA	NA	NA			
4. Bias in the measurement of the outcome	Low	The outcome measure is observer-reported and involves little or no judgement.	N	PN	NI	PN	NA		
5. Bias in the selection of the reported results	Some concerns		NI	NI	PN				
OVERALL risk of bias	**High**								

Study ID. Hassan 2012	**Outcome domain.** global symptoms **Assessments**. lung function, global symptoms		**Comparison.** Buteyko versus inactive control **Design.** parallel (individually randomised)						
Domain	**Judgment**	**Explanation** (for concerns that lead to high or some concerns about RoB)	**Response to signalling questions**						
			SQ1	SQ2	SQ3	SQ4	SQ5	SQ6	SQ7
1. Bias arising from the randomisation process	High	No information provided on the randomisation process beyond participants being "randomly assigned" into two groups.	NI	NI	N				
2. Bias due to deviations from the intended intervention	Some concerns	Intervention group received Buteyko and comparator no intervention (i.e. no placebo, no sham or no 'active' standard care), so it is likely that participants were aware of their assigned intervention. People delivering the intervention were likely aware of the participants' assigned intervention because the allocation sequence was likely not concealed.	PY	PY	NI	NA	NA	Y	NA
3. Bias due missing outcome data	Low	I: 20 (evident from the degrees of freedom in tables with the pre-post summary statistics)/20 C: 20 (evident from the degrees of freedom in tables with the pre-post summary statistics)/20	Y	NA	NA	NA			
4. Bias in the measurement of the outcome	Low		N	PN	PY	PY	PN		
5. Bias in the selection of the reported results	Some concerns		NI	NI	PN				
OVERALL risk of bias	**High**								

Study ID. **Jain 2023**	**Outcome domain.** physical function (activity limitations) **Assessments**. physical function, EFMH		**Comparison.** Buteyko versus inactive control **Design.** parallel (individually randomised)						
Domain	**Judgment**	**Explanation** (for concerns that lead to high or some concerns about RoB)	**Response to signalling questions**						
			SQ1	SQ2	SQ3	SQ4	SQ5	SQ6	SQ7
1. Bias arising from the randomisation process	Low	Sequence generated using computer generated random numbers. Block randomisation used (indicated in abstract), but block size not reported. Sequentially Numbered, Opaque, Sealed, Envelopes (SNOSE) used. Ambiguity in the text as to whether the same person who generated the sequence also alllocated the individuals to the treatments.	Y	PY	N				
2. Bias due to deviations from the intended intervention	Some concerns	Intervention group received Buteyko and comparator no intervention (i.e. no placebo, no sham or no 'active' standard care), so it is likely that participants were aware of their assigned intervention. People delivering the intervention were aware of the participants' assigned intervention because they were the trial researchers.	Y	Y	NI	NA	NA	Y	NA
3. Bias due missing outcome data	Low	I: 20/20 (0%) C: 20/20 (0%)	Y	NA	NA	NA			
4. Bias in the measurement of the outcome	High	Participants' knowledge of the intervention they received could have influenced their response. Participants were likely to have had a prior belief about the benefits of Buteyko compared to usual care that were likely to influence the outcome.	PN	PN	Y	PY	PY		
5. Bias in the selection of the reported results	Some concerns		NI	PN	NI				
OVERALL risk of bias	**High**								

Study ID. **Jain 2023**	**Outcome domain.** EFMH - anxiety **Assessments**. physical function, EFMH		**Comparison.** Buteyko versus inactive control **Design.** parallel (individually randomised)						
Domain	**Judgment**	**Explanation** (for concerns that lead to high or some concerns about RoB)	**Response to signalling questions**						
			SQ1	SQ2	SQ3	SQ4	SQ5	SQ6	SQ7
1. Bias arising from the randomisation process	Low	Sequence generated using computer generated random numbers. Block randomisation used (indicated in abstract), but block size not reported. Sequentially Numbered, Opaque, Sealed, Envelopes (SNOSE) used. Ambiguity in the text as to whether the same person who generated the sequence also alllocated the individuals to the treatments.	Y	PY	N				
2. Bias due to deviations from the intended intervention	Some concerns	Intervention group received Buteyko and comparator no intervention (i.e. no placebo, no sham or no 'active' standard care), so it is likely that participants were aware of their assigned intervention.	Y	Y	NI	NA	NA	Y	NA

Study ID. **Jain 2023**	**Outcome domain.** EFMH - anxiety **Assessments**. physical function, EFMH		**Comparison.** Buteyko versus inactive control **Design.** parallel (individually randomised)						
Domain	**Judgment**	**Explanation** (for concerns that lead to high or some concerns about RoB)	**Response to signalling questions**						
			SQ1	SQ2	SQ3	SQ4	SQ5	SQ6	SQ7
		People delivering the intervention were aware of the participants' assigned intervention because they were the trial researchers.							
3. Bias due missing outcome data	Low	I: 20/20 (0%) C: 20/20 (0%)	Y	NA	NA	NA			
4. Bias in the measurement of the outcome	High	Participants' knowledge of the intervention they received could have influenced their response. Participants were likely to have had a prior belief about the benefits of Buteyko compared to usual care that were likely to influence the outcome.	PN	PN	Y	PY	Y		
5. Bias in the selection of the reported results	High	Another measure of anxiety (Hindi version of Depression, Anxiety and Stress Scale 42 (DASS 42) is listed in the registry entry, but not reported in the trial report.	NI	PY	NI				
OVERALL risk of bias	**High**								

Study ID. **Mohamed 2019**	**Outcome domain.** physical function (activity limitations) **Assessments**. physical function, global symptoms		**Comparison.** Buteyko versus inactive control **Design.** parallel (individually randomised)						
Domain	**Judgment**	**Explanation** (for concerns that lead to high or some concerns about RoB)	**Response to signalling questions**						
			SQ1	SQ2	SQ3	SQ4	SQ5	SQ6	SQ7
1. Bias arising from the randomisation process	High	No information provided on the randomisation process beyond participants being "randomly assigned" into two groups. However, the design is noted to be a "quasi-experimental research design" perhaps suggesting that the allocation was not random.	NI	NI	N				
2. Bias due to deviations from the intended intervention	Some concerns	Intervention group received Buteyko and comparator no intervention (i.e. no placebo, no sham or no 'active' standard care), so it is likely that participants were aware of their assigned intervention. People delivering the intervention were aware of the participants' assigned intervention because they were the trial researchers.	PY	Y	NI	NA	NA	Y	NA
3. Bias due missing outcome data	Low	I: 50/50 (0% missing) C: 50/50 (0% missing)	Y	NA	NA	NA			
4. Bias in the measurement of the outcome	High	Participants' knowledge of the intervention they received could have influenced their response. Participants were likely to have had a prior belief about the benefits of Buteyko compared to usual care that were likely to influence the outcome. Participants met with the researcher for assessment	PN	PN	Y	PY	PY		

Study ID. Mohamed 2019	**Outcome domain.** physical function (activity limitations) **Assessments**. physical function, global symptoms		**Comparison.** Buteyko versus inactive control **Design.** parallel (individually randomised)						
Domain	**Judgment**	**Explanation** (for concerns that lead to high or some concerns about RoB)	**Response to signalling questions**						
			SQ1	SQ2	SQ3	SQ4	SQ5	SQ6	SQ7
		of their asthma symptoms weekly and help with continuity of the intervention.							
5. Bias in the selection of the reported results	Some concerns		NI	NI	PN				
OVERALL risk of bias	**High**								

Study ID. Mohamed 2019	**Outcome domain.** global symptoms **Assessments**. physical function, global symptoms		**Comparison.** Buteyko versus inactive control **Design.** parallel (individually randomised)						
Domain	**Judgment**	**Explanation** (for concerns that lead to high or some concerns about RoB)	**Response to signalling questions**						
			SQ1	SQ2	SQ3	SQ4	SQ5	SQ6	SQ7
1. Bias arising from the randomisation process	High	No information provided on the randomisation process beyond participants being "randomly assigned" into two groups. However, the design is noted to be a "quasi-experimental research design" perhaps suggesting that the allocation was not random.	NI	NI	N				
2. Bias due to deviations from the intended intervention	Some concerns	Intervention group received Buteyko and comparator no intervention (i.e. no placebo, no sham or no 'active' standard care), so it is likely that participants were aware of their assigned intervention. People delivering the intervention were aware of the participants' assigned intervention because they were the trial researchers.	PY	Y	NI	NA	NA	Y	NA
3. Bias due missing outcome data	Low	I: 50/50 (0% missing) C: 50/50 (0% missing)	Y	NA	NA	NA			
4. Bias in the measurement of the outcome	High	Participants' knowledge of the intervention they received could have influenced their response. Participants were likely to have had a prior belief about the benefits of Buteyko compared to usual care that were likely to influence the outcome. Participants met with the researcher for assessment of their asthma symptoms weekly and help with continuity of the intervention.	PN	PN	Y	PY	PY		
5. Bias in the selection of the reported results	Some concerns		NI	NI	PN				
OVERALL risk of bias	**High**								

Study ID. **Mohamed 2022**	**Outcome domain.** physical function (activity limitations) **Assessments**. physical function, global symptoms		**Comparison.** Buteyko versus inactive control **Design.** parallel (individually randomised)						
Domain	**Judgment**	**Explanation** (for concerns that lead to high or some concerns about RoB)	**Response to signalling questions**						
			SQ1	SQ2	SQ3	SQ4	SQ5	SQ6	SQ7
1. Bias arising from the randomisation process	Some concerns	Papers with one of 3 Tx groups written on it, and then the papers were placed in a box from which children chose one to determine which Tx they received. Described by investigators as 'closed envelope method' but unclear if the envelopes were ordered in any way, Tx group was visible when envelopes were drawn etc. Have answered no information because the information required to confirm appropriate use of envelopes for allocation concealment is not reported. Some differences in nighttime symptoms, use of SABA. No clear problem with randomisation, especially given smaller sample size in each group.	PY	NI	PN				
2. Bias due to deviations from the intended intervention	Some concerns	Two different breathing techniques and a control. Participants were children, so would not seek the treatment themselves. Their parents might, but there is no reason to think they would have sought one breathing intervention over the other. Can't tell if control group participants (esp. parents) sought active treatement. Practioner involvement is unlikely to matter because intervention delivery was done in the first 3-5 days over a 3 month period where the Tx was self-administered. No loss to followup.	PY	PY	NI	NA	NA	Y	NA
3. Bias due missing outcome data	High	All participants are reported as having completed final follow-up. However, for one group (pranayama) the trialists report that participants who did not comply with 15% of the treatment were excluded from the analysis. Despite this, all groups have the same number of participants at baseline and follow-up. Given that participants from the pranayama group were excluded if they did not complete treatment, we assume the same could apply to Buteyko. This missingness could depend on the true value (e.g those who found the Tx ineffective did not continue)	NI	PN	PY	PY			
4. Bias in the measurement of the outcome	High	Researchers' collected data in interviews with children using validated measures at baseline and 3 month follow-up. Nothing to indicate differences in ascertainment between intervention and control group. No information to indicate that the 'reasearchers' were unaware of Tx group, and self report measure by children, who would have been aware.	N	PN	PY	Y	PY		

Study ID. **Mohamed 2022**	**Outcome domain.** physical function (activity limitations) **Assessments**. physical function, global symptoms		**Comparison.** Buteyko versus inactive control **Design.** parallel (individually randomised)						
Domain	**Judgment**	**Explanation** (for concerns that lead to high or some concerns about RoB)	**Response to signalling questions**						
			SQ1	SQ2	SQ3	SQ4	SQ5	SQ6	SQ7
		Both measures are self-reported symptom measures. Researchers interviewed children, and the control group received no intrevention, so it seems likely that the self-report could be influenced. Given the reminders to ensure compliance, this could raise expectations in the effects of the intervention, and children may wish to please the researchers, so judged likely to be influenced by knowledge of the intervention.							
5. Bias in the selection of the reported results	Some concerns	The measures seem to be what would be expected for the study. No indication there would be other measures. Length of follow up is typical. Summary statistics are reported for all responses, so unlikely that there is select non-reporting of analyses or data.	NI	PN	PN				
OVERALL risk of bias	**High**								

Study ID. **Opat 2000**	**Outcome domain.** global symptoms **Assessments**. global symptoms, EFMH, HR-QoL		**Comparison.** Buteyko versus inactive control **Design.** parallel (individually randomised)						
Domain	**Judgment**	**Explanation** (for concerns that lead to high or some concerns about RoB)	**Response to signalling questions**						
			SQ1	SQ2	SQ3	SQ4	SQ5	SQ6	SQ7
1. Bias arising from the randomisation process	Some concerns	Sequence generated using random numbers. No information provided on the recruitment process and whether the allocation sequence was concealed prior to randomisation. Although the means in the variables measuring medication use at baseline seem quite large, there is large variability in these outcomes.	Y	NI	PN				
2. Bias due to deviations from the intended intervention	Low	Buteyko delivered via video, and the control group also received a video with different content. Participants were only informed that they were taking part in a "drug-free asthma therapy", but were not told the exact details of what their treatment involved.	PN	PY	N	NA	NA	Y	NA
3. Bias due missing outcome data	Some concerns	I: 13/18 (28%) C: 15/18 (17%) Incomplete information is provided on participants with missing outcome data for this outcome. In the Buteyko group, 1 participant in the Buteyko group is due to an administration error, but for the other 4 participants, it is not clear as to whether this may be due to the true value. However, the data for 2 of these 4 participants is available for other outcomes.	N	N	PY	PN			

Study ID. Opat 2000	**Outcome domain.** global symptoms **Assessments**. global symptoms, EFMH, HR-QoL		**Comparison.** Buteyko versus inactive control **Design.** parallel (individually randomised)						
Domain	**Judgment**	**Explanation** (for concerns that lead to high or some concerns about RoB)	**Response to signalling questions**						
			SQ1	SQ2	SQ3	SQ4	SQ5	SQ6	SQ7
4. Bias in the measurement of the outcome	Low	Participants were the outcome assessors, but were only informed that they were taking part in a "drug-free asthma therapy", and were not told the exact details of what their treatment involved.	N	N	PN	NA	NA		
5. Bias in the selection of the reported results	Some concerns	Length of follow up is typical.	NI	PN	NI				
OVERALL risk of bias	**Some concerns**								

Study ID. Opat 2000	**Outcome domain.** EFMH - wellbeing **Assessments**. global symptoms, EFMH, HR-QoL		**Comparison.** Buteyko versus inactive control **Design.** parallel (individually randomised)						
Domain	**Judgment**	**Explanation** (for concerns that lead to high or some concerns about RoB)	**Response to signalling questions**						
			SQ1	SQ2	SQ3	SQ4	SQ5	SQ6	SQ7
1. Bias arising from the randomisation process	Some concerns	Sequence generated using random numbers. No information provided on the recruitment process and whether the allocation sequence was concealed prior to randomisation. Although the means in the variables measuring medication use at baseline seem quite large, there is large variability in these outcomes.	Y	NI	PN				
2. Bias due to deviations from the intended intervention	Low	Buteyko delivered via video, and the control group also received a video with different content. Participants were only informed that they were taking part in a "drug-free asthma therapy", but were not told the exact details of what their treatment involved.	PN	PY	N	NA	NA	Y	NA
3. Bias due missing outcome data	Low	I: 16/18 (11%) C: 16/18 (11%)	Y	NA	NA	NA			
4. Bias in the measurement of the outcome	Low	Participants were the outcome assessors, but were only informed that they were taking part in a "drug-free asthma therapy", and were not told the exact details of what their treatment involved.	N	N	PN	NA	NA		
5. Bias in the selection of the reported results	Some concerns	Length of follow up is typical.	NI	PN	NI				
OVERALL risk of bias	**Some concerns**								

Study ID. **Opat 2000**	**Outcome domain.** HR-QoL **Assessments**. global symptoms, EFMH, HR-QoL		**Comparison.** Buteyko versus inactive control **Design.** parallel (individually randomised)						
Domain	**Judgment**	**Explanation** (for concerns that lead to high or some concerns about RoB)	**Response to signalling questions**						
			SQ1	SQ2	SQ3	SQ4	SQ5	SQ6	SQ7
1. Bias arising from the randomisation process	Some concerns	Sequence generated using random numbers. No information provided on the recruitment process and whether the allocation sequence was concealed prior to randomisation. Although the means in the variables measuring medication use at baseline seem quite large, there is large variability in these outcomes.	Y	NI	PN				
2. Bias due to deviations from the intended intervention	Low	Buteyko delivered via video, and the control group also received a video with different content. Participants were only informed that they were taking part in a "drug-free asthma therapy", but were not told the exact details of what their treatment involved.	PN	PY	N	NA	NA	Y	NA
3. Bias due missing outcome data	Low	I: 16/18 (11%) C: 16/18 (11%)	Y	NA	NA	NA			
4. Bias in the measurement of the outcome	Low	Participants were the outcome assessors, but were only informed that they were taking part in a "drug-free asthma therapy", and were not told the exact details of what their treatment involved.	N	N	PN	NA	NA		
5. Bias in the selection of the reported results	Some concerns	Length of follow up is typical.	NI	PN	NI				
OVERALL risk of bias	**Some concerns**								

Study ID. **Prem 2013**	**Outcome domain.** lung function **Assessments**. physical function, lung function, global symptoms, EFMH, HR-QoL		**Comparison.** Buteyko versus inactive control **Design.** parallel (individually randomised)						
Domain	**Judgment**	**Explanation** (for concerns that lead to high or some concerns about RoB)	**Response to signalling questions**						
			SQ1	SQ2	SQ3	SQ4	SQ5	SQ6	SQ7
1. Bias arising from the randomisation process	Low	The authors do not specificially mention the method used to generate the allocation sequence, but note that they used block randomisation, which suggests a random component. They use sequentially numbered, sealed, opaque envelopes and note that the randomisation procedure is undertaken by an 'independent observer' not involved in conducting the intervention and collecting the outcome measures. Even although it is not clear if the block size is fixed, and therefore, the allocation could be determined, this is unlikely given the independence of the person randomising participants.	PY	PY	N				
2. Bias due to deviations from the intended intervention	Some concerns	Per-protocol analysis was undertaking, excluding participants from the Buteyko group who did comply with the exercises.	PY	Y	PN	NA	NA	N	N

Study ID. Prem 2013	**Outcome domain.** lung function **Assessments**. physical function, lung function, global symptoms, EFMH, HR-QoL		**Comparison.** Buteyko versus inactive control **Design.** parallel (individually randomised)						
Domain	**Judgment**	**Explanation** (for concerns that lead to high or some concerns about RoB)	**Response to signalling questions**						
			SQ1	SQ2	SQ3	SQ4	SQ5	SQ6	SQ7
		Only one participant was excluded from the Buteyko group due to non-compliance.							
3. Bias due missing outcome data	Low	I: 39/40 (3%) C: 40/40 (0%)	Y	NA	NA	NA			
4. Bias in the measurement of the outcome	Low	Outcome assessor was blinded to group allocation for pulmonary function test.	N	N	Y	N	NA		
5. Bias in the selection of the reported results	Some concerns	The measures seem to be what would be expected for the study. No indication there would be other measures. Length of follow up is typical.	NI	PN	PN				
OVERALL risk of bias	**Some concerns**								

Study ID. Prem 2013	**Outcome domain.** physical function (activity limitations) **Assessments**. physical function, lung function, global symptoms, EFMH, HR-QoL		**Comparison.** Buteyko versus inactive control **Design.** parallel (individually randomised)						
Domain	**Judgment**	**Explanation** (for concerns that lead to high or some concerns about RoB)	**Response to signalling questions**						
			SQ1	SQ2	SQ3	SQ4	SQ5	SQ6	SQ7
1. Bias arising from the randomisation process	Low	The authors do not specificially mention the method used to generate the allocation sequence, but note that they used block randomisation, which suggests a random component. They use sequentially numbered, sealed, opaque envelopes and note that the randomisation procedure is undertaken by an 'independent observer' not involved in conducting the intervention and collecting the outcome measures. Even although it is not clear if the block size is fixed, and therefore, the allocation could be determined, this is unlikely given the independence of the person randomising participants.	PY	PY	N				
2. Bias due to deviations from the intended intervention	Some concerns	Per-protocol analysis was undertaking, excluding participants from the Buteyko group who did comply with the exercises. Only one participant was excluded from the Buteyko group due to non-compliance.	PY	Y	PN	NA	NA	N	N
3. Bias due missing outcome data	Low	I: 39/40 (3%) C: 40/40 (0%)	Y	NA	NA	NA			
4. Bias in the measurement of the outcome	High	Patient reported outcome, and it was likely the patient was aware of the intervention they were receiving. Participants may have had a prior belief about the benefits of Buteyko compared to usual care that were likely to influence the outcome.	N	N	Y	PY	PY		

Study ID. **Prem 2013**	**Outcome domain.** physical function (activity limitations) **Assessments**. physical function, lung function, global symptoms, EFMH, HR-QoL		**Comparison.** Buteyko versus inactive control **Design.** parallel (individually randomised)					
Domain	**Judgment**	**Explanation** (for concerns that lead to high or some concerns about RoB)	**Response to signalling questions**					
			SQ1	SQ2	SQ3	SQ4	SQ5	SQ6	SQ7
5. Bias in the selection of the reported results	Some concerns	The measures seem to be what would be expected for the study. No indication there would be other measures. Length of follow up is typical.	NI	PN	PN				
OVERALL risk of bias	**High**								

Study ID. **Prem 2013**	**Outcome domain.** global symptoms **Assessments**. physical function, lung function, global symptoms, EFMH, HR-QoL		**Comparison.** Buteyko versus inactive control **Design.** parallel (individually randomised)					
Domain	**Judgment**	**Explanation** (for concerns that lead to high or some concerns about RoB)	**Response to signalling questions**					
			SQ1	SQ2	SQ3	SQ4	SQ5	SQ6	SQ7
1. Bias arising from the randomisation process	Low	The authors do not specificially mention the method used to generate the allocation sequence, but note that they used block randomisation, which suggests a random component. They use sequentially numbered, sealed, opaque envelopes and note that the randomisation procedure is undertaken by an 'independent observer' not involved in conducting the intervention and collecting the outcome measures. Even although it is not clear if the block size is fixed, and therefore, the allocation could be determined, this is unlikely given the independence of the person randomising participants.	PY	PY	N				
2. Bias due to deviations from the intended intervention	Some concerns	Per-protocol analysis was undertaken, excluding participants from the Buteyko group who did comply with the exercises. Only one participant was excluded from the Buteyko group due to non-compliance.	PY	Y	PN	NA	NA	N	N
3. Bias due missing outcome data	Low	I: 39/40 (3%) C: 40/40 (0%)	Y	NA	NA	NA			
4. Bias in the measurement of the outcome	High	Patient reported outcome, and it was likely the patient was aware of the intervention they were receiving. Participants may have had a prior belief about the benefits of Buteyko compared to usual care that were likely to influence the outcome.	N	N	Y	PY	PY		
5. Bias in the selection of the reported results	Some concerns	The measures seem to be what would be expected for the study. No indication there would be other measures. Length of follow up is typical.	NI	PN	PN				
OVERALL risk of bias	**High**								

Study ID. Prem 2013	**Outcome domain.** EFMH - wellbeing **Assessments**. physical function, lung function, global symptoms, EFMH, HR-QoL		**Comparison.** Buteyko versus inactive control **Design.** parallel (individually randomised)						
Domain	**Judgment**	**Explanation** (for concerns that lead to high or some concerns about RoB)	**Response to signalling questions**						
			SQ1	SQ2	SQ3	SQ4	SQ5	SQ6	SQ7
1. Bias arising from the randomisation process	Low	The authors do not specificially mention the method used to generate the allocation sequence, but note that they used block randomisation, which suggests a random component. They use sequentially numbered, sealed, opaque envelopes and note that the randomisation procedure is undertaken by an 'independent observer' not involved in conducting the intervention and collecting the outcome measures. Even although it is not clear if the block size is fixed, and therefore, the allocation could be determined, this is unlikely given the independence of the person randomising participants.	PY	PY	N				
2. Bias due to deviations from the intended intervention	Some concerns	Per-protocol analysis was undertaking, excluding participants from the Buteyko group who did comply with the exercises. Only one participant was excluded from the Buteyko group due to non-compliance.	PY	Y	PN	NA	NA	N	N
3. Bias due missing outcome data	Low	I: 39/40 (3%) C: 40/40 (0%)	Y	NA	NA	NA			
4. Bias in the measurement of the outcome	High	Patient reported outcome, and it was likely the patient was aware of the intervention they were receiving. Participants may have had a prior belief about the benefits of Buteyko compared to usual care that were likely to influence the outcome.	N	N	Y	PY	PY		
5. Bias in the selection of the reported results	Some concerns	The measures seem to be what would be expected for the study. No indication there would be other measures. Length of follow up is typical.	NI	PN	PN				
OVERALL risk of bias	**High**								

Study ID. Prem 2013	**Outcome domain.** HR-QoL **Assessments**. physical function, lung function, global symptoms, EFMH, HR-QoL		**Comparison.** Buteyko versus inactive control **Design.** parallel (individually randomised)						
Domain	**Judgment**	**Explanation** (for concerns that lead to high or some concerns about RoB)	**Response to signalling questions**						
			SQ1	SQ2	SQ3	SQ4	SQ5	SQ6	SQ7
1. Bias arising from the randomisation process	Low	The authors do not specificially mention the method used to generate the allocation sequence, but note that they used block randomisation, which suggests a random component. They use sequentially numbered, sealed, opaque envelopes and note that the randomisation procedure is undertaken by an 'independent observer' not involved in conducting the intervention	PY	PY	N				

Study ID. Prem 2013	**Outcome domain.** HR-QoL **Assessments**. physical function, lung function, global symptoms, EFMH, HR-QoL		**Comparison.** Buteyko versus inactive control **Design.** parallel (individually randomised)						
Domain	**Judgment**	**Explanation** (for concerns that lead to high or some concerns about RoB)	**Response to signalling questions**						
			SQ1	SQ2	SQ3	SQ4	SQ5	SQ6	SQ7
		and collecting the outcome measures. Even although it is not clear if the block size is fixed, and therefore, the allocation could be determined, this is unlikely given the independence of the person randomising participants.							
2. Bias due to deviations from the intended intervention	Some concerns	Per-protocol analysis was undertaking, excluding participants from the Buteyko group who did comply with the exercises. Only one participant was excluded from the Buteyko group due to non-compliance.	PY	Y	PN	NA	NA	N	N
3. Bias due missing outcome data	Low	I: 39/40 (3%) C: 40/40 (0%)	Y	NA	NA	NA			
4. Bias in the measurement of the outcome	High	Patient reported outcome, and it was likely the patient was aware of the intervention they were receiving. Participants may have had a prior belief about the benefits of Buteyko compared to usual care that were likely to influence the outcome.	N	N	Y	PY	PY		
5. Bias in the selection of the reported results	Some concerns	The measures seem to be what would be expected for the study. No indication there would be other measures. Length of follow up is typical.	NI	PN	PN				
OVERALL risk of bias	**High**								

Study ID. Vagedes 2021	**Outcome domain.** lung function **Assessments**. lung function, global symptoms		**Comparison.** Buteyko versus inactive control **Design.** parallel (individually randomised)						
Domain	**Judgment**	**Explanation** (for concerns that lead to high or some concerns about RoB)	**Response to signalling questions**						
			SQ1	SQ2	SQ3	SQ4	SQ5	SQ6	SQ7
1. Bias arising from the randomisation process	Some concerns	Authors provide no information on the method used to generate the allocation sequence aside from saying that the children were randomised. Sealed envelopes used to conceal allocation, with children choosing an envelope in the presence of a research assistant not involved in the preparation of the envelopes. However, no information is provided as to whether the envelopes were opaque or whether details of participants were recorded prior to opening the envelopes. There is SD deviation difference in FEV1 at baseline; however this seems consistent with a chance imbalance given there are not large differences for many of the other baseline variables (Table 1).	NI	NI	N				

Study ID. Vagedes 2021	**Outcome domain.** lung function **Assessments**. lung function, global symptoms		**Comparison.** Buteyko versus inactive control **Design.** parallel (individually randomised)					
Domain	**Judgment**	**Explanation** (for concerns that lead to high or some concerns about RoB)	**Response to signalling questions**					
			SQ1	SQ2	SQ3	SQ4	SQ5	SQ6	SQ7
2. Bias due to deviations from the intended intervention	Some concerns	Not clear what information parents and children received when consenting to participate in the study and whether those allocated to the TAU group may have sought Buteyko. Analyses undertaken on an intention-to-treat basis.	Y	Y	NI	NA	NA	Y	NA
3. Bias due missing outcome data	Low	I: 14/16 (13%) C: 15/16 (6%)	Y	NA	NA	NA			
4. Bias in the measurement of the outcome	Low	Outcome assessors were probably not blinded to group allocation ("No further blinding was performed."). However, the outcome measure is observer-reported and involves little or no judgement.	N	PN	Y	PN	NA		
5. Bias in the selection of the reported results	Some concerns	Registry entry lists 'peak flow', but with no further details of specific instruments. However, FEV1 was the preferred measure in the review, and data is available for this outcome.	NI	PN	NI				
OVERALL risk of bias	**Some concerns**								

Study ID. Vagedes 2021	**Outcome domain.** global symptoms **Assessments**. lung function, global symptoms		**Comparison.** Buteyko versus inactive control **Design.** parallel (individually randomised)					
Domain	**Judgment**	**Explanation** (for concerns that lead to high or some concerns about RoB)	**Response to signalling questions**					
			SQ1	SQ2	SQ3	SQ4	SQ5	SQ6	SQ7
1. Bias arising from the randomisation process	Some concerns	Authors provide no information on the method used to generate the allocation sequence aside from saying that the children were randomised. Sealed envelopes used to conceal allocation, with children choosing an envelope in the presence of a research assistant not involved in the preparation of the envelopes. However, no information is provided as to whether the envelopes were opaque or whether details of participants were recorded prior to opening the envelopes. There is SD deviation difference in FEV1 at baseline; however this seems consistent with a chance imbalance given there are not large differences for many of the other baseline variables (Table 1).	NI	NI	N				
2. Bias due to deviations from the intended intervention	Some concerns	Not clear what information parents and children received when consenting to participate in the study and whether those allocated to the TAU group may have sought Buteyko. Analyses undertaken on an intention-to-treat basis.	Y	Y	NI	NA	NA	Y	NA

Study ID. **Vagedes 2021**	**Outcome domain.** global symptoms **Assessments**. lung function, global symptoms		**Comparison.** Buteyko versus inactive control **Design.** parallel (individually randomised)						
Domain	**Judgment**	**Explanation** (for concerns that lead to high or some concerns about RoB)	**Response to signalling questions**						
			SQ1	SQ2	SQ3	SQ4	SQ5	SQ6	SQ7
3. Bias due missing outcome data	Low	I: 14/16 (13%) C: 15/16 (6%)	Y	NA	NA	NA			
4. Bias in the measurement of the outcome	High	Patient reported outcome ascertained through interview and it is likely that the outcome assessor was not blinded to the group allocation ("No further blinding was performed."). Beliefs in the benefits of Buteyko by the interviewer may have influenced the responses from children.	N	PN	PY	PY	PY		
5. Bias in the selection of the reported results	Some concerns	Registry entry lists 'peak flow', but with no further details of specific instruments. However, the data for the preferred measure of lung function and other outcomesis available.	NI	PN	NI				
OVERALL risk of bias	**High**								

Study ID. **Zeng 2019**	**Outcome domain.** global symptoms **Assessments**. global symptoms		**Comparison.** Buteyko versus inactive control **Design.** parallel (individually randomised)						
Domain	**Judgment**	**Explanation** (for concerns that lead to high or some concerns about RoB)	**Response to signalling questions**						
			SQ1	SQ2	SQ3	SQ4	SQ5	SQ6	SQ7
1. Bias arising from the randomisation process	Some concerns	Authors state that the allocation sequence was computer generated. No information is provided on whether or how the sequence was concealed. Baseline characteristics similar (importantly the baseline of the outcome).	Y	NI	N				
2. Bias due to deviations from the intended intervention	Some concerns		Y	Y	NI	NA	NA	Y	NA
3. Bias due missing outcome data	Low	I: 27/29 (7%) C: 24/27 (11%)	Y	NA	NA	NA			
4. Bias in the measurement of the outcome	High	Participants may have had a prior belief about the benefits of Buteyko compared to usual care that were likely to influence the outcome.	N	PN	Y	PY	PY		
5. Bias in the selection of the reported results	Low	Outcome was the preferred measure in the review, and data was available for this outcome. Length of follow up is typical. We re-analysed the data from the individual participant data provided.	Y	N	N				
OVERALL risk of bias	**High**								

Appendix G. Differences between the protocol and the review

Changes from the protocol and methods not implemented

	Section	Planned method	Change	Details (text, rationale or both)
	A1. Objectives A1.1.3	In our protocol, we planned an overall synthesis across any condition for each outcome domain.	Not done	The plan to synthesise across conditions was a contingency for reviews that included a large number of studies examining effects diverse conditions. This was not the case for this review. As such, at the prioritisation step, the NHMRC endorsed a proposal to structure and report the summary and synthesis by population group, without reporting an overall analysis across conditions.
	A1. Objectives A1.1.3	We planned to examine the effects of Buteyko compared to "evidence-based" treatments, in the exceptional circumstance that there were studies at low risk of bias that could be combined in a synthesis.	Not possible	With the exception of pranayama, no two studies in the same population had the same active comparator. Pranayama was not considered to be an evidence-based treatment.
	A3.1 Selection of studies	We had planned to pilot title and abstract screening by three reviewers.	Change in process	We piloted title and abstract screening by two reviewers.
	B1.2 Measure of treatment effects	We planned to use Cohen's guiding rules for SMDs where 0.2 represents a small effect, 0.5 a moderate effect, and 0.8 a large effect.	We used a single threshold for an important effect (0.2) and did not interpret effect size.	**Revised text (and rationale).** Because we were concerned that bias may be leading to exaggerated effect sizes, we chose not to describe the size of effect (i.e. we did not interpret effects as small, moderate or large) because this is likely to be misleading. **Implications.** This has no implications for the certainty of evidence because our a priori plan was to assess certainty in relation to whether there was an important effect or not (i.e. in relation to a threshold for an important difference of an SMD of 0.2), not our certainty in the magnitude of effect (trivial, small, moderate or large).
	B1.2 Measure of treatment effect	Where a valid and reliable minimal important difference (MID) is available for a familiar measure of relevance to the population groups in the meta-analysis, we will re-express the SMD in units of the measure and interpret the effect in relation to the MID if feasible to do so.	We did not re-express SMDs in units of a familiar measure	**Rationale**. We followed GRADE and Cochrane guidance which recommends use of SMD for interpreting continuous outcomes in the absence of well-established MIDs. In addition, using SMDs provided a consistent basis for interpretation across all results.
	B2.4 Sensitivity analysis	Analysis to examine if the meta-analysis estimates were robust to the meta-analysis mode, assumptions made to enable inclusion of results in the meta-analysis, and the impact of excluding studies at risk of bias.	Could not be done	**Revised text.** There were too few studies to undertake these analyses.
	B2.4 Sensitivity analysis	Our stated method was to undertake and report sensitivity analyses in which we excluded	Terminology corrected (not a change to protocol)	**"**Unclear risk of bias" is the terminology used in the original ROB tool. Updated ROB2 terminology replaces this wording with "some concerns".

	Section	Planned method	Change	Details (text, rationale or both)
		"trials judged to be at an overall high or unclear risk of bias."		
	B2.5 GRADE assessments – risk of bias	As per B2.4 we did not use the term 'some concerns' when describing our approach to rating down for risk of bias	Terminology corrected (not a change to protocol)	The use of 'some concerns' is consistent with the ROB2 tool. Our approach to GRADE is consistent with that for sensitivity analyses where downgrades of -1 are considered where the majority of studies are rated as 'some concerns' or studies with the majority of weight in the analysis are rated as 'high risk of bias'. Downgrades of -2 are made where most or all studies are at high risk of bias. Decisions not to rate down in these circumstances would be warranted if sensitivity analyses showed removal of studies at risk of bias did not materially alter the effect estimate.
	B2.6 Interpretation of findings	Our endorsed protocol stated that we would report "a plain language statement interpreting the evidence for each comparison and outcome, following GRADE guidance for writing informative statements". We did not specify which option would be used for very low certainty evidence (i.e. give the direction of the effect, or limit to a statement that the 'evidence is very uncertain').	Prior to submission of the draft report, NTWC advised not to include direction of effect for very low certainty evidence.	The decision not to interpret very low certainty results was made independently by the NTWC to ensure a consistent and clear interpretation of findings across Natural Therapy Review reports.
	B2.2 Summary and synthesis when meta-analysis is not possible	For a particular comparison, if we are unable to analyse most of the effect estimates (due to incomplete reporting of effects and their variances, variability in the effect measures across the studies), we will consider alternative synthesis method.	Other synthesis methods not used. We report available data if interpretable.	**Rationale.** Where possible, we report available data and present the studies on the meta-analyses. We do not include these studies in another synthesis because the data are incompletely reported and any interpretation thereof would be inconsistent with that for other results.

Appendix H. Response to comments from the Methodological review

Methodological review (or peer review) was conducted to appraise the methodological quality and assess the appropriateness of reporting for this systematic review (including appendices).

For reporting, the methodological review assessed the systematic review against the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) Checklist (2020) and where applicable, the MECIR (Methodological Expectations of Cochrane Intervention Reviews) manual to ensure the systematic review was designed and conducted in accordance with:

- NHMRC's Developing your Guideline module in NHMRC's Guidelines for Guidelines Handbook
- Cochrane Handbook for Systematic Reviews of Interventions (updated 2022)
- GRADE guidance and GRADE working group criteria for determining whether the GRADE approach was used (GRADE handbook).

Assessment included the application of criteria for considering studies for the review and synthesis, search methods, data extraction and analysis, assessment of risk of bias of studies, assessment of the certainty of evidence using GRADE, and the interpretation and summary of findings.

The systematic review (including appendices) has been updated to reflect the amendments suggested by methodological review and NHMRC's Natural Therapies Working Committee, where appropriate. In summary, updates included additional information and/ or clarification of the Plain Language Summary, Executive Summary, Results sections and Appendices, including:

- Clarifications to the definition of comparator interventions, especially usual care.
- GRADE judgements clarified and confirmed where appropriate.
- Clarifications to the PRISMA diagram. Rewording in various parts of the report for clarity and to align with other reports.
- Requests to change the implications for practice/research to standardise across natural therapies

A detailed record of responses to all comments indicating changes that were made was provided to NHMRC together with the amended Report and Appendices documents.

Appendix I. Abbreviations

Below is a list of abbreviations used in the report. Abbreviations for outcome measures are in a table following the list.

AMED: Allied and Complementary Medicine Database

BBT: Buteyko breathing technique

CAM: complementary and alternative medicine

CENTRAL: Cochrane Central Register of Controlled Trials

CINAHL: Cumulative Index of Nursing and Allied Health Literature

CI: confidence interval

CM: Complementary Medicine

COMET: Core Outcome Measures in Effectiveness Trials

DEFF: design effect

EFMH: Emotional functioning and mental health

EUROPE PMC: Europe PubMed Central

GRADE: Grading of Recommendations, Assessment, Development and Evaluation

Grp. Group

HR-QoL: health-related quality of life

ICC: intra-cluster correlation
ICD-11: International Classification of Diseases 11th Revision
ICTRP: International Clinical Trials Registry Platform
MA: Meta-analysis
MeSH: Medical Subject Headings
MID: minimal important difference
NR: not reported
NHMRC: National Health and Medical Research Council
NRSI: non-randomised study of interventions
NTREAP: Natural Therapies Review Expert Advisory Panel
NTWC: Natural Therapies Working Committee
PICO: population, intervention, comparator, outcome
PRACI: Practitioner Research and Collaboration Initiative
PRISMA: Preferred Reporting Items for Systematic review and Meta-Analyses
PRISMA-P: Preferred Reporting Items for Systematic review and Meta-Analyses Protocols
PROSPERO: International prospective register of systematic reviews
RCT: randomised controlled trial
REML: restricted maximum likelihood estimator
ROB: risk of bias
RR: risk ratios
SD: standard deviations
SMD: standardised mean difference
TIDieR: Template for Intervention Description and Replication
TGA: Therapeutic Goods Administration
UK: United Kingdom

Abbreviations for measures reported in this review

Abbreviation	Measure
6MWD	6 minute walk distance test
ACQ	Asthma Control Questionnaire
AQLQ	Asthma Quality of Life Questionnaire
ASC, GINA	Asthma symptom control, Global Initiative for Asthma
ASD	Asthma Symptom Diary
ETDQ-7	Eustachian Tube Dysfunction Questionnaire
GAD-7	General Anxiety Disorder-7
NAEPP, PSAASS	NAEPP Patient self-assessment of asthma symptom severity
PEFR	Peak Expiratory Flow Rate
RPE	Borg Rate of Perceived Exertion Scale