Systematic review of evidence on the clinical effectiveness of Buteyko

Protocol prepared by

Cochrane Australia

15 September 2023

Trusted evidence.

Informed decisions.

Better health.

Contents

In April 2022 Cochrane Australia was contracted by the National Health and Medical Research Council (NHMRC) to design and undertake the systematic review described in this protocol. This systematic review is one of several independent contracted evidence evaluations being undertaken to update the evidence underpinning the 2015 *Review of the Australian Government Rebate on Natural Therapies for Private Health Insurance* (2015 Review) by the Department of Health and Aged Care (Department). The design and conduct of the review will be done in collaboration with the Office of NHMRC (ONHMRC), the NHMRC's Natural Therapies Working Committee (the Committee) and Department's Natural Therapies Review Expert Advisory Panel (NTREAP).

Authors and contributors to the protocol

Sue Brennan[1]* sue.brennan@monash.edu	Senior Evidence Officer responsible for leading the review. Led design of the review (including the analytic framework) and writing of the protocol, with contributions from other authors as described.
Anneliese Synnot[1] anneliese.synnot@monash.edu	Drafted background on Buteyko (including characterising the interventions and scope of practice) and conducted background research on conditions treated and outcomes.
Steve McDonald[1] steve.mcdonald@monash.edu	Designed and developed the search strategy and wrote the search methods. Will conduct and report the search, and contribute to acquisition of data from studies.
Melissa Murano[1] melissa.murano@monash.edu	Conducted background research on Buteyko (including characterising the interventions, conditions treated and outcomes) to inform the analytic framework for the review and data extraction. Will co-lead review, develop data collection tools and contribute to acquisition of data.
Joanne McKenzie[2] joanne.mckenzie@monash.edu	Designed and wrote the analysis plan and method for reporting treatment effects. Wrote the section on Assessment of biases due to missing results. Provided statistical advice on risk of bias assessment and interpretation. Planned contributions to the analysis and interpretation.

[1] Cochrane Australia, Evidence Synthesis Qualitative & Implementation Methods Unit, School of Public Health and Preventive Medicine, Monash University, St Kilda Rd, Melbourne, Australia
[2] Biostatistics, Data Analytics/Modelling and Health Economics Unit, School of Public Health and Preventive Medicine, Monash University
* contact author

Declarations of interest

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

Funding

This review is funded by the Australian Government Department of Health and Aged Care via the NHMRC, under Contract Reference ID: 2021-22C021.

1. Background

In 2015, the Australian Government conducted a *Review of the Australian Government Rebate on Natural Therapies for Private Health Insurance (2015 Review).* Underpinned by systematic reviews of evidence for each natural therapy, one of the findings from the 2015 Review was that there was no clear scientific evidence that Buteyko was effective. The National Health and Medical Research Council (NHMRC) has been engaged by the Department of Health and Aging (Department) to update the evidence underpinning the 2015 Review. This evidence evaluation of Buteyko is one of a suite of independent contracted systematic reviews that will inform the Review of the Australian Government Rebate on Private Health Insurance for Natural Therapies 2019-20 (2019-20 Review) (1).

The Buteyko Breathing Technique (Buteyko) is one of a number of systematised breathing interventions used to improve respiratory health and related conditions (2-4). These systematised breathing interventions include specific breathing techniques (nasal breathing, pursed-lips breathing, deep breathing) and approaches that combine multiple techniques such as Buteyko, yoga (which uses pranayama breathing exercises), the Papworth technique (which uses diaphragmatic breathing exercises),and breathing gymnastics (2, 3). Breathing techniques (or exercises) and retraining are widely used to treat breathing pattern disorders, especially as part of non-pharmacological care for hyperventilation (overbreathing) and dysfunctional breathing (3-6). Buteyko is one such approach devised in the 1950s by Konstantin Buteyko and introduced in Australia in the early 1990s (7) and later in the decade to other Western countries.

Australian data are lacking on the prevalence and frequency of consultation with Buteyko practitioners or routine use of the technique. The main source of information about the rates of consultation with complementary medicine practitioners in Australia is a cross-sectional survey conducted as part of the Practitioner Research and Collaborative Initiative (PRACI) (8). The 2017 PRACI survey of Australian adults found that about a third of all respondents (36%; 726/2025 respondents) had consulted at least one complementary therapist in the last 12 months, however Buteyko was not among the therapies examined.

1.1 Description of the intervention

Buteyko has been described as a health education program involving breathing techniques, posture, health and lifestyle guidelines, with the aim of returning breathing to an optimal pattern (9, 10). It may include relatively conventional breathing techniques (e.g. nasal breathing / inspiration) alongside techniques for which the scientific basis and safety for some groups has been questioned (e.g. mouth taping, long breath holds) (3). While the choice of specific breathing techniques differs between practitioners, and is usually individualised to the patient, Buteyko typically involves a structured set of daily exercises focused on “reduced-volume” breathing (relaxed diaphragm breathing), breath-holding techniques and nasal breathing (inspiration and exhalation (2, 11). Reduced breathing exercises focus on reducing tidal volume (breath size in both the inhalation and exhalation phases). Breath-holding techniques include the control pause (used at the beginning and end of the exercises to assess breathing) and if appropriate, an individually tailored “extended” pause (to progressively increase the time the patient can hold their breath), which may be used as a symptom relief and/or breathing retraining tool. Patients may be taught to clear the nasal passages with breath-holding techniques and are encouraged to nasal breathe at all times, including during sleep and exercise (2, 11). Mouth taping has been advocated by some as a way of ensuring nasal breathing while sleeping (3).

Mode of administration and dose

Buteyko is usually taught over several sessions by practitioners trained in the technique (12). In Australia, learning Buteyko typically involves patients attending a minimum of five sessions with a trained Buteyko practitioner (10, 13). Sessions may be offered online, face-to-face, individually or in small groups (10, 14). Self-directed online courses are also available (15); however, the Buteyko Institute recommends against self-instruction (10). Patients are initially encouraged to practice Buteyko breathing techniques daily, with a typical exercise routine lasting 30 minutes. Thereafter, the techniques are used as needed for symptom control during rest, sleeping, exercising, speaking, eating and performing daily activities with the aim to retrain a normal breathing pattern (1, 2, 10, 16).

Practitioners of Buteyko and regulation

Buteyko education is delivered by trained Buteyko practitioners. Training for practitioners varies from short, intensive courses targeted to health professionals (e.g. physiotherapists and nurses), to several month courses for people who are not health professionals (2, 12, 13). The practice and teaching of Buteyko is not regulated by the Australian Health Practitioner Regulation National Law, which means there is no requirement for professional registration of Buteyko practitioners (17). In Australia, the Buteyko Institute of Breathing & Health conducts Buteyko practitioner training and provides registration and accreditation for practitioner members of the Institute (13), but it is unclear what proportion of Australian Buteyko practitioners this covers.

1.2 How Buteyko might work

Buteyko aims to address dysfunctional breathing patterns and hyperventilation (breathing that is too fast or deep), restoring a more natural pattern (11, 12). Konstantin Buteyko theorised that hyperventilation, or over-breathing, was the main underlying cause of many diseases, including asthma and sleep disorders. He suggested that this hyperventilation leads to hypocapnia (a decrease in blood carbon dioxide levels), triggering bronchospasm in the case of people with asthma. As such, the breathing exercises he devised focus on reducing over-breathing, and on under-breathing if appropriate, with the goal of raising carbon dioxide levels and achieving bronchodilation without medication (11, 18). The underlying premise that low carbon dioxide levels are a primary cause of respiratory symptoms lacks supporting evidence, and evidence to the contrary exists for some aspects of the mechanism proposed by Buteyko (3).

A number of additional mechanisms of action for Buteyko have been proposed. These include improved breathing biomechanics, altering respiratory muscle and nerve functioning, and psychophysiological effects, such as disrupting the feedback loop between anxiety and shortness of breath (2, 16).

1.3 Description of conditions for which Buteyko is used

Buteyko is primarily used to treat respiratory conditions, especially chronic conditions like asthma, dysfunctional breathing (hyperventilation syndrome), and chronic obstructive pulmonary disease (COPD) (4, 5, 16, 19). Other conditions treated include those for which breathing techniques are suggested by Buteyko therapists as having potential to relieve symptoms where an underlying pattern of abnormal breathing may be a contributing factor. These conditions include sleep disorders (especially sleep apnoea), allergies affecting the respiratory system, sinusitis, breathing abnormalities (e.g. chronic mouth breathing in children), and anxiety disorders (2, 12). Although Buteyko has been used to treat other conditions such as diabetes, attention deficit hyperactivity disorder and dental disorders, the appropriateness of such uses has been

questioned (3, 16). Data are lacking about the conditions most frequently treated using Buteyko in Australia.

The British Thoracic Society's guideline for the management of asthma includes a recommendation that breathing exercise programs (including face-to-face physiotherapist taught) may be offered as an adjuvant to pharmacological treatment for adults (20). While the recommendation is based on evidence from trials of several different systematised breathing interventions, Buteyko was among the treatments evaluated. This recommendation contrasts with the Australian Asthma Handbook which lists Buteyko as a complementary therapy with 'insufficient or conflicting evidence' to recommend it as an effective therapy for people with asthma (21).

1.4 Why it is important to do this review

This systematic review will inform the Australian Government's Natural Therapies Review 2019-20, which is evaluating evidence of the clinical effectiveness of 16 therapies (including Buteyko). The conclusion from the evidence evaluation conducted on Buteyko for the 2015 Review was that "[i]n people with asthma, Buteyko breathing technique may potentially reduce bronchodilator use compared with inactive control but has no consistent significant effect on pulmonary function, asthma symptoms or quality of life...there is insufficient evidence to support the clinical use of Buteyko breathing technique for the management of asthma.

Conclusions from the 2015 Review were unable to be drawn about the effectiveness of Buteyko breathing technique for conditions other than asthma" (22)(pp 54-55).

The 2015 evidence evaluation used overview methods, synthesising results from two systematic reviews published between 2008 and June 2013. All the primary studies (N = 7) included in these systematic reviews were published before 2008. Since the completion of the original evidence evaluation, there have been additional published trials of Buteyko although the number of trials remains small. In contrast to the 2015 Buteyko evidence evaluation, which was limited to evidence from randomised trials included in existing systematic reviews, this review will examine evidence from eligible primary studies (randomised trials and non-randomised studies) published from database inception until the date of the last search for this systematic review.

2. Objectives

The overall objective of this systematic review is to examine the evidence for the clinical effectiveness of Buteyko in preventing and/or treating injury, disease, medical conditions or preclinical conditions (1). The review will focus on outcomes (and underlying conditions) for which Buteyko is commonly sought or prescribed in Australia, and which are relevant to the 2019-20 Review of the Private Health Insurance rebate.

The specific objectives of the review follow (framed as questions). Examples of potentially relevant outcome domains and conditions are included to illustrate the breadth of questions to be addressed in the synthesis. These questions will be refined through a staged prioritisation process (Methods, Figure 1) to align with priorities for the 2019-20 Review, ensure a consistent approach across the evidence evaluations of natural therapies (where appropriate), and make best use of available evidence.

Primary objectives

1. What is the effect of *Buteyko* compared to *no Buteyko (inactive controls)* (see section 3.1.3 Comparisons) among people with any condition, pre-condition, injury or risk factor on

outcomes for which Buteyko is indicated? (for example, what is the effect on breathing patterns, health-related quality of life, global symptoms, or sleep quality, irrespective of the underlying condition?)

Secondary objectives

2. For each specified condition, pre-condition, injury or at-risk population, what is the effect of *Buteyko* compared to no Buteyko on outcomes for which Buteyko is indicated? (for example, what are the effects on health-related quality of life among people with asthma, people with sleep apnoea, and people with anxiety disorders?)

3. What are the effects of *Buteyko* compared to *evidence-based 'gold standard' treatments*? (see 3.1.3 Types of interventions - Comparisons)

4. What evidence exists examining the effects of *Buteyko* compared to other active comparators? (i.e. not a 'gold standard')

3. Methods

Methods reported in this protocol are based on the *Cochrane Handbook for Systematic Reviews of Interventions* (23). The GRADE approach will be used to summarise and assess the certainty of evidence arising from this review (see Section 3.3.9 for details). GRADE methods are widely used in systematic reviews and guideline development to ensure a systematic, transparent and common approach to interpreting results (24). The protocol is reported in accordance with the PRISMA-P statement (25, 26) with consideration given to the extensively updated guidance for reporting methods for systematic review in the PRISMA 2020 statement (27, 28). The review has been prospectively registered on the International prospective register of systematic reviews (PROSPERO ID CRD42023466774). Any changes to the protocol (including methods not used) will be documented when reporting the methods of the completed review.

The methods for this review are designed to accommodate the breadth of evidence about the effects of Buteyko relevant to the 2019-20 Review, and ensure a consistent approach with the other evidence evaluations of natural therapies (where appropriate). To achieve this, we will follow the staged approach summarised in **Figure 1** and elaborated in subsequent sections. We begin with an initial analytic framework (step 1) that will be refined through a prioritisation process. To facilitate this process, we will screen studies against the review eligibility criteria and compile an aggregate list of outcomes for each population group, derived from the included studies and organised by the initial framework (step 2). No identifying information will be included (i.e. no study-level information, results, references, number of studies etc). The NTWC and NTREAP will review the list in order to prioritise outcomes and advise on the final framework for the synthesis (step 3), which will be finalised (step 4) prior to proceeding with the review (step 5).

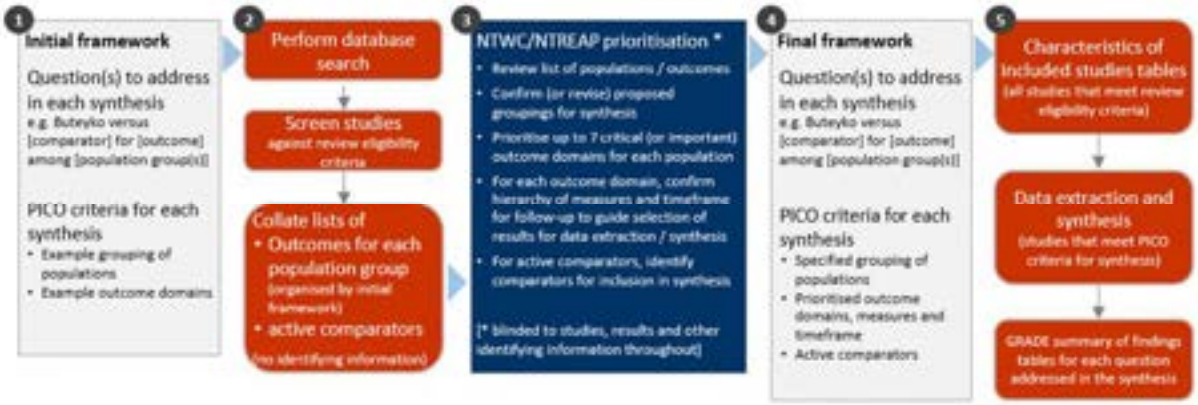

Figure 1. Staged approach for developing the analytic framework for this review

Figure 2 shows the initial analytic framework for the review. Example populations and outcome domains are included to convey the breadth of the review, and illustrate possible population and outcome groups for synthesis. These are indicative and not intended to be exhaustive. The framework was informed by research on the outcomes (and underlying conditions) for which Buteyko is commonly sought or prescribed, a scoping search of studies evaluating Buteyko, the wider literature on Buteyko, and consideration of frameworks for classifying disease and outcomes (29, 30). Details for each PICO (population, intervention, comparator, outcomes) element follow (Section 3.1).

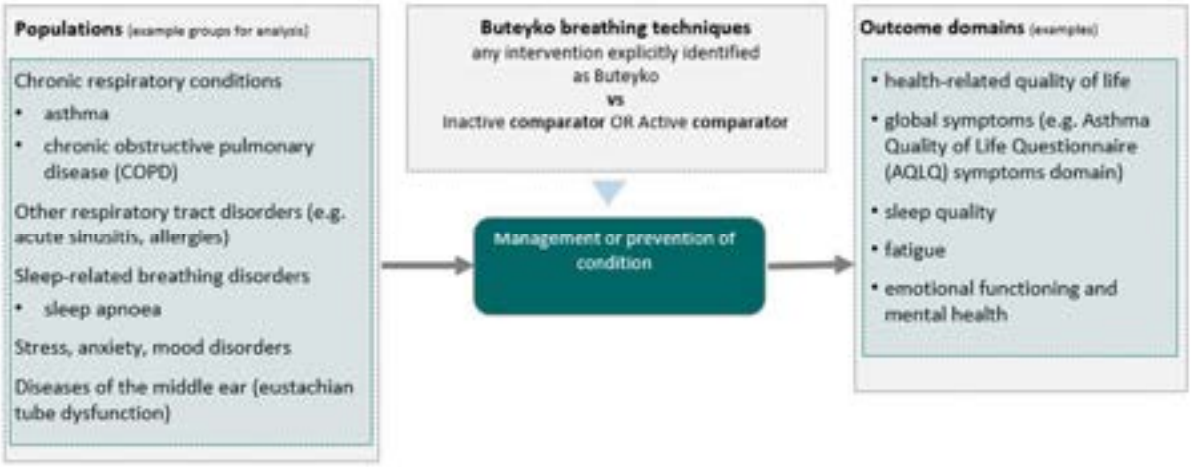

Figure 2. Initial analytic framework for the review and synthesis

3.1 Criteria for considering studies for this review

3.1.1 Types of studies

Randomised controlled trials (RCTs) and non-randomised studies of interventions (NRSIs) are eligible for inclusion in the review.

Randomised controlled trials including individually and cluster randomised, and cross-over trials are eligible for inclusion in the review.

Controlled trials in which the allocation sequence does not include a truly random element, was predictable, or was not adequately concealed from investigators, will be considered alongside RCTs when assessing risk of bias (i.e. using the Cochrane ROB 2 tool) and in the synthesis as long as there was an attempt to have some kind of 'randomisation' to groups. These studies are

sometimes referred to as 'quasi-random'. Examples include studies using methods for sequence generation based on alternation, dates (of birth or admission) and patient record numbers (31). This is to ensure a consistent approach with other natural therapies systematic reviews.

Non-randomised studies of interventions (NRSIs) with specific design features are eligible for inclusion in the review. In line with current Cochrane guidance, only NRSIs with design features that are suitable for estimating a causal effect will be eligible. While study design labels will be used as an aid to communicating about eligible designs and for use in the review (Table 1), eligibility decisions will be based on assessment of the specific design features of each study rather than the label used by the study authors (see checklist Appendix 2) (32, 33).

Eligible designs are those in which the following features are present.

- The intervention may be allocated to individuals or clusters. We anticipate that Buteyko (or the control) will be allocated to individuals in most studies, although clustering is possible in these studies given the way in which Buteyko can be delivered (i.e. the same practitioner may teach the intervention to multiple participants).
- 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) (31).
- 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.

Table 1. Study design labels to be used in the review with indicative features

Study design label	Thumbnail sketch of indicative features (extracts from (33, 34))
Nonrandomised controlled trial (NRCT)	Study in which allocation to intervention and comparator is not random or quasi-random (including where based on the practitioners' or participants' preferences) and is applied by research personnel. Participants are followed up from the start of intervention up to a later time for ascertainment of outcomes of interest (33, 34).
Prospective cohort study (PCS)	Study in which subjects are identified prospectively and classified as having received the intervention or comparator of interest on the basis of the prospectively collected information. Participants are followed up from the start of intervention up to a later time for ascertainment of outcomes of interest (33, 34).
Retrospective cohort study (RCS)	Study in which subjects are identified from historic records and classified as having received the intervention or comparator of interest on the basis of the historic information. Participants are followed up from the start of intervention up to a later time for ascertainment of outcomes of interest (33, 34).
Controlled Interrupted time series (CITS)	Studies that "collect longitudinal data measured at an aggregate level (across participants within one or more units), with several measurement times before implementation of the intervention, and several measurement times after implementation of the intervention." (34) Must

	have a contemporaneous control in which the intervention is not implemented to be eligible.
Controlled before-and after-study (CBA)	"Studies in which: (i) units are non-randomly allocated to a group that receives an intervention or to an alternative group that receives nothing or a comparator intervention; and (ii) at least one measurement of the outcome variable is made in both groups before and after implementation of the intervention…" (34)

Exclusions. Historical case control (no contemporaneous control), uncontrolled before-after studies (no contemporaneous control), and cross-sectional studies (not suitable for establishing a causal effect because both the outcome and exposure to the intervention are measured at the same time). Case-control studies are also ineligible because they are only suited to examining rare outcomes (which are unlikely to be among the core outcomes for the conditions for which Buteyko is used) and harms (which are outside the scope of this review).

Date and language restrictions.

There are no restrictions on publication date.

Potentially eligible studies published in languages other than English will not be included in the review, but will be listed according to whether they are likely to be eligible or whether eligibility cannot be determined (see 3.3.1 Selection of studies). The impact of excluding these studies will be considered in the assessment of bias due to missing results (see 3.3.8 Assessment of reporting bias, 3.3.9 Summary of findings tables and assessment of the certainty of the body of evidence).

3.1.2 Types of participants

Studies involving participants with any disease, medical condition, injury, or preclinical condition are eligible for the review. This includes healthy participants with clearly-identified risk factors (e.g. biomedical, health behaviours, or other). There are no restrictions on age or other demographic factors.

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

We will operationalise this as follows:

- The risk factor(s) for the condition that Buteyko is used to prevent must be part of the eligibility criteria for the trial or reported in the baseline data (e.g. in a trial aiming to prevent workplace-related anxiety, eligible participants must report symptoms of burnout on trial entry or have higher than normal anxiety symptom scores at baseline), and
- There must be 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; for example, a diagnosis of anxiety in a trial that aims to prevent workplace anxiety)

We expect that studies will include participants that fall within broad population groups, such as those shown in **Figure 2**. These are 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 will be made through the prioritisation process

(**Figure 1**). This process may lead to changes and additions to the population groups (i.e. broader, narrower or new groups).

Exclusions. Healthy populations seeking health improvement. This includes healthy participants using Buteyko to improve performance skills and enhance general well-being. For borderline decisions, the NTWC will be given the study PICO, aims and any information about potential risk factors reported by the trialists (without results or information that identifies the study) and asked to adjudicate.

Studies that include both healthy participants and participants eligible for the review, will be included if separate data are available or a majority of participants meet the review eligibility criteria as per guidance in the Cochrane handbook (35). For the latter, we will consider implications for the applicability of study findings in the GRADE assessment.

While studies involving any population will be included in the review (except for the specific exclusions above), if the number of eligible studies for synthesis is unmanageable, the synthesis may be limited to populations (conditions) most relevant to the use of Buteyko in Australia. Such decisions will be made through a similar process to that outlined in Figure 1, by the NTWC without knowledge of study results or other identifying information. The decisions will be guided by data about practice in the Australian context (e.g. practitioner or patient surveys that report reasons for use in Australia). Studies excluded from the synthesis to limit scope will be included in an evidence inventory (Objective 5).

3.1.3 Types of interventions

For the purpose of this review, Buteyko is 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 Buteyko from related modalities (especially other systematised breathing interventions that use similar techniques to Buteyko), and the likelihood of identifying studies in which the defining techniques and principles of Buteyko are incompletely reported, studies will be included if:

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

It is expected that the majority of studies will involve participants undertaking education in Buteyko techniques. Except for the specific exclusions below, Buteyko interventions will be eligible irrespective of:

- whether the study examines the effects of undertaking a series of educational sessions or the routine use of Buteyko,
- the specific breathing techniques used by the therapist,
- mode of delivery (individual or group; face-to-face or virtual),
- whether the intervention is guided by a teacher or self-directed (the latter possible when trained individuals use Buteyko in daily life),
- the training or qualifications of the teacher or practitioner,
- the setting in which Buteyko 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').

More details about each of these intervention features is considered in Section 3.3.2 Data extraction and Appendix 3.

Comparisons

1. Buteyko *versus* any inactive comparator (placebo/sham, no intervention, wait list control, usual care).
2. Buteyko *versus* evidence-based gold standard treatment(s) (see below for selection method)
3. Buteyko *versus* other active comparators (for inclusion in evidence inventory only, not the synthesis – See below)

Comparisons 1 and 2 will be addressed in separate syntheses (meta-analyses). Each synthesis will address a broad outcome domain with studies grouped within by population group (where appropriate; see Figure 2 for examples). Where a study includes multiple arms, with at least one eligible comparator (e.g. a placebo control arm), we will include the eligible comparison(s). Co-interventions are eligible providing the same co-intervention is in both arms of the comparison.

For comparison 2, evidence-based gold standard treatments will be identified through the prioritisation process (Figure 1, step 3). We will provide the NTWC and NTREAP with a list of active comparators identified from included studies (step 2). Studies with active comparators will not contribute to the synthesis except in the exceptional circumstance where the NTWC considers that the comparator intervention is an accepted, evidence-based 'gold standard' of care for the population in the studies, and there are studies suitable for conducting a synthesis (meta-analysis) (i.e. comparable PICO criteria, low risk of bias). These judgements will be made blinded to the studies and study results, to the fullest extent possible. For studies involving other active comparators, we will provide an inventory of available evidence, tabulating a brief description of the characteristics of the Population, Intervention, Comparator, and Outcomes (PICO) for each study.

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

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

3.1.4 Types of outcomes

Any study that measures a health outcome will be eligible for the review, however the synthesis will be limited to outcomes of importance for decision-makers as identified through the process in Figure 1 and detailed below. Outcomes eligible for this review are those that align with the reasons why Buteyko is sought by patients and prescribed by practitioners. In principle, this may include any patient-important outcome that helps elucidate the effects of Buteyko 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 are shown in Figure 2. The example outcome domains are intended to illustrate the breadth of outcomes likely to be important for understanding the effects of Buteyko across conditions, as identified from the background literature on Buteyko.

The grouping of related outcomes within each domain (Figure 2) is based on ICD11 codes and the COMET outcome taxonomy (29, 30). These systems provide a widely agreed and understood structure for categorising different outcomes

Prioritisation and selection of outcomes for summary and synthesis

Outcome prioritisation. To accommodate the breadth of relevant outcomes, the outcome domains and population-specific outcomes for inclusion in the synthesis will be determined through the prioritisation process (Figure 1).

To prioritise the most important outcomes for this review:

- We will compile a list of specific outcomes from included studies and example outcome measures (without results or identification of studies).
- Outcomes in the list will be categorised by the outcome domains and population groups in Figure 2. Outcomes that fall outside the proposed outcome domains will also be listed.
- The NTWC will be asked to indicate whether each of the listed outcome domains (or specific outcomes) is critical, important or of limited importance for understanding the effects of Buteyko on each population group. Only critical and important outcomes will be considered in the review.

Outcome selection. From each study, we will select only one outcome per outcome domain for data extraction (results), risk of bias assessment and inclusion in the summary and synthesis. The exception is for timeframe of measurement (see below).

For each outcome domain, we anticipate that there may be considerable multiplicity of results arising as follows.

(1) Across studies, both the specific outcomes and methods used to measure each outcome will vary.
(2) Within studies, results may be reported for multiple outcomes within a domain (e.g. sleep quantity, daytime functioning), multiple measures (e.g. visual analogue scale; overall and subscale scores from the Pittsburgh Sleep Quality Index), at multiple timepoints, or combinations of all three.

We will address this by using the following approach to select one outcome per domain from each study and by using the standardised mean difference (SMD) to quantify the effects of Buteyko. Using the SMD enables results to be combined for meta-analysis irrespective of the measure used, thus ensuring that any study that reports an outcome within a domain can be included in the analysis (see 3.3.9).

- An initial hierarchy of population-specific outcomes and measures will be presented to the NTWC for discussion and approval (e.g., a hierarchy of HR-QoL outcomes and measures for asthma).

- Where possible, the initial hierarchy will be based on outcome hierarchies used in published Cochrane reviews, systematic reviews of measures that provide evidence of the relevance and validity of measures, and core outcome sets.
- We will also seek advice on the most relevant time point for outcome measurement. This is likely to be at the end of the intervention period.
- We will follow the agreed hierarchy of population-specific outcomes measures and timepoints to select the most relevant and valid measure of each outcome domain available from each study for inclusion in the synthesis.

Exclusions:

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

Studies will not be excluded from the review based on outcome, except where it is possible to confirm unequivocally (e.g., from a registry record or a priori protocol) that the study only measured experience of care, safety, quality, or economic outcomes.

3.2 Search methods for identification of studies

3.2.1 Electronic searches

We will conduct searches of the Cochrane Central Register of Controlled Trials (CENTRAL), MEDLINE (via PubMed), Embase, Emcare, AMED, CINAHL and Europe PMC (for preprints). Details of completed or ongoing studies will be sought from ClinicalTrials.gov and the WHO International Clinical Trials Registry Platform. Searches will not be limited by date, language, or format of publication.

3.2.2 Selection of search terms

Buteyko is a subject heading in Embase and Emcare (Buteyko Breathing) and CINAHL (Buteyko Method) but not the other databases. In MEDLINE, Buteyko does not automatically translate to other MeSH terms since it is not listed as an entry term.

We investigated potential MeSH terms by looking at a sample of nine MEDLINE records (all RCTs) with Buteyko in the title. This revealed a lack of specific terms applied to this intervention. The two most common MeSH terms were Breathing Exercises and Asthma / Therapy. However, these terms capture a much broader scope of interventions than just Buteyko, and when combined with the randomised trial filter, retrieve over 7,000 records. Hence our MEDLINE search does not contain any MeSH terms.

We have found one protocol (19) but no published systematic reviews that focus solely on Buteyko. We looked at the search strategies of two recent systematic reviews of breathing exercises or retraining for asthma, which include studies of Buteyko (4, 36). Although these reviews include broad terms, such as breathing exercises and breathing retraining, studies categorised as Buteyko by the review authors use the word Buteyko in the title/abstract. Based on this analysis, we have opted to only include "buteyko" as a search term. The rationale being that research studies conducted to evaluate Buteyko will use the term Buteyko in the title/abstract of the record rather than non-specific or generic descriptions of breathing techniques that may or may not be classified as Buteyko.

Given the relatively small number of records we are likely to retrieve (~250) we will not include study design terms in any of the search strategies. (See Appendix 1 for proposed search strategies.)

3.3.3 Searching other resources

The 2015 overview of Buteyko identified two systematic reviews that included seven RCTs. These RCTs will be added to the records we screen, along with those from other systematic reviews retrieved by our search.

To safeguard against missing potentially eligible studies, we will also screen the included studies of systematic reviews indexed in MEDLINE with Breathing Exercises as a major MeSH term published from 2020 onwards. We have also contacted the author of the Buteyko systematic review protocol (referred to above) to request details of eligible studies they have identified.

We will search Google Scholar using the terms “Buteyko breathing” and “Buteyko method” in the title.

We will screen studies provided by the public and key stakeholders (via Department of Health and Aged Care), NTREAP and NTWC for eligibility. Where these groups recommend particular systematic reviews, we will examine the references to identify potentially eligible studies.

We will examine the reference lists of all included studies to identify additional trials or studies, and conduct forwards citation searching (i.e. looking for studies that have cited included studies) using Citation Chaser (https://estech.shinyapps.io/citationchaser/).

For studies included in the review, we will conduct supplementary searches to check whether any are listed as fraudulent or retracted, or have errata and comments associated with them. This will involve retrieving the PubMed record linked to each study and checking for the following Publication Types: Retracted Publication, Retraction of Publication, Expression of Concern, Published Erratum, and Comment. For reports of studies not indexed in PubMed, we will check the source database. We will also search the Retraction Watch Database (http://retractiondatabase.org/) using the term Buteyko. As appropriate, we will consult Chapter 4 (section 4.4.6) of the Cochrane Handbook and section 3.9 of Technical Supplement for additional guidance.

3.3 Data collection and analysis

3.3.1 Selection of studies

Records from the database and register searches will be imported into EndNote and duplicates removed. All remaining records will be imported into Covidence for screening. Records submitted through the Department, NTREAP or the Committee will be screened to confirm that the type of study is eligible, then non-duplicate records will be imported into Covidence for screening alongside other studies.

We will pilot guidance for title and abstract screening on a sample of 50 records to ensure the eligibility criteria are being applied consistently by three reviewers (SB, MM, SM). If needed, we will amend the screening guidance (but not the eligibility criteria) to enhance consistency. All records at title and abstract stage will be independently screened by at least two reviewers. All records selected for full-text screening will be reviewed independently by two reviewers. Disagreements at either stage of screening will be resolved by consensus among members of the review team.

Where disagreement cannot be resolved, advice will be sought from the NTWC (which will be provided with PICO characteristics for the de-identified study).

Studies confirmed as meeting the eligibility criteria, but for which results are not available in a published report, will be included in a list of 'ongoing studies'.

The following will be included in a list of 'studies awaiting classification'.

- Studies that are only published as abstracts or for which a full report is not available (i.e. we will not seek further information from study authors to confirm eligibility).
- Studies identified by, or submitted to, the review team after the date of the last search.
- Studies confirmed as likely to be eligible, but for which no English language translation of the full-text publication is available. Studies for which eligibility cannot be confirmed following translation of the title and abstract using Google translate, will be listed separately (Figure 3).

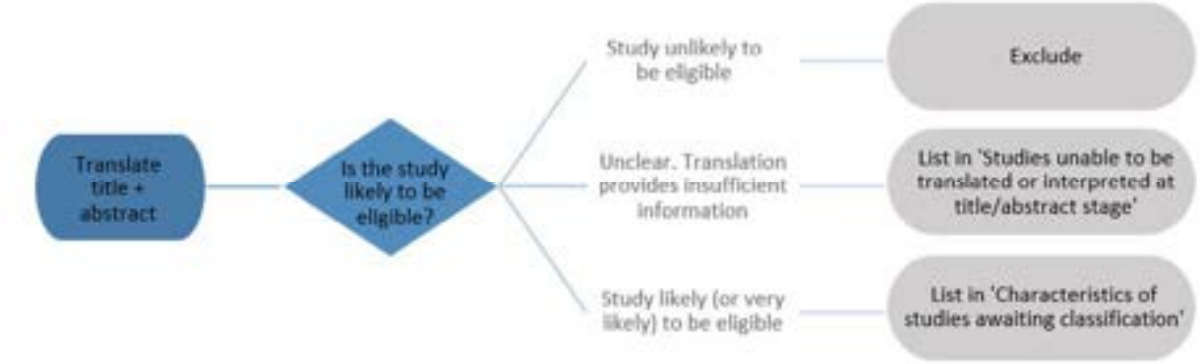

Figure 3. Flowchart showing handling of studies in languages other than English (reproduced from NHMRC framework for natural therapies systematic reviews (37)).

Studies that do not meet the eligibility criteria will be excluded and the reason for exclusion will be recorded at full-text screening. These studies will be included in a 'Characteristics of excluded studies' table in which the reason for exclusion is reported.

The search and study selection steps will be summarised in a PRISMA flow diagram.

For studies that originated from the call for evidence, NTREAP, or the Committee, we will record and report exclusion decisions irrespective of whether the study was excluded during title and abstract screening or full text review. We will document the flow of these studies through the review in the PRISMA flow chart and annotate tables with the source.

Dealing with duplicate and companion publications

Multiple publications to the same study (e.g. protocols, trial registry entries, trial reports) will be identified and linked in Excel prior to the data extraction stage. Each study will be given a unique identifier and all linked records cited in the final report. Records will be matched using trial registry numbers. Where these are not available, we will consider author names, trial name, trial location(s) and number of participants.

3.3.2 Data extraction and management

Three authors (MM, SB or SM) will pre-test the data extraction and coding form on 3-5 studies (as needed to achieve consistent coding) using REDCap electronic data capture tools hosted at

Monash University (38, 39). These studies will be purposefully selected from the included studies to cover the diversity of data types anticipated in the review. One author (SB) will review the extracted and coded pilot data for completeness, accuracy and consistency. Where needed, advice will be sought from a clinical advisor and biostatistician (JM) to ensure data are extracted as planned. Revisions to the data extraction form and guidance will be made as required to maximise the quality and consistency of data collection.

For each included study, one review author (MM, SB, or SM) will extract study characteristics and quantitative data using a pre-tested data extraction and coding form, with a 10% random sample extracted by a second author (with further sampling if needed until 80% agreement is achieved). For studies extracted by a single author, a second author (MM, SB, or SM) will independently verify the quantitative data. Discrepancies will be resolved through discussion, and advice sought from a third reviewer, the biostatistician (JM), if agreement cannot be reached or for more complex scenarios. NTWC will be asked to adjudicate on borderline decisions (e.g. whether an otherwise healthy population has 'risk factors').

We will extract 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 will be matched to an index reference; code as index paper, protocol, registry entry, results paper 1, 2, …)
    - Study name, location, commencement date, and trial registration number
    - Study design (categorised as randomised trial ['individually randomised', 'cluster randomised', 'cross-over', or 'other'] or non-randomised study of intervention ['non-randomised controlled trial', 'prospective cohort study', 'retrospective cohort study', 'controlled interrupted time series study', or 'controlled before-after study']).
    - Funding sources and funder involvement in study.
    - Financial and non-financial interests declared by investigators.

2. Study design

    - Study design label: coded according to the design identified after applying the checklist of design features (Appendix 2)
    - Detailed coding of design features of each study using the checklist of design features (Appendix 2)

3. Characteristics of each intervention group (including comparator groups)

    - Characteristics of the intervention structured by domains of the Template for Intervention Description and Replication (TIDIER) checklist (40) (see Appendix 3 for TIDIER domains, codes and an example of coding for Buteyko).
    - Number of participants: randomised to each group, at follow up for selected outcome, and included in analysis and reasons for loss to follow-up
    - For NRSIs, details of the methods used to collect information on the intervention allocated or received by each participant (i.e. the information used to classify intervention status)

4. Characteristics of participants

    - Participant eligibility criteria (verbatim or precis)
    - Age (mean, range)
    - Sex
    - Population group: coded using categories specified in the final analytic framework for the review (e.g. chronic respiratory conditions; other respiratory tract disorders (e.g. allergies); sleep-related breathing disorders; stress, anxiety and mood disorders)
    - Condition: specific underlying condition as described in study (e.g. asthma, sleep apnoea, eustachian tube disfunction), ICD11 code, information about severity (if relevant)
    - Other characteristics of importance within the context of each study

5. Outcomes assessed and results

    - Outcomes measured (a list of all outcomes, categorised according to the broad domains specified in the final analytic framework for the review, or as 'other' if none of the outcome domains apply)
    - For outcomes selected for inclusion in the summary and synthesis of results:
        - Outcome domain: categorised according to the broad domains specified in the final analytic framework for the review (e.g. health-related quality of life, global symptoms, sleep quality, fatigue, emotional functioning / mental health)
        - Outcome as described in the included study (verbatim or precis)
        - Measurement method (e.g. Asthma Quality of Life Questionnaire (AQLQ), Asthma Control Questionnaire (ACQ)), information required to interpret measures (scale range and direction, minimally important difference) and time point (exact, and time-frame categorised as 'immediate' or 'longest follow-up')
        - Results including: summary statistics by group (means and standard deviations, or number of events for cognitive outcomes that have been dichotomised, and sample size), estimates of intervention effect (e.g. mean differences (or adjusted mean differences for NRSIs), confidence intervals, t-values, p-values, or risk ratios/odds ratios for binary outcomes). For adjusted estimates, we will extract information on the analysis method, how confounding was adjusted, and which confounders were adjusted for (applies primarily to NRSIs).
        - Data required to support risk of bias judgements (see Section 3.3.3) (31, 41, 42)

3.3.3 Assessment of risk of bias of included studies

Assessment of risk of bias in RCTs

We will assess the risk of bias in included studies using the revised Cochrane 'Risk of bias' tool (RoB 2) for randomised trials (31, 42) or the ROBINS-I tool for non-randomised studies of interventions(34, 41, 43). The assessment will be made for each critical (or important) outcome included in the synthesis. Our assessment will be based on the effect of assignment to the intervention.

RoB 2 addresses five domains:

1. bias arising from the randomisation process;
2. bias due to deviations from intended interventions;
3. bias due to missing outcome data;
4. bias in measurement of the outcome;
5. bias in selection of the reported result.

For cluster trials and cross-over trials, we will use the variants of the RoB 2 tool specific for each design (31, 44). Non-randomised clinical trials that do not include the study design features that we have specified for a 'randomised' trial will be assessed with ROBINS-I.

ROBINS-I addresses seven domains:

1. bias due to confounding
2. bias in selection of participants into the study
3. bias in classification of interventions
4. bias due to deviations from the intended intervention
5. bias due to missing data
6. bias in measurement of outcomes
7. bias in the selection of the reported result

In assessing risk of bias due to confounding (domain 1), we will answer the optional question about the predicted direction of bias (to determine whether residual confounding favours the control, thus lessening concerns that any observed beneficial effects are due to bias). In our overall judgement for domains 1 and 2, we will consider whether large effects or a dose response mitigate concerns about bias due to confounding or selection of participants into the study (41, 45).

To promote concordance, the assessment will be piloted by three review authors (MM, SB, SMc) on 3-5 studies until consistent judgments are achieved across a range of scenarios. One review author (MM, SB or SMc) will then apply the tool to the selected results from each study following the RoB 2 or ROBINS-I guidance (31, 41), and a second author will verify the assessments (SB or SMc). Supporting information and justifications for judgements for each domain (i.e. for ROB 2: low, some concerns, high risk of bias; for ROBINS-I: low, moderate, serious, critical risk of bias) will be recorded. We will derive an overall summary of the risk of bias from each assessment, following the algorithm in the RoB 2 and ROBINS-I guidance (31, 41). Disagreement between review authors will be resolved through discussion, and a third review author (SB, SM or JM) will adjudicate where agreement cannot be reached.

For NRSIs, if we judge a result to be at "critical" risk of bias on the first domain (bias due to confounding), we will not assess other domains, since the overall risk of bias for the result would be "critical" by default. Results judged to be at "critical" risk of bias overall will excluded from the summary and syntheses of results, and will not contribute to our conclusions.

Selection of effects for ROB assessment and synthesis. When multiple effects of the intervention using different approaches are presented in a study report, we will select one effect for inclusion in the meta-analysis and for RoB assessments. For randomised trials, the selected effect will be 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 are excluded (31). For NRSIs, the selected effect will be the estimate adjusted for critical confounding domains.

Pre-specification of confounding factors and co-exposures

Confounding domains are “prognostic variables (factors that predict the outcome of interest)” that also predict whether an individual receives one or the other interventions of interest (41). ROBINS-I defines important confounding domains as those “for which, in the context of [a specific] study, adjustment is expected to lead to a clinically important change in the estimated effect of the [exposure]”. It is recommended that users applying ROBINS-I should consider in advance the confounding factors and co-interventions that have the potential to lead to bias in included studies. Due to the diversity of conditions and outcomes eligible for this review, confounding domains and co-exposures will be identified after the prioritisation process (Figure 1) but prior to assessment of studies.

3.3.4 Measures of treatment effect

We anticipate that many of the outcomes will be continuous (e.g. sleep quality score; activity limitation score for HR-QoL), and that varying measurement instruments will be used to measure the same underlying construct across the studies. For this reason, we will quantify the effects of Buteyko using the standardised mean difference (SMD) (implementing the Hedges’ adjusted *g* version). In trials where a continuous measure has been dichotomised (e.g. sleep quality score is dichotomised into ‘improved’ or ‘no improvement’) and analysed as binary outcomes, we will re-express reported, or calculated, odds ratios as SMDs (46). For dichotomous (e.g. proportion who experienced an exacerbation), count (e.g. number of asthma exacerbations) and time-to-event (e.g. time to first exacerbation) outcomes, we will quantify the effects of Buteyko using risk ratios (RR), rate ratios, and hazard ratios, respectively, where possible.

Thresholds for interpreting the effects. Following the GRADE minimally contextualised approach for interpreting findings, all interpretations will be based on where the combined estimate of effect (point estimate) lies in relation to a pre-specified threshold for an important effect (i.e. whether an important effect is observed or not) and the direction of effect (beneficial or harmful) (47). Given the wide range of conditions. outcomes and measurement methods expected in this review, it is not possible to specify thresholds for interpreting the size of effect for each outcome measure. We, therefore, plan to use Cohen’s guiding rules for interpreting SMDs (0.2 represents a small effect, 0.5 a moderate effect, and 0.8 a large effect) (48). In practice, our interpretation will be based on whether there was an important effect or not (47, 49, 50), with an SMD of 0.2 standard units set as the threshold for an important difference. If the SMD falls within the prespecified range of -0.2 to 0.2 (i.e. within both thresholds), the effect of Buteyko will be considered to be no different from control. An SMD above 0.2 or below -0.2 will be interpreted as an important effect.

Where a valid and reliable minimal important difference (MID) is available for a familiar and commonly used 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 (48). For dichotomous outcomes, we will use a relative risk increase of 25% (or above) and a relative risk reduction of 20% as the thresholds for an important effect. We will also express relative effects in absolute terms, and consider whether the absolute risk difference with intervention still indicates an important effect, seeking advice from the NTWC if decisions are borderline.

3.3.5 Unit of analysis issues

In this review, unit of analysis issues may arise from non-standard designs (cluster trials, cross-over trials) or from trials with more than two eligible intervention groups. In the following we outline the methods for making adjustments when necessary. Any adjustments will be documented (e.g. assumed intra-cluster correlation and average cluster size). We will also report when necessary adjustments were unable to be made due to missing information.

For cluster randomised trials that have not appropriately accounted for correlation in observations within clusters, we will attempt a re-analysis. We will do this by inflating the variance of the intervention estimates by a design effect (DEFF). The DEFF is calculated from two quantities – an intra-cluster correlation (ICC) and the average cluster size. Estimates of ICC will be imputed from other cluster trials included in the review, where possible, or by using external estimates from empirical research (e.g. Bell 2013 (51)). The average cluster size will be calculated from reported information in the trial.

For cross-over trials where an appropriate paired analysis is not available, we will attempt to approximate a paired analysis by imputing missing statistics (e.g. correlation). Estimates of the missing statistics will be imputed from other cross-over trials included in the review, where possible, or by using external estimates from empirical research (e.g. Balk 2012 (52)).

For trials where more than one comparison from the same trial is eligible for inclusion in the same meta-analysis (e.g. instructors with different training in Buteyko; different numbers of lessons), we will combine intervention groups, where it makes sense to do so; otherwise, we will appropriately reduce the sample size so that the same participants do not contribute more than once.

3.3.6 Dealing with missing data

We will not contact trial authors to obtain missing information (e.g. study characteristics, description of conduct of the study) or aggregate level statistics (e.g. missing standard deviations). However, we will attempt 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). When standard deviations cannot be calculated from available statistics, but interquartile ranges or ranges are reported, we will use the formula in Wan et al (53) to estimate approximate standard deviations. When neither of the above methods are possible, we will impute the standard deviation using the average standard deviation across trials included in the same meta-analysis that have used the same measurement tool. When means are missing, but medians are reported, we will use the formula in Wan et al (53) to estimate approximate means.

Our approach for dealing with missing outcome data within the primary trials will be through sensitivity analyses, where trials judged to be at a high or unclear risk of bias will be excluded (Section 3.3.9 Data synthesis). Risk of bias 'due to missing outcome data' is considered within the overall bias judgement for a trial. A similar approach will be taken with NRSI, excluding studies judged to be at serious risk of bias.

3.3.7 Assessment of heterogeneity

We will assess statistical heterogeneity of the intervention effects visually by inspecting the overlap of confidence intervals on the forest plots, formally test for heterogeneity using the χ^2 test (using a significance level of α=0.1), and quantify heterogeneity using the I^2 statistic (54).

3.3.8 Assessment of biases due to missing results

We will use 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 (termed 'known-unknowns') and the risk of missing studies (termed 'unknown-unknowns') (55). We will use this framework to guide our assessments of whether there is 'undetected' or 'suspected' reporting bias for each of the comparisons in our GRADE assessment (Section 3.3.10 Summary of findings tables and assessment of certainty of the body of evidence).

In assessing 'known-unknowns', we will determine what studies meeting the inclusion criteria for the particular meta-analysis have missing results through examination of the publication's methods section, trial registry entry (if available), and trial or study protocol (if available). We will make an assessment as to whether the missing result was potentially due 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). We will also assess the impact of missing results from studies reported in languages other than English that were judged as being likely to meet the eligibility criteria for each synthesis (see Section 3.1.1 Types of studies; Section 3.3.1 Selection of studies).

In assessing 'unknown-unknowns', we will judge whether the studies not identified were likely to have results eligible for inclusion (e.g. for broad outcome domains such as 'pain', it is likely that for particular conditions, missing studies would have been eligible for inclusion). We will use funnel plots and contour enhanced funnel plots to examine whether there is evidence of small study effects (56). If there is funnel plot asymmetry, we will undertake a sensitivity analysis to compare the combined effect estimated from the random-effects model (primary analysis) with that estimated from a fixed (common) effect model. If the random-effects estimate is importantly larger than the fixed-effect estimate, with no explanation for the difference (e.g. differences in clinical populations or intensity of the delivery of intervention between small and large trials, differences in risk of bias between small and large trials), then we will downgrade for 'suspected' reporting bias.

3.3.9 Data synthesis

Meta-analysis

Separate comparisons will be set up based on outcome domains agreed in the final framework (see Figure 2). These comparisons will be stratified in two ways. First, by the population groups in the final framework (see Figure 2 and 3.1.2 Types of participants for indicative groupings). This approach to structuring the meta-analysis will yield an overall estimate of the effect of Buteyko for the outcome (review objectives 1, 2 and 4), as well as estimates within each population group (review objective 2). Subgroup analysis by population group will allow examination of whether these population groups explain any observed statistical heterogeneity in the intervention effects (Section 3.3.9 Subgroup analysis). Second, by study design, where we will separately combine estimates from randomised trials and non-randomised studies. We will not calculate an overall estimate of the effect of Buteyko across these study design strata.

We will combine the effects using a random effects meta-analysis model, since we expect there to be clinical and methodological diversity across the trials that may lead to statistical heterogeneity.

These analyses will use the restricted maximum likelihood estimator (REML) of between trial heterogeneity variance and the Hartung-Knapp-Sidik-Jonkman confidence interval method.

Forest plots will be used to visually depict the intervention effect estimates and their confidence intervals. Forest plots will be stratified by condition and risk of bias (within population group).

Summary and synthesis when meta-analysis is not possible

Available effect estimates (95% confidence intervals, p-values), details of scales (direction and range), risk of bias assessments, and intervention characteristics will be tabulated. Tables will be ordered by outcome domain, population group and risk of bias assessment.

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 methods, such as calculating summary statistics of the effect estimates, combining p-values, or vote counting based on the direction of effect (57). Our choice of method will be determined by the available data (e.g. summary statistics if data permit; other methods if the data are more limited).

Subgroup analysis and investigation of heterogeneity

We will undertake a subgroup analysis to examine whether population group explains any observed statistical heterogeneity in the intervention effects (see Figure 2 and 3.1.2 Types of participants for indicative population groupings). In addition, we will consider whether the inclusion of intervention components other than breathing techniques (posture, lifestyle guidelines) explains any observed statistical heterogeneity in the intervention effects. For this analysis we will stratify the meta-analysis, grouping studies that use breathing techniques without other components separately from those that use breathing techniques and other components.

Sensitivity analyses

We plan to undertake and report sensitivity analyses examining if the meta-analysis estimates are robust to the:

- *meta-analysis model*. In addition to fitting a random-effects model, we will fit fixed effect models. This analysis will be undertaken to investigate the impact of any small-study effects.
- *inclusion of trials judged to be at an overall high or unclear risk of bias.* We will exclude trials judged to be at an overall high or unclear risk of bias.

Results of the sensitivity analyses will be tabulated, including the meta-analysis estimate (and its confidence interval), along with details of the original and sensitivity analysis assumptions.

3.3.10 Summary of findings tables and assessment of certainty of the body of evidence

We will prepare 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 author (SB) will use 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 3.3.4 Measures of treatment effect). In accordance with detailed GRADE guidance (45, 50, 58), an overall GRADE of high, moderate, low or very low certainty will be reported for each result based on whether there are serious, very serious or no concerns in relation to each of the following domains.

1. Risk of bias. We will assess the overall risk of bias across all studies contributing to each synthesised result, considering the weight studies rated at high risk of bias contribute to

the analysis. Serious, very serious or extremely serious concerns are more likely if studies at risk of bias contribute considerable weight in the analysis and sensitivity analyses indicate that removing these studies changes the size of the effect (see 3.3.9 Sensitivity analyses) (45, 58).

2. Inconsistency. We will assess whether there is 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), statistical measures that quantify and test for heterogeneity (I^2 statistic, χ^2 test), results of subgroup analyses (see 3.3.7 Assessment of heterogeneity), and 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) (59). To enhance our interpretation of whether inconsistency is important, we will also examine the prediction interval, considering whether it included values that lead to a different conclusion than an assessment based on the confidence interval (60). Where a result is based on a single study, inconsistency will not be rated.

3. Imprecision. We will assess whether the confidence interval for each pooled effect estimate crosses our threshold for an important effect (e.g. including a small effect and little or no difference, which would lead to different interpretations) and, for large effects, whether the sample size meets the optimal information size (based on number of events for binary outcomes; sample size for continuous outcomes). In judging imprecision, we will use our thresholds for a small but important effect as specified in 3.3.4 Measures of treatment effect. We will rate down for serious imprecision when the confidence interval crosses one threshold, very serious when two thresholds are crossed, and extremely serious when the confidence interval is so wide that the estimate is considered uninterpretable (47, 49).

4. Indirectness. We will assess whether there are important differences between the characteristics of studies included in each synthesis and the question we are 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 delivery of Buteyko in the study compared to Australia that are likely to influence the size of effect, such as interventions in which the number or duration of sessions in which participants learn Buteyko is substantially less than that recommended (see Background, 'Mode of administration').

5. Publication bias. Our judgement of suspected publication bias will be based on assessment of bias due to missing results (see 3.3.8). In these assessments, we will also consider the potential impact on each synthesised result of excluding studies in languages other than English.

6. Upgrading domains (large effect size, dose response gradient, opposing plausible residual confounding). For NRSIs, we will consider large effects, any dose response, and opposing residual confounding as part of the ROBINS-I assessment (judging whether any of these factors mitigate concerns that observed effects are due to residual confounding) (45, 58). For this reason, upgrading of NRSIs will not be considered. There is no precedent for rating up the evidence from randomised trials, however in principle, these domains apply to any body of evidence so are included here for completeness.

Using GRADE decision rules, we will derive an overall GRADE for the certainty of evidence for each result included in the summary of findings table (24, 45, 58). 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 serious concerns on any GRADE domain that reduces confidence that Buteyko has at least a small effect (as determined by the pre-specified thresholds) (24, 45, 58).

A result from a body of evidence comprised of NRSIs will also begin as 'high' certainty, with the same rules for rating down used as for randomised trials. The option to rate down by -3 may be considered if there are extremely serious concerns about risk of bias (especially for NRSIs) or imprecision (where results are compatible with both important benefit and important harm) (49, 61).

Summary of findings tables will be prepared using the GRADEpro GDT software (www.gradepro.org). The tables will include:

- estimates of the effects of Buteyko reported as standardised mean differences, and for binary outcomes relative and absolute effects
- the overall GRADE (rating of certainty) and an explanation of the reason(s) for rating down (or up) (62)
- 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 (63).

Where results for a comparison and outcome from both randomised trials and NRSIs are available, these will be presented in the same row of the summary of findings table providing that the certainty of evidence available from randomised trials is not lowered by inclusion of NRSIs at higher risk of bias than the trials (64). This approach will enable an overall conclusion to be drawn from the body of evidence where appropriate to do so. Doing so may also ameliorate concerns about indirectness (if direct evidence is available from NRSIs) and imprecision (if high or moderate certainty evidence is available from NRSIs). If the certainty of evidence is lowered by inclusion of NRSIs at higher risk of bias than RCTs, or if results from NRSIs and RCTs are not coherent (i.e. both leading to an interpretation of a small but important improvement), then we will report the results in separate rows with separate GRADEs and interpretations.

We will present the four levels of certainty of evidence in summary of findings tables with the following symbols and interpretations.

- High (⨁⨁⨁⨁): further research is very unlikely to change the confidence in the estimate of effect
- Moderate (⨁⨁⨁⊖): further research is likely to have an important impact in the confidence in the estimate of effect
- Low (⨁⨁⊖⊖): 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 (⨁⊖⊖⊖): any estimate of effect is very uncertain

4. References

1. National Health and Medical Research Council. Statement of requirement: Evidence evaluations for review of natural therapies (tranche two). 2020.
2. Courtney R. Buteyko breathing method. In: Chaitow L, Bradley D, Gilbert G, editors. Recognising and treating breathing disorders: A multidisciplinary approach. 2nd ed: Churchill Livingstone; 2014. p. 241-7.
3. Illidi C, Romer L, Johnson M, Williams N, Rossiter H, Casaburi R, et al. Distinguishing science from pseudoscience in commercial respiratory interventions: an evidence-based guide for health and exercise professionals. European Journal of Applied Physiology. 2023.
4. Santino TA, Chaves GSS, Freitas DA, Fregonezi GAF, Mendonça K. Breathing exercises for adults with asthma. Cochrane Database of Systematic Reviews. 2020(3).
5. Jones M, Harvey A, Marston L, O'Connell NE. Breathing exercises for dysfunctional breathing/hyperventilation syndrome in adults. Cochrane Database of Systematic Reviews. 2013(5).
6. Bott J, Blumenthal S, Buxton M, Ellum S, Falconer C, Garrod R, et al. Guidelines for the physiotherapy management of the adult, medical, spontaneously breathing patient. Thorax. 2009;64(Suppl 1):i1-i52.
7. Holland AE, Hill CJ, Jones AY, McDonald CF. Breathing exercises for chronic obstructive pulmonary disease. Cochrane Database of Systematic Reviews. 2012(10).
8. Steel A, McIntyre E, Harnett J, Foley H, Adams J, Sibbritt D, et al. Complementary medicine use in the Australian population: Results of a nationally-representative cross-sectional survey. Sci Rep. 2018;8(1):17325.
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. The Buteyko Institute of Breathing and Health Incorporated. Buteyko Institute Method [Internet]. Fitzroy, VIC2018 [Available from: https://buteyko.info/buteyko-method/.
11. Bruton A, Lewith GT. The Buteyko breathing technique for asthma: a review. Complementary therapies in medicine. 2005;13(1):41-6.
12. Austin G. Buteyko technique use to control asthma symptoms. Nursing Times. 2013;109(16):16-7.
13. The Buteyko Institute of Breathing and Health Incorporated. Practitioner training [Internet]. Fitzroy, VIC2018 [Available from: https://buteyko.info/practitioner-training/.
14. Buteyko Health & Breathing Pty Ltd. Course Schedule [Internet]. Fitzroy, VICn.d. [Available from: https://buteykoairways.com.au/learn-buteyko/course-schedule/.
15. Buteyko Clinic International. [Buteyko] Self study online courses [Internet]. Moycullen, GAL: Buteyko Clinic International; 2022 [Available from: https://buteykoclinic.com/buteyko-clinic-online-courses/.
16. Courtney R. Strengths, weaknesses, and possibilities of the Buteyko breathing method. Biofeedback. 2008;36(2):59-63.
17. Australian Health Practitioner Regulatory Association. Public consultation on clearer regulation of medical practitioners who provide complementary and unconventional medicine and emerging treatments. Available from: https://www.ahpra.gov.au/ Access date: 4 February 2021 2021.
18. 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(6):789-807.
19. Mendonça K, Collins S, Santos TZ, Chaves G, Leite S, Santino TA, et al. Buteyko method for people with asthma: a protocol for a systematic review and meta-analysis. BMJ Open. 2021;11(10):e049213.
20. Scottish Intercollegiate Guidelines Network, British Thoracic Society. British guideline on the management of asthma. Edinburgh (UK): Healthcare Improvement Scotland; 2019 (revised). Report No.: 978-1-909103-70-2.
21. National Asthma Council Australia. Australian Asthma Handbook [Internet]. Melbourne, VIC: National Asthma Council Australia; 2023 [Available from: www.asthmahandbook.org.au.
22. Australian Government Department of Health. Review of the Australian Government Rebate on Natural Therapies for Private Health Insurance. 2015.
23. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al., editors. Cochrane Handbook for Systematic Reviews of Interventions version 6.1 (updated September 2022). Cochrane. Available from www.training.cochrane.org/handbook2022.
24. Schunemann HJ, Brozek J, Guyatt G, Oxman AD, editors. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. Hamilton, Canada: McMaster University; 2013.
25. 25. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Systematic reviews. 2015;4(1):1.
26. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. Bmj. 2015;349:g7647.
27. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Bmj. 2021;372:n71.
28. Page MJ, Moher D, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. Bmj. 2021;372:n160.

29. 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.
30. 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.
31. 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 20212019.
32. Reeves B, Deeks J, Higgins J, Shea B, Tugwell P, Wells G. Chapter 24: Including non-randomized studies on intervention effects. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. Cochrane Handbook for Systematic Reviews of Interventions version 63 (updated February 2022) Cochrane, 2022 Available from wwwtrainingcochraneorg/handbook2022.
33. 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.
34. Sterne J, Hernán M, McAleenan A, Reeves B, Higgins J. Chapter 25: Assessing risk of bias in a non-randomized study. In: Higgins JPT TJ, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editor. Cochrane Handbook for Systematic Reviews of Interventions version 63 (updated February 2022) Cochrane, 2022 Available from wwwtrainingcochraneorg/handbook2022.
35. 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: Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Welch V, editors. Cochrane Handbook for Systematic Reviews of Interventions version 63 (updated February 2022) Cochrane, 2022 Available from wwwtrainingcochraneorg/handbook. 2nd ed. Chichester (UK): John Wiley & Sons; 2022.
36. Kadam S, Prem V, Ughreja RA. Effect of breathing re-training on end-tidal carbon-di-oxide in patients with asthma: a systematic review and meta-analysis. J Asthma. 2023:1-10.
37. National Health and Medical Research Council. Draft framework for protocols for systematic reviews of randomised controlled trials and non-randomised studies of interventions. 2020.
38. 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. Journal of Biomedical Informatics. 2009;42(2):377-81.
39. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, et al. The REDCap consortium: Building an international community of software platform partners. Journal of Biomedical Informatics. 2019;95:103208.
40. 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.
41. 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.
42. Sterne J, Savovic J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. Bmj. 2019;366:l4898.
43. Sterne JA, Hernan MA, Reeves BC, Savovic J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. Bmj. 2016;355:i4919.
44. Higgins J, Eldridge S, Li T. Chapter 23: Including variants on randomized trials. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al., editors. Cochrane Handbook for Systematic Reviews of Interventions version 63 (updated February 2022) Cochrane, 2022 Available from wwwtrainingcochraneorg/handbook2022.
45. Schunemann HJ, Cuello C, Akl EA, Mustafa RA, Meerpohl JJ, Thayer K, et al. GRADE guidelines: 18. How ROBINS-I and other tools to assess risk of bias in nonrandomized studies should be used to rate the certainty of a body of evidence. J Clin Epidemiol. 2019;111:105-14.
46. Higgins J, Li T, Deeks Je. Chapter 6: Choosing effect measures and computing estimates of effect. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. Cochrane Handbook for Systematic Reviews of Interventions version 63 (updated February 2022) Cochrane, 2022 Available from wwwtrainingcochraneorg/handbook2022.
47. Zeng L, Brignardello-Petersen R, Hultcrantz M, Siemieniuk RAC, Santesso N, Traversy G, et al. GRADE guidelines 32: GRADE offers guidance on choosing targets of GRADE certainty of evidence ratings. J Clin Epidemiol. 2021;137:163-75.
48. Schünemann HJ, Vist GE, Higgins J, Santesso N, Deeks JJ, Glasziou P, et al. Chapter 15: Interpreting results and drawing conclusions. In: Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Welch V, editors. Cochrane Handbook for Systematic Reviews of Interventions version 63 (updated February 2022) Cochrane, 2022 Available from wwwtrainingcochraneorg/handbook. 2nd ed. Chichester (UK): John Wiley & Sons; 2022.
49. Zeng L, Brignardello-Petersen R, Hultcrantz M, Mustafa RA, Murad MH, Iorio A, et al. GRADE Guidance 34: update on rating imprecision using a minimally contextualized approach. J Clin Epidemiol. 2022;150:216-24.
50. Brennan SE, Johnston RV. Research Note: Interpreting findings of a systematic review using GRADE methods. Journal of Physiotherapy. 2023.
51. Bell ML, McKenzie JE. Designing psycho-oncology randomised trials and cluster randomised trials: variance components and intra-cluster correlation of commonly used psychosocial measures. Psychooncology. 2013;22(8):1738-47.

52. Balk EM, Earley A, Patel K, Trikalinos TA, Dahabreh IJ. Empirical Assessment of Within-Arm Correlation Imputation in Trials of Continuous Outcomes. Methods Research Report. (Prepared by the Tufts Evidence-based Practice Center under Contract No. 290-2007-10055-I.) AHRQ Publication No. 12(13)-EHC141-EF. Rockville, MD: Agency for Healthcare Research and Quality. November 2012. www.effectivehealthcare.ahrq.gov/reports/final.cfm. AHRQ Methods for Effective Health Care2012.
53. Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. BMC Med Res Methodol. 2014;14:135.
54. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med. 2002;21(11):1539-58.
55. Page MJ, Higgins JPT, Sterne JA. Chapter 13: Assessing risk of bias due to missing results in a synthesis. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al., editors. Cochrane Handbook for Systematic Reviews of Interventions version 63 (updated February 2022) Cochrane, 2022 Available from wwwtrainingcochraneorg/handbook2022.
56. 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(10):991-6.
57. McKenzie JE, Brennan SE. Chapter 12: Synthesizing and presenting findings using other methods. In: Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Welch V, editors. Cochrane Handbook for Systematic Reviews of Interventions version 63 (updated February 2022) Cochrane, 2022 Available from wwwtrainingcochraneorg/handbook. 2nd ed. Chichester (UK): John Wiley & Sons; 2022.
58. Schünemann HJ, Higgins J, Vist GE, Glasziou P, Akl EA, Skoetz N, et al. Chapter 14: Completing 'Summary of findings' tables and grading the certainty of the evidence. In: Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Welch V, editors. Cochrane Handbook for Systematic Reviews of Interventions version 63 (updated February 2022) Cochrane, 2022 Available from wwwtrainingcochraneorg/handbook. 2nd ed. Chichester (UK): John Wiley & Sons; 2022.
59. Guyatt G, Zhao Y, Mayer M, Briel M, Mustafa R, Izcovich A, et al. GRADE Guidance 36: Updates to GRADE's approach to addressing inconsistency. J Clin Epidemiol. 2023.
60. Nikolakopoulou A, Higgins JPT, Papakonstantinou T, Chaimani A, Del Giovane C, Egger M, et al. CINeMA: An approach for assessing confidence in the results of a network meta-analysis. PLOS Medicine. 2020;17(4):e1003082.
61. Piggott T, Morgan RL, Cuello-Garcia CA, Santesso N, Mustafa RA, Meerpohl JJ, et al. Grading of Recommendations Assessment, Development, and Evaluations (GRADE) notes: extremely serious, GRADE's terminology for rating down by three levels. Journal of Clinical Epidemiology. 2020;120:116-20.
62. Santesso N, Carrasco-Labra A, Langendam M, Brignardello-Petersen R, Mustafa RA, Heus P, et al. Improving GRADE Evidence Tables part 3: Guidance for useful GRADE certainty in the evidence judgments through explanatory footnotes. J Clin Epidemiol. 2016.
63. Santesso N, Glenton C, Dahm P, Garner P, Akl EA, Alper B, et al. GRADE guidelines: 26. informative statements to communicate the findings of systematic reviews of interventions. Journal of Clinical Epidemiology. 2019.
64. Cuello-Garcia CA, Santesso N, Morgan RL, Verbeek J, Thayer K, Ansari MT, et al. GRADE guidance 24 optimizing the integration of randomized and non-randomized studies of interventions in evidence syntheses and health guidelines. J Clin Epidemiol. 2022;142:200-8.

Appendix 1: Database and register search strategies

Cochrane Central Register of Controlled Trials

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

MEDLINE via Ovid

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

Embase via Ovid

#	Search strategy
1	Buteyko Breathing/
2	buteyko.af. [af=all fields]
3	or/1-2

AMED via Ovid

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

Emcare via Ovid

#	Search strategy
1	Buteyko Breathing/
2	buteyko.af. [af=all fields]
3	or/1-2

CINAHL Plus via EBSCOhost

#	Search strategy
1	SU Buteyko Method
2	TX buteyko
3	S1 OR S2

Europe PMC

#	Search strategy
1	(TITLE:"buteyko") OR (ABSTRACT:"buteyko") OR (("buteyko") AND (SRC:"PPR"))

ClinicalTrials.gov and WHO ICTRP

Buteyko (not limited to any specific field)

Buteyko for any health condition: protocol for a systematic review (PROSPERO ID. CRD42023466774)

Appendix 2: Study design features checklist

The following checklist is from Reeves et al (33). The checklist is intended to be used as a tool for confirming eligibility of studies for the review and describing the design features of each included study. In this case, it is used to indicate how studies will be labelled in the review.

- Responses for each design were completed for this systematic review (not taken from the paper) to indicate the minimum features required for inclusion of each type of study (labelled 'Yes').
- Alternatives to these minimum features are labelled 'Possible'.
- In most cases, the combination of features labelled 'Yes' distinguishes the type of study (i.e. NRCT versus cohort).

Study design feature	Feature of				
	NRCT	Prospective cohort	Retrospective cohort	Controlled ITS	CBA
1. Was the intervention/comparator (answer 'yes' to more than one item, if applicable):					
• allocated to (provided for/administered to/chosen by) individuals?	Yes (or clusters)	Yes	Yes	Possibly	No
• allocated to (provided for/administered to/chosen by) clusters of individuals?	Possibly	No	No	Yes	Yes
• clustered in the way it was provided (by practitioner or organizational unit)?	Possibly	Possibly	Possibly	Possibly	Possibly
2. Were outcome data available (answer 'yes' to only one item):					
• after intervention / comparator only (same individuals)?	Yes	Yes	Yes	No	No
• after intervention/comparator only (not all same individuals)?	No	No	No	No	No
• before (once) AND after intervention/comparator (same individuals)?	Possibly	Possibly	Possibly	No	Possibly
• before (once) AND after intervention/comparator (not all same individuals)?	No	No	No	No	Yes
• multiple times before AND multiple times after intervention/comparator (same individuals)?	Possibly	Possibly	Possibly	Possibly	No (ITS)
• multiple times before AND multiple times after intervention/comparator (not all same individuals)?	No	No	No	Yes	No (ITS)
3. Was the intervention effect estimated by (answer 'yes' to only one item):					
• change over time (same individuals at different time-points)?	No	No	No	No	No
• change over time (not all same individuals at different time-points)?	No	No	No	No	No
• difference between groups (of individuals or clusters receiving either intervention or comparator)?	Yes	Yes	Yes	Yes	Yes
4. Did the researchers aim to control for confounding (design or analysis) (answer 'yes' to only one item):					
• using methods that control in principle for any confounding?	No	No	No	Yes	No
• using methods that control in principle for time invariant unobserved confounding?	No	No	No	No	Yes

Study design feature	Feature of				
	NRCT	Prospective cohort	Retrospective cohort	Controlled ITS	CBA
• using methods that control only for confounding by observed covariates?	Yes [exclude if not]	Yes	Yes	Possibly	Possibly
5. Were groups of individuals or clusters formed by (answer 'yes' to more than one item, if applicable):					
• *randomization?*	No	No	No	No	No
• *quasi-randomization?*	No	No	No	No	No
• some other action of researchers?	Yes	Possibly	Possibly	Possibly	No
• *time differences?*	No	No	No	Yes	No
• location differences?	Possibly	Possibly	Possibly	Possibly	Possibly
• healthcare decision makers/practitioners?	Possibly	Possibly	Possibly	Possibly	Yes
• participants' preferences?	Possibly	Yes	Yes	No	Possibly
• on the basis of outcome? e	No	No	No	No	No
• some other process? (specify)	Possibly	Possibly	Possibly	Possibly	Possibly
6. Were the following features of the study carried out after the study was designed (answer 'yes' to more than one item, if applicable):					
• characterization of individuals/clusters before intervention?	Yes	Yes	Possibly	Possibly	Possibly
• actions/choices leading to an individual/cluster becoming a member of a group?	Yes	Yes	Possibly	Possibly	Possibly
• assessment of outcomes?	Yes	Yes	Possibly	Possibly	Possibly
7. Were the following variables measured before intervention (answer 'yes' to more than one item, if applicable):					
• potential confounders?	Yes	Yes	Yes	Possibly	Possibly
• outcome variable(s)?	Possibly	Possibly	Possibly	Yes	Yes

NRCT: non-randomised controlled trial; ITS: interrupted time series; CBA controlled before-after studies

Appendix 3: TIDIER domains and example of application in Buteyko SR

Characteristic	Description	Standard headings (bolded) and codes illustrated with an example*
Brief name	Name or phrase	Buteyko
Why	Rationale, theory or goal of essential elements	Relief of anxiety
What	Procedures, activities or processes including enabling/supporting activities	Participants were taught three main procedures: control pause, slow breathing, combining control pause with slow breathing. Participants were instructed to practice respiratory exercises in daily activities; they received fortnightly calls to monitor adherence and provide guidance and advice.
	Materials (physical or informational)	**CO-INTERVENTIONS (code all that apply):** written information leaflet, educational material, other [medication; psychiatric counselling], none reported **
When and how much	Number of times delivered, over what period of time, number of sessions, their schedule, and their duration, intensity or dose	Instruction: number and duration of sessions not reported; fortnightly calls. Respiratory exercises: 15-20 minutes per day, at least 4 days per week for 1 month (possibly up to 2 months, but unclear).
How	Modes of delivery ()	**CODE** (choose all that apply): individual lessons, group lessons; in person, online, other [specify]; not reported [individual seems likely]
Who provided	Intervention provider by category and their expertise, background and training	**Administered by a provider**: [Y/N] **Self-administered**: [Y/N] **Provider** (as identified by authors, code all that apply) ○ Buteyko practitioner ○ Other natural therapist [specify] ○ Nurse (clinically qualified) ○ Allied Health [specify e.g. physiotherapist] ○ General practitioner/ physician [specify] ○ Research staff ○ Other [specify] ○ Not reported **Highest level of training** (for health-care providers only): postgraduate, bachelor, diploma, certificate, other [specify], not reported **Trained in Buteyko?** [Y/N/not reported] **Details**: certified in Buteyko procedures
Where	Type of location where intervention occurred	Community-based (home)
Tailoring	If personalised, titrated or adapted describe what, why, when and how	n/a
Modifications	Not collected. Any modifications are likely to be part of protocol (i.e. tailoring)	
How well	Not collected. Used for questions about *adherence* to intervention (not assignment to intervention)	

*Example from doi:10.1589/jpts.34.247
**Codes highlighted in grey are the codes that apply to this example