The efficacy and safety of medicinal cannabis in adult populations: A protocol for an overview of reviews

July 2022

Kathryn Lambe

Áine Teahan

Olivia Cagney

Caitríona Lee

Jean Long

Disclaimer

Any views expressed in this protocol are those of the authors and not necessarily those of the Minister for Health, the Department of Health or Health Research Board.

Published by:

Health Research Board, Dublin

Health Research Board
Grattan House
67-72 Lower Mount Street
Dublin 2
D02 H638
Ireland

t 353 1 234 5000
f 353 1 661 2335
e hrb@hrb.ie
w www.hrb.ie

Table of Contents

List of tables

List of abbreviations

Abbreviation	Explanation
AMSTAR	A MeaSurement Tool to Assess systematic Reviews
CBD	Cannabidiol
GRADE	Grading of Recommendations, Assessment, Development and Evaluations
HSE	Health Service Executive
MCAP	Medical Cannabis Access Programme
MeSH	Medical Subject Headings
PICO	Population/Intervention/Comparator/Outcome
THC	Tetrahydrocannabidiol

1 Background

In January 2017, the Health Products Regulatory Authority (HPRA), at the request of the Minister for Health, convened an expert working group to review the potential medical use of cannabis. The outcome of this review was a report titled "Cannabis for Medical Use: A Scientific Review"[1].

The HPRA advised that any programme to make cannabis available for medical purposes should recognise patient need but be evidence-based. It advised that access to cannabis should be permitted under a controlled access programme for the treatment of patients with one of three stated conditions, who have failed to respond to all other previous treatments, namely:

- Spasticity associated with multiple sclerosis resistant to all standard therapies and interventions whilst under expert medical supervision
- Intractable nausea and vomiting associated with chemotherapy, despite the use of standard anti-emetic regimes whilst under expert medical supervision[2]
- Severe, refractory (treatment-resistant) epilepsy that has failed to respond to standard anticonvulsant medications whilst under expert medical supervision

This recommendation was made on the basis that there was "at least modest evidence that cannabis may be effective" for these conditions [1] p4. Clinical guidelines were published in 2019 [3]. The legislation to establish the Medical Cannabis Access Programme (MCAP) was also enacted that year. The MCAP was added to the Health Service Executive (HSE) Service Plan 2021 and is currently operated by the Primary Care Reimbursement Service.

The Department of Health has received many representations and communications seeking to have the scope of the programme expanded to include other conditions, including chronic pain, fibromyalgia, anxiety, and endometriosis, among others. There is significant media, political, and public interest in this topic.

1.1 Purpose of the review

This evidence review, examining the efficacy and safety of cannabis-based treatments for a range of conditions will be prepared by the Evidence Centre with the aim of supporting the planned 2022 review of the MCAP, including decisions on what conditions are included in the MCAP. The synthesis will also be used to respond to the many communications the Department receives each year on the prescribing of cannabis-based products and will support the Department's position as to what clinical indications are suitable for access to cannabis-based products.

2 Review questions

Q1 What is the evidence for the clinical efficacy of medicinal cannabis in the treatment of the conditions / clinical indications of interest among adults?

Q2 What is the evidence for the safety of medicinal cannabis in the treatment of the conditions / clinical indications of interest among adults?

2.1.1 Scope

The conditions of interest include but are not limited to:

- Inflammatory disorders, including endometriosis

- Sleep disorders
- Parkinson's disease
- Anxiety
- Depression
- Severe refractory epilepsy

The clinical indications of interest include but are not limited to:

- Chronic pain
- Cancer-related pain and appetite-related symptoms
- Appetite-related symptoms due to HIV/AIDS
- Spasticity associated with multiple sclerosis
- Nausea/ vomiting associated with chemotherapy

The above conditions / clinical indications have been selected because the Department of Health has specified them as being of particular interest; however, others will be included in the review if they are found in the literature. The review protocol will be reported in accordance with the Preferred Reporting Items for Systematic reviews and Meta-Analyses PRISMA-P reporting guidance for review protocols and registered on PROSPERO [4].

3 Review design

This evidence review will comprise an overview of reviews (umbrella review).

We chose an overview of reviews for two reasons. First, our scoping searches indicated that the literature is already populated with a number of systematic reviews that are relevant to our review questions. The available reviews vary in design and conduct and comprise both Cochrane and non-Cochrane reviews. Therefore, it would be inappropriate to undertake an original systematic review while ignoring the existing evidence base in systematic reviews. According to Aromataris *et al.,* "if current, multiple, good quality, systematic reviews exist about a given topic or question, any reviewer should reconsider the need to conduct yet another review addressing the same issue. Rather, these [existing reviews] may be the basis to conduct an Umbrella Review [overview of reviews] and summarize or synthesize the findings of systematic reviews already available" [5] p365.

Second, to inform policy decisions around the scope of the Medical Cannabis Access Programme in Ireland, the Department of Health requires information about the efficacy and safety of cannabis-based treatments for a very wide range of conditions/clinical indications. The efficiencies offered by this "hybrid" approach allow for the review to cover the full scope of conditions of interest, which would not be possible with a traditional systematic review in the available time.

3.1 Eligibility criteria

The eligibility criteria for reviews to be included are outlined in Table 1.

Regarding population, the scope of the overview of reviews will be limited to adult patients only, as considerations for adolescent and paediatric patients present different complexities and access may be channelled through separate systems and healthcare providers. Syntheses of data from paediatric patients aged 12 years and under will be excluded. Syntheses from systematic reviews with mixed adult

and adolescent (aged 13-17 years) patients will be excluded if adolescents make up 20% or more of the sample.

Regarding outcomes, misuse or diversion of prescribed products will not be included under adverse events and will not be examined in this review. We believe that a review of primarily randomised controlled trials, in which supply of medical cannabis is tightly controlled and follow-up is generally only short- or medium-term, will not capture these outcomes as effectively as other study designs (e.g. patient registries). Therefore, we have chosen not to explore these outcomes, rather than present only a narrow and potentially unrepresentative slice of data on misuse or diversion.

The outcomes listed are intentionally wide-ranging so as not to exclude any relevant outcomes that may be examined in the literature. As characterised by Lunny *et al.* "Overviews of systematic reviews synthesise the results of multiple systematic reviews. Overviews are typically broader in scope than systematic reviews and may examine different interventions for the same condition, the same intervention for different conditions, or the same intervention for the same condition but focusing on different outcomes" [6] p2.

Regarding date, the date range 2010 – present was chosen to capture systematic reviews from the last 12 years. Based on expert guidance, we expect that this will yield primary research conducted in the last 30 years, which comprehensively covers the period since the first medical cannabis access programme was launched (in Canada in 2001) [5].

Regarding language, only English-language reviews will be included in the final analysis. The databases to be searched (see Section 3.2.4) index primarily English-language material. No language limit will be used in the search strategy. Relevant reviews in non-English languages will be excluded during full-text screening and listed among the excluded studies in an appendix to the final report.

Table 1 Eligibility criteria for overview of reviews

Domain	Inclusion	Exclusion
Population	Adult patients (aged 18 and above) Adolescent patients (aged 13-17 years), provided that they comprise no more than 20% of the sample	Paediatric patients (aged 12 and under) Populations of unspecified age Animals
Intervention	Cannabis-based medicinal products containing natural or synthetic CBD or THC or CBD or THC derivatives	Cannabis for recreational use Cannabis for medicinal use without prescription/medical supervision Systematic reviews including interventions not focused on cannabis-based medicinal products.
Comparator	Other cannabis-based medicinal products/doses/regimens Placebo Any relevant alternative treatment Usual / standard care No treatment	Systematic reviews of studies with no comparator
Outcome	Reduction in relevant symptoms Changes in quality of life Relevant adverse events Withdrawal/complications	Patient satisfaction
Study design	Systematic review of randomised controlled trials and/or prospective longitudinal cohort studies	Systematic reviews of non-randomised trials

		Systematic reviews based on searches of only one bibliographic database Systematic reviews that do not present a full search strategy Systematic reviews without a quality assessment/risk of bias assessment of their included studies, or systematic reviews that used an inappropriate tool for assessment Systematic reviews of descriptive epidemiological studies or case-control studies Systematic reviews where it is not possible to extract data based on outcomes of interest; Systematic reviews where it is not possible to extract data based on study designs of interest Narrative (non-systematic) reviews Primary studies
Date	2010 – May 2022	Pre-2010
Language	English	Non-English languages

3.1.1 Overlapping reviews

To address the issue of overlapping primary studies in this overview of reviews, we will calculate the corrected covered area as a measure of overlap. This approach is recommended by Pieper *et al.* who contend that "all producers of overviews should analyse the overlaps and report their analysis. Reporting should be done even if the amount of overlap is small and unlikely to have an impact on the conclusion. Otherwise, consumers will not know whether there is no meaningful overlap or if the authors simply did not account of it. Consequently, overlaps should be reported by default" [7] p374-375.

For each outcome, the corrected covered area is calculated as follows:

$$Corrected\ cover\ area = \frac{N - r}{r\ c - r}$$

where *N* is the number of included primary publications (including double counting) in the evidence synthesis, *r* is the number of unique primary publications, and *c* is the number of reviews.

3.2 Identifying research evidence

3.2.1 Search approach

A single search for the overview of reviews will be used to answer the two research questions outlined in Section 2. The search plan will have two components: searches of bibliographic databases and grey literature resources, which will be followed up with a range of supplementary search strategies. Aromataris *et al.* recommend that a broad search be used, to maximise the opportunity to capture relevant reviews [5].

3.2.2 Search concepts

The concepts used in the search for this overview of reviews will directly relate to the review questions. The primary concept used in the search will be cannabinoid compounds. The search terms for this

concept will include cannabis-related controlled vocabulary (such as Emtree or Medical Subject Headings/MeSH, for example, 'Cannabis/', 'exp Cannabinoids/', 'Medical Marijuana/') and free terms/keywords (for example, CBD, THC, Nabilone, or Cesamet).

The search will not be limited to specifically or exclusively medical/medicinal cannabis. This is because, while the overview of reviews will examine therapeutic use of cannabis, the terms used in titles, abstracts and indexed keywords to describe studies may not specify the term 'medical cannabis'. The search will aim to capture as much as possible of the available review literature on cannabis administered for therapeutic purposes.

Test searches captured a wide range of literature, including much non-clinical material on, for example, chemical properties of cannabis, rather than the use of cannabis by humans and returned very large numbers of results. The timeframe and practical logistics of this review project would not allow for screening of very large numbers of irrelevant results. For this reason, some clinical terms were tested for use in the free term/keyword Ovid MEDLINE searches to focus the search on cannabis administered therapeutically, for example, the combination of cannabis terms with such terms as '(clinical$ or therap$ or medic$ or trial$ or patient$ or placebo$ or randomi$ or adminis$ or ameliorat$ or treat$ or manage$ or tolerat$ or intervention$ or prescrib$)).mp.' The results excluded by this method were examined for relevancy and it was found that the material excluded using these terms related to non-clinical work on cannabis, such as drug discovery, or horticultural studies. Using these keyword terms did not affect the number of results returned using MeSH terms, which made up the majority of the results. Refining the search in this manner will not be possible for all of the search resources, given the technical limitations of many non-database search resources.

For the overview of reviews, the study type of interest is the systematic review. Because of this, a study design-related limit will be used for the search. We will use specific controlled vocabulary relating to systematic reviews. We will also use wider review-related free terms (including terminology relating to systematic review methodology that we would expect to be included in titles, abstracts and indexed terms, for example, 'risk of bias', or 'handsearch'). This is because systematic reviews are sometimes described using other terminology, such as evidence syntheses, and may not be described in the texts using words such as 'systematic'.

We will not include review outcomes or treatment indications as concepts, in order to capture as much relevant research as possible. If we search for specific outcomes and treatment indications, we will be less likely to find new outcomes; for example, if we search specifically for medical cannabis and epilepsy, we would be unlikely to capture research on medical cannabis and endometriosis (or any other unspecified concept) [8,9].

3.2.3 Search strategies

The initial strategy will be constructed in Ovid MEDLINE, using MeSH (Medical Subject Headings) thesaurus terms and 'free' natural language terms (free terms or keywords). Search terms will be sourced for Ovid MEDLINE via the MeSH Browser and PubReminer. Terms will also be captured by examining abstract and index terms of relevant papers. This strategy will be adapted for use in other databases, using thesaurus terms where available. Abbreviated searches will be used for information resources for which structured searching will be difficult or impossible. The search strategies will be peer-reviewed by another information specialist, in line with best practice [10]. An example search strategy for Ovid MEDLINE for the overview of reviews is uploaded in the PROSPERO search section.

No limits for population age will be applied, to ensure that no reviews with relevant information are inadvertently excluded. No language limit will be applied, so that relevant studies in non-English languages may be identified but not included in the final analysis.

Two limits will be included in the search strategies:

- **Study type:** Only systematic reviews are required.
- **Date:** A date limit will be set from 2010 to the date of the search.

We will use a date limit of 2010-2022 (see Section 3.1) as a search limit. As outlined in the Joanna Briggs umbrella review guidance, this should capture primary studies from the previous 30 years [5], in this case from 1992 to the current year of 2022.

3.2.4 Search resources

A wide range of search resources will be used to maximise the chance of capturing as much relevant material as possible. To this end, the resources selected for the database base stage of searching include biomedical/clinical/social and public health literature databases, systematic review resources, grey literature resources, search engines, preprint and protocol resources. A final set of searches will be carried out after supplemental searches to capture any newly-published relevant reviews – the resources for these searches will include Ovid MEDLINE, the Cochrane Library and Google Scholar.

The resources to be searched for the three stages of the literature search for the review are listed below.

Stage 1 Database searches June 2022:

Resources: Ovid MEDLINE; Ovid Embase; Ovid PsycInfo; EBSCO CINAHL; EBSCO SocINDEX with Full Text; LILACS; SciELO; Agency for Healthcare Research and Quality (AHRQ); Database of Abstracts of Reviews of Effects (DARE); Dopher; JBI Evidence Syntheses; International HTA database; Health Evidence (McMaster University); Health Systems Evidence (McMaster University); Core.ac.uk; BASE: Bielefeld Academic Search Engine; International Association for Cannabinoid Medicines (IACM) databank; DuckDuckgo.com; Google Scholar; Osf.io; Researchsquare; MedRxiv/BioRxiv; PROSPERO; Follow up of existing relevant guidelines and websites to check for referenced reviews

Stage 2 Supplemental searches Dec-Jan 2022

Forward and backward 'chasing' of reference and citation lists of papers selected for inclusion from the screening process; Follow-up of review protocols retrieved from the screening process to capture their associated reviews – results already included or excluded will not be added; Reference checking of published guidelines relating to medical cannabis

Stage 3 Final database search Feb 2022

Ovid MEDLINE; Cochrane Library; Google Scholar

Limits: Date limit: 2010-2022; Language limit: No English language limit will be imposed on the search but only English language reviews will be included in the final synthesis – relevant non-English reviews will be recorded in an Appendix

With respect to the databases to be included, recent work by Goosen *et al.* found that a combination of MEDLINE, Epistemonikos and reference searching was suitable for health-related searches for umbrella reviews [11]. This study also found that, while reviews were predominantly found in database searches, some reviews were found on websites rather than as indexed journal articles. To avoid missing these

types of reviews, search engines such as Google Scholar and DuckDuckGo.com will be included to maximise the opportunity to capture reviews published outside of the traditional indexed journals [11]. Search engines have been shown to be of some use when carrying out literature searches for material which may not be published in traditional formats [12–14]. The disadvantages of using these resources must be considered when using them, in that the algorithms used to create the resources are not available and even geographic location of the searcher can affect the results captured [15]. In the guidance by Aromataris *et al.*, it is suggested that at least two or three grey literature sources should be included [5]. The first stage of searches began in June 2022.

3.2.5 Supplementary search strategies

The main database searches will be supplemented by screening articles in the reference lists of included articles (backward citation chasing) and screening articles that have cited included articles (forward citation chasing). This step of the process will be carried out once a set of papers is included from the full-text screening stage of the database searches. Reference snowballing has previously been shown to be helpful when searching for systematic review (but the study designs were variable in this work) [16]. This step of the review will use Google Scholar to retrieve citations of reviews to be included, and Dimensions to retrieve reference lists where possible. Citation counts from Google Scholar can vary and are known not be completely accurate, including duplicates and errors. These references and citations will be screened using the same inclusion criteria as for the database search results.

An updated search will be carried out on selected databases (Ovid MEDLINE, Cochrane Library, Google Scholar) after all other elements of the search have been completed, to ensure the most up-to-date coverage possible.

3.3 Review selection

Eppi-Reviewer [17] will be used to manage the screening process. First, each review's title and abstract will be reviewed against the eligibility criteria in Table 1 by two independent reviewers. The two reviewers will then compare their included and excluded reviews and resolve any disagreements through discussion. Any reviews not excluded at this stage will be sourced for full-text screening, again by two independent reviewers. Following full-text screening, any reviews that meet inclusion criteria or do not meet exclusion criteria will be considered for inclusion. The review selection process will be presented in a complete Preferred Reporting Items for Systematic reviews and Meta-Analyses flowchart [18], showing the numbers of reviews excluded at each stage of screening.

3.4 Data extraction

We will use an amended version of the Joanna Briggs Institute data extraction form [5] (see Appendix A) for systematic reviews and research syntheses to extract data from each included systematic review. The extracted data includes citation details, objectives of the review, participants, setting, interventions, comparators, search information, study date range, number of primary studies, study design, risk of bias tool used, risk of bias assessment including publication bias, analysis methods, outcomes assessed, and results by outcome(s). Our amendments to the tool include additional notes, to ensure that all reviewers undertaking extraction make decisions using the same parameters, and additional items for extraction to capture data to be used in quality assessment (see Section 3.5).

Data extraction will be carried out by one reviewer and validated by another.

Data will be extracted at the level of the included systematic reviews only, not at the level of the primary studies included therein. Following expert guidance, extraction and presentation of data will be limited to

the findings presented by the included systematic reviews; while primary studies included in the systematic reviews may be retrieved to check the accuracy of extraction by systematic review authors where necessary, extra data will not be extracted directly from the primary studies [5].

Full descriptive characteristics of the systematic review will be presented in tabular format in the appendices of the review.

3.5 Quality assessment

The AMSTAR 2 instrument [19] will be used to assess the quality and risk of bias of each included systematic review. The AMSTAR 2 instrument has been used in one previous HRB evidence review and allows for the appraisal of systematic reviews of both randomised and non-randomised studies of healthcare interventions, which makes it highly appropriate for this review. Two reviewers will independently apply the instrument to each included systematic review. Discrepancies in scores will be resolved through discussion.

The AMSTAR 2 instrument contains 16 items. The original text of the items will be used. However, having piloted the tool and used it in a previous HRB evidence review, we have made a number of adjustments. Please note that these adjustments are not intended to fundamentally alter the items, but merely to provide more explicit guidance and ensure that all reviewers make decisions using the same parameters:

- The scoring of Items 1, 4, and 8 has been adjusted to provide consistent and more stringent judgement of the parameters being scrutinised.
- For items 1-4, 8, 9, and 11-16, we have added text to further explain and clarify what is required for each parameter.

The adapted instrument is included in Appendix B (see Table 5).

Shea *et al.* recommend defining critical domains before beginning appraisal of a systematic review; these are domains in which identification of weaknesses should undermine confidence in the results of the review. According to Shea *et al.* "responses to AMSTAR 2 items should not be used to derive an overall score. We accept that an overall score may disguise important weaknesses that should diminish confidence in the results of a systematic review, and we recommend that users adopt the rating process based on identification of critical domains, or some variation based on these principles." [19] p6

In the absence of clear definitions from Shea *et al.*, we regard a **critical domain** as a fundamental characteristic of a study design that is essential for its validity (e.g., adequate randomisation in a randomised controlled trial, excessive loss to follow-up in a cohort study). We regard a **non-critical weakness** as a weakness or failing in a non-critical domain. We regard a **critical flaw** as a weakness or failing in a critical domain. We regard a **fatal flaw** as a failing in a critical domain that renders the study ineligible for inclusion (see Table 1).

Shea *et al.* suggest seven critical domains in the AMSTAR 2 instrument that reviewers may use to identify important weaknesses or flaws in systematic reviews. However, reviewers can change some of these domains depending on the focus of their overview. Reflecting our exclusion criteria (see Table 1), we will exclude reviews that do not meet the criteria in domains 2 (adequacy of the literature search) and 4 (risk of bias from individual studies included in the review). We have identified eight rather than seven critical domains (see Appendix B Table 6 for selected domains and justifications).

We will also allocate each included systematic review a confidence rating using the schema from Shea *et al.* [19], shown in 2.

Table 2 Rating overall confidence in the results of individual systematic reviews

Score	Criteria
High	No or one non-critical weakness: the systematic review provides an accurate and comprehensive summary of the results of the available studies that address the question of interest
Moderate	More than one non-critical weakness*: the systematic review has more than one weakness but no critical flaws. It may provide an accurate summary of the results of the available studies that were included in the review
Low	One critical flaw with or without non-critical weaknesses: the review has a critical flaw and may not provide an accurate and comprehensive summary of the available studies that address the question of interest
Critically low	More than one critical flaw with or without non-critical weaknesses: the review has more than one critical flaw and should not be relied on to provide an accurate and comprehensive summary of the available studies
*Downgrade	*Multiple non-critical weaknesses may diminish confidence in the review, and it may be appropriate to move the overall appraisal down from moderate to low confidence.

Source: Shea *et al.* 2017 [19]

4 Synthesis

4.1 Collecting and presenting data on descriptive characteristics of included systematic reviews

As described in Section 3.4, we will use the Joanna Briggs Institute data extraction form [5] (see Appendix A) for systematic reviews and research syntheses to extract review characteristics data from each included systematic review. Data extraction will be carried out by one reviewer and validated by another.

Descriptive data on the review characteristics will be documented in tabular form. For each included systematic review, the extracted data will be presented in two formats: a high-level summary taking account of the quality of evidence, presented in the main report, and detailed structured summaries, presented in the appendices of the main report. PICO and other study characteristics will be extracted and presented in the appendices to demonstrate to the reader why each study was included.

The main report will also present information on the overlap of primary papers evaluating the same intervention for the same outcomes across one or more systematic reviews using the Pieper *et al.* corrected cover area method [7].

4.2 Collecting, analysing, and presenting outcome data

Gates *et al.* [20] describe a number of challenges in synthesising findings from multiple systematic reviews, including heterogeneity of outcome measures, procedural variation at the level of individual systematic reviews, multiple comparisons and discordant results and conclusions across different systematic reviews.

The outcomes specified *a priori* are intentionally wide-ranging so as not to exclude any relevant outcomes that may be examined in the literature on any of the conditions/clinical indications of interest. As characterised by Lunny *et al.* "Overviews of systematic reviews synthesise the results of multiple systematic reviews. Overviews are typically broader in scope than systematic reviews and may examine

different interventions for the same condition, the same intervention for different conditions, or the same intervention for the same condition but focusing on different outcomes" [6] p2.

As described in Section 3.4, we will use the Joanna Briggs Institute data extraction form [5] (see Appendix A) for systematic reviews and research syntheses to extract outcome data from each included systematic review. Data extraction will be carried out by one reviewer and validated by another.

We will extract and compile all relevant data from each included systematic review pertaining to the assessment and reporting of outcomes in each review. Findings for outcomes related to efficacy (e.g. reduction in relevant symptoms, changes in quality of life) and safety (e.g., relevant adverse events, withdrawal/complications) will be presented separately, in accordance with the two research questions (see Section 2).

4.3 Assessing the quality of evidence of outcome data

4.3.1 The Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach

The Grading of Recommendations, Assessment, Development and Evaluations (GRADE) system is framework recommended by the Cochrane Handbook [21] to facilitate the transparent rating of quality of evidence for systematic reviews. While the AMSTAR 2 instrument described in Section 3.5 rates the *methodological quality* of individual systematic reviews, the GRADE approach is used to rate the *quality of the body of evidence* for each outcome across all studies. To illustrate the distinction, a systematic review can be of high methodological quality (e.g., with comprehensive searching, rigorous data extraction, and appropriate synthesis techniques) but identify only low-quality evidence for the outcomes of interest (e.g., a lack of randomised controlled trials, studies with small sample sizes).

Under the GRADE system, the initial certainty of the evidence is determined based on study design, with randomised controlled trials providing a high degree of certainty and observational studies providing a low degree of certainty. The level of certainty is then adjusted upwards or downwards based on a number of factors. Ultimately, a body of evidence related to an outcome receives one of four grades: high, moderate, low, or very low, reflecting the level of confidence we may have that the true effect is similar to (or substantially different from) the estimate of the effect.

4.3.2 Challenges of applying GRADE to overviews of reviews

The GRADE approach has been traditionally applied to rating the quality of evidence in single systematic reviews, primarily reviews that include a meta-analysis. However, there is a lack of consensus on how best to apply a GRADE assessment when undertaking an overview of reviews. The following extract from Gates *et al.* [20] elaborates these difficulties:

"It may not be possible or appropriate to simply extract existing GRADE appraisals from the included systematic reviews. The reviews might not include GRADE appraisals for the outcomes or populations of interest or be missing details on each of the GRADE considerations. Different systematic reviews with the same studies that have made different decisions about handling data (analysis) and appraising study quality may come to different GRADE conclusions, especially related to the study limitations, consistency, and precision domains. Different raters across systematic reviews could come to different conclusions, due to the subjectivity of the GRADE approach. If re-doing the GRADE for each systematic review, authors are likely to encounter difficulty due to an absence of guidance on how to apply GRADE in the context of an overview, incomplete reporting at the level of the systematic review, and a lack of familiarity with the contributing primary studies." [20] p16

These difficulties notwithstanding, we believe that it is important to assess the quality of evidence in this overview of reviews, given the intended purpose of the review to inform decision-making by the Department of Health in relation to the scope of Ireland's MCAP. GRADE is the framework recommended by the Cochrane Handbook to facilitate the transparency rating of quality of evidence. However, following a 2016 study attempting to apply GRADE in an overview of reviews, Pollock *et al.* [22] concluded that "Within our overview, reviewers found that current GRADE guidance was insufficient to make reliable and consistent judgments." [22] p1

In an effort to overcome some of these challenges to applying GRADE in an overview of reviews, Pollock *et al.* developed a modified algorithm to grade the quality of evidence in their overview. Our approach to applying GRADE will be based on this algorithm. We will apply the modified algorithm to all reviews included in our overview of reviews. If individual included reviews have applied the original GRADE assessment, we will refrain from using these assessments; this is because we want to avoid re-reporting potential conflicting uses of the original instrument by different review teams. Additionally, the original instrument is comparatively more subjective than the more objective modified algorithm, and we want to avoid mixing the two.

4.3.3 Pollock *et al.*'s modified GRADE algorithm

Pollock *et al.*'s [22] algorithm for applying GRADE to an overview of reviews is based on four criteria:

1. The number of participants within the analysis considering imprecision based on sample size and confidence intervals around outcomes of interest
2. The risk of bias within the trials contributing participants to the analysis with respect to randomisation and blinding
3. The statistical inconsistency or heterogeneity within the analysis, as determined by I^2; and
4. The methodological quality of the review as determined by the selection of critical factors from the quality assessment tool. These can be adapted depending on the subject matter of the review.

Each review starts with a ranking of 'high' certainty. The ranking may then be downgraded 1 level for serious methodological concerns (sample size between 100 and 199 participants; high risk of bias in randomization and blinding for > 75% included studies; high heterogeneity (I^2 > 75%); and 'No' on one of these AMSTAR items: a priori research design, comprehensive literature search, duplicate study selection, or duplicate study abstraction). The ranking may be downgraded 2 levels for very serious concerns (sample size < 100 participants and 'No' on two or more of these AMSTAR 2 items: a priori research design, comprehensive literature search, duplicate study selection, or duplicate study abstraction).

We have modified the criteria to be applied to rate the overall quality of each systematic review. While we acknowledge that our modifications will make it less straightforward to compare our findings to those of other overviews of reviews, we believe that this change is necessary because the criteria nominated by Pollock *et al.* were based on their use of the original AMSTAR. As we are using AMSTAR 2 to assess the methods quality of each review, the eight criteria we have nominated are more appropriate to our assessment than the four nominated by Pollock *et al.* Our rationale for our choice of the eight criteria is presented in Appendix B (see Table 6). Our nominated criteria (critical domains) are:

- Research questions and inclusion criteria for the review include the components of PICO (item 1)
- Protocol registered before commencement of the review (item 2)
- Adequacy of the literature search (item 4)

- Risk of bias and publication bias based on primary studies being included in the systematic review (item 9, covered as a separate item in Pollock *et al.* 2016 criteria)
- Appropriateness of meta-analytical methods (item 11)
- If meta-analysis was performed, did the review authors assess the potential impact of risk of bias in individual studies on the results of the meta-analysis or other evidence synthesis (item 12)?
- Consideration of risk of bias when interpreting the results of the review (item 13, covered as a separate item in Pollock *et al.* 2016 criteria)
- Did the review authors provide a satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review (item 14, covered as a separate item in Pollock *et al.* 2016 criteria)?

These modifications are modest and do not materially change the principles of the formula nominated by Pollock *et al*. A full elaboration of how we intend to apply the GRADE algorithm are outlined in the formula below in Table 3.

Table 3 Formula for applying GRADE level of evidence to reviews included in overview of reviews using modified Pollock et al. algorithm

AREA ASSESSED Method of assessment	**IMPRECISION (BASED ON SAMPLE SIZE)** Adequate number of participants included in the pooled analysis	**RISK OF BIAS (TRIAL QUALITY)** Proportion of study participants included in the pooled analysis from primary trials or studies judged to have low risk of bias for randomisation and observer blinding	**INCONSISTENCY** Statistical heterogeneity or inconsistency (e.g., assessed by I^2 or Q statistic)	**RISK OF BIAS (REVIEW QUALITY)** Responses to five AMSTAR questions 1, 2, 4, 11, 12 (see bulleted list above)	**STUDY DESIGN**
No downgrade (no serious limitations)	≥ 200 and included design effect for clustering (that there are more than 1 restoration per participant	≥ 75% of study participants included in the pooled analysis from primary trials or studies judged to have low risk of bias for randomisation and observer blinding	$I^2 \leq 75\%$	5/5 are all "yes" (i.e., low ROB)	Randomised study design
Downgrade 1 level (serious limitations)	100-199	< 75% of study participants included in the pooled analysis from primary trials or studies judged to have low risk of bias for randomisation and observer blinding	$I^2 > 75\%$	4/5 are "yes" and 1 is "partial" or "no" on AMSTAR 2	Non-randomised or cohort study design
Downgrade 2 levels (very serious limitations)	1-99			3/5 are "yes" and remainder are "partial" or "no" on AMSTAR 2	
Notes		If ROB for individual trials is not reported within the review, we can assume that less than 75% of participants had low ROB.	If only one trial contributed to analysis, no downgrade; if I^2 not reported, assumed to be greater than 75%.		

Source: Adapted from Pollock *et al.* 2016 [22]

The number of downgrades that can be applied using the modified algorithm ranges from 0-6 and on this basis, ratings can be applied using the standard GRADE level of evidence. Table 44 displays the system we will use to determine the rating of levels of evidence in our overview of reviews.

Table 4 Classification of GRADE level of evidence to overview of reviews from number of downgrades determined using the Pollock et al. modified algorithm

GRADE level of evidence	Number of downgrades (derived from objective assessment)
High	Score awarded when 0 downgrades are applied
Moderate	Score awarded when 1 or 2 downgrades are applied
Low	Score awarded when 3 or 4 downgrades are applied
Very low	Score awarded when 5 or 6 downgrades are applied

Source: Adapted from Pollock *et al.* 2016 [22]

For example, one downgrade may be applied to a review where inconsistency/heterogeneity is not or cannot be dealt with appropriately. Two downgrades may be applied where there is imprecision, based on inadequate sample size within the pooled analysis. Two downgrades may be applied if the review quality or risk of bias is one of the critical domains.

4.4 Interpreting outcome data and drawing conclusions

Gates *et al.* [20] describe a number of challenges in synthesising findings from multiple systematic reviews, including heterogeneity of outcome measures, procedural variation at the level of individual systematic reviews, multiple comparisons and discordant results and conclusions across different systematic reviews.

To address these challenges, we will use the Six Item Framework proposed by Lunny *et al.* [23] to synthesise our interpretations and conclusions. We will therefore:

1. Elaborate our interpretation and conclusions,
2. Summarise the results from included systematic reviews,
3. Assess and report on heterogeneity,
4. Assess and report on risk of bias in the reviews,
5. Assess and report on overlap of primary studies included in more than one systematic review, and
6. Assess and report on discordant results, interpretations, and conclusions among the included reviews.

5 Deviations from protocol

Any deviations from the protocol will be clearly noted in the final report, along with the rationale for the deviation. Any published versions of the protocol (e.g., PROSPERO) will also be updated.

6 References

1 Health Products Regulatory Authority. Cannabis for medical use: A scientific review. Dublin, Ireland: Health Products Regulatory Authority, 2017. Available from: https://www.hpra.ie/docs/default-source/publications-forms/newsletters/cannabis-for-medical-use---a-scientific-review.pdf

2 Lefebvre C, Glanville J, Briscoe S, *et al.* Technical supplement to chapter 4: Searching for and selecting studies. In: Thomas J, Chandler J, Cumpston M, *et al.*, eds. *Cochrane Handbook for Systematic Reviews of Interventions Version 6.3 (updated February 2022)*. Cochrane, 2019. 67–107. Available from: https://doi.org/10.1002/9781119536604.ch4

3 Department of Health. Clinical guidance on cannabis for medical use. 2020. Available from: https://assets.gov.ie/46697/f1efaeedcfda4d258c6908f86e86056f.pdf

4 Shamseer L, Moher D, Clarke M, *et al.* Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ* 2015;349:g7647. Available from https://doi.org/10.1136/bmj.g7647

5 Aromataris E, Fernandez R, Godfrey C, *et al.* Umbrella Reviews. In: Aromataris E, Munn Z, eds. *JBI Manual for Evidence Synthesis*. JBI; 2020. Available from https://doi.org/10.46658/JBIMES-20-11

6 Lunny C, Brennan SE, McDonald S, *et al.* Toward a comprehensive evidence map of overview of systematic review methods: paper 1—purpose, eligibility, search and data extraction. *Syst Rev* 2017;6:231. Available from https://doi.org/10.1186/s13643-017-0617-1

7 Pieper D, Antoine S-L, Mathes T, *et al.* Systematic review finds overlapping reviews were not mentioned in every other overview. *J Clin Epidemiol* 2014;67:368–75. Available from https://doi.org/10.1016/j.jclinepi.2013.11.007

8 Frandsen TF, Nielsen MFB, Eriksen MB. Avoiding searching for outcomes called for additional search strategies: A study of Cochrane review searches. *J Clin Epidemiol* 2022;S0895435622001378. Available from https://doi.org/10.1016/j.jclinepi.2022.05.015

9 Tsujimoto Y, Tsutsumi Y, Kataoka Y, *et al.* Around ten percent of most recent Cochrane reviews included outcomes in their literature search strategy and were associated with potentially exaggerated results: A research-on-research study. *J Clin Epidemiol* 2022;141:74–81. Available from https://doi.org/10.1016/j.jclinepi.2021.08.030

10 McGowan J, Sampson M, Salzwedel DM, *et al.* PRESS Peer Review of Electronic Search Strategies: 2015 guideline statement. *J Clin Epidemiol* 2016;75:40–6. Available from https://doi.org/10.1016/j.jclinepi.2016.01.021

11 Goossen K, Hess S, Lunny C, *et al.* Database combinations to retrieve systematic reviews in overviews of reviews: a methodological study. *BMC Med Res Methodol* 2020;20:138. Available from https://doi.org/10.1186/s12874-020-00983-3

12 Briscoe S. Web searching for systematic reviews: a case study of reporting standards in the UK Health Technology Assessment programme. *BMC Res Notes* 2015;8:153. Available from https://doi.org/10.1186/s13104-015-1079-y

13 Haddaway NR, Collins AM, Coughlin D, *et al.* The role of Google Scholar in evidence reviews and Its applicability to grey literature searching. *PLoS ONE* 2015;10:e0138237. Available from https://doi.org/10.1371/journal.pone.0138237

14 Stansfield C, Dickson K, Bangpan M. Exploring issues in the conduct of website searching and other online sources for systematic reviews: how can we be systematic? *Syst Rev* 2016;5:191. Available from https://doi.org/10.1186/s13643-016-0371-9

15 Cooper C, Lorenc T, Schauberger U. What you see depends on where you sit: The effect of geographical location on web-searching for systematic reviews: A case study. *Res Synth Methods* 2021;12:557–70. Available from https://doi.org/10.1002/jrsm.1485

16 Horsley T, Dingwall O, Sampson M. Checking reference lists to find additional studies for systematic reviews. *Cochrane Database Syst Rev*, 2011;8:MR000026. Available from https://doi.org/10.1002/14651858.MR000026.pub2

17 Thomas J, Graziosi S, Brunton J, *et al.* EPPI-Reviewer: advanced software for systematic reviews, maps and evidence synthesis [software]. 2022. Available from https://eppi.ioe.ac.uk/cms/Default.aspx?tabid=2967

18 Page MJ, McKenzie JE, Bossuyt PM, *et al.* The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *PLoS Med* 2021;18:e1003583. Available from https://doi.org/10.1371/journal.pmed.1003583

19 Shea BJ, Reeves BC, Wells G, *et al.* AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ* 2017;j4008. Available from https://doi.org/10.1136/bmj.j4008

20 Gates M, Gates A, Guitard S, *et al.* Guidance for overviews of reviews continues to accumulate, but important challenges remain: a scoping review. *Syst Rev* 2020;9:254. Available from https://doi.org/10.1186/s13643-020-01509-0

21 Schünemann HJ, Higgins JP, Vist GE, *et al.* Completing 'Summary of findings' tables and grading the certainty of the evidence. In: Higgins JPT, Thomas J, Chandler J, *et al.*, eds. *Cochrane Handbook for Systematic Reviews of Interventions*. Wiley 2019. 375–402. Available from https://doi.org/10.1002/9781119536604.ch14

22 Pollock A, Farmer SE, Brady MC, *et al.* An algorithm was developed to assign GRADE levels of evidence to comparisons within systematic reviews. *J Clin Epidemiol* 2016;70:106–10. Available from https://doi.org/10.1016/j.jclinepi.2015.08.013

23 Lunny C, Brennan SE, McDonald S, *et al.* Toward a comprehensive evidence map of overview of systematic review methods: paper 2—risk of bias assessment; synthesis, presentation and summary of the findings; and assessment of the certainty of the evidence. *Syst Rev* 2018;7:159. Available from Https://doi.org/10.1186/s13643-018-0784-8

Appendix A Joanna Briggs Institute Data Extraction Form for Review for Systematic Reviews and Research Syntheses

Parameter	Extraction items
First author and year of publication	
Objectives Report exact review question(s) and page number	• PICO elements reported in Introduction/Methods • Study objectives • Exact review question and page number
Participants (characteristics and numbers) The defining characteristics of the participants in studies included in the research syntheses/review should be detailed, for example this may include diagnostic criteria, age, or ethnicity. The total number of participants that inform the outcomes relevant to the umbrella review question from all studies included studies should be presented.	For whole sample and subgroups: • Number of participants • Age • Gender • Details of clinical diagnosis/indications
Setting/context Details of the setting of interest such as acute care, primary health care, or the community or a geographical location should be included. For some umbrella reviews, particularly those that draw upon qualitative research syntheses, the context that underpins the review question will be important to clearly reveal to the reader and may include but is not limited to consideration of cultural factors such as geographic location and specific racial or gender based interests.	• Countries (alphabetic order) • Setting (university, public or private clinic) • Other relevant features of setting
Description of Interventions/ phenomena of interest Clear, succinct details of the interventions or phenomena of interest should be presented as described by systematic review author(s), including the type of intervention, the frequency, and/or intensity of the intervention. A statement of the phenomena of interest is also required where applicable.	• Exact definition of the intervention as per authors • Dose and regimen • Administration methods • Cannabis-related properties • Comparator • Timeframe for follow-up
Databases and sources searched The number of sources searched should be reported. Though this will have been considered during critical appraisal of the research synthesis, reporting to the reader of the review will allow rapid and easy comparison between differences across included reviews and also consideration of potential for publication bias in the event that no formal analysis has been conducted. Where possible the names of databases and	• Number and names of databases • Other sources • Grey literature • Reference chasing Yes/No

Parameter	Extraction items
sources should be listed (i.e. if <5-10). The search range of each database should also be included.	• Expert consultation Yes/No • Dates • Search limits • Justifications for search limits • Other searches • Protocol prepared Yes/No • If yes, published Yes/No, if yes, number and link • Search strategy / key words provided • Screening completed in duplicate Yes/No • If yes, rate of agreement • Extraction completed in duplicate Yes/No • If Yes, rate of agreement • Funding of review • Conflicts of interest of review • How conflicts of interest were managed •
Date Range (years) of included studies The date range spanning from the earliest study that informs the included research synthesis to the latest should be reported. This is important information that allows for consideration of the currency of the evidence base not necessarily reflected in the year of publication of the research synthesis. If this is not readily identifiable in the table of study characteristics provided by the included synthesis, it should be discerned by scanning the date range of publications through the results section of the included systematic review.	• Exact years for included studies
Number of primary studies included in the systematic review Summary descriptive details of the included studies in the research synthesis should be reported. This includes the number of studies in the included research synthesis, the types of study designs included in the research synthesis, for example randomized controlled trials, prospective cohort study, phenomenology, ethnography etc., and also the country of origin of the included studies. The latter is important to allow the reader of the review to consider the external validity and generalizability of the results presented.	• Number of studies • Research designs • Number of studies by study design • Study years • Funding of included studies

Parameter	Extraction items
	• Conflicts of interest of included studies
Types of studies included	• Planned study designs to be included • Reasons for including only RCTs/prospective cohort studies • List of excluded studies at full text and reasons for exclusion
Country of origin of included studies	• Country names in alphabetic order
Appraisal instruments used The instrument or tool used to assess risk of bias, rigour or study quality should be reported along with some summary estimate of the quality of primary studies in the included research synthesis. For example, for umbrella reviews that use the Jadad Scale, a mean score for quality may be reported whereas for checklist appraisals, reporting of cut-off score or any ranking of quality should be reported. An example of the latter would be exclusion of studies that score <3/10, and inclusion of four moderate quality studies (4-6/10) and two high quality studies (7-10/10).	• Full name of tools used For RCTs, record Yes/No for appraisal instrument assessment of: • Concealment of allocation • Blinding of assessors • Sequence allocation (individual vs group randomisation) • Selective reporting For prospective cohort studies: • Confounding • Selection bias • Exposure and outcomes • Selective reporting
Appraisal ratings	• Number of studies by high risk of bias, medium and low • Number of studies out of total number of studies that were at low risk of bias for randomization and at low risk of bias for outcome ascertainment • Authors' exact comments on risk of bias and how it affected analysis and quality of evidence • Graphical or statistical test fro publication bias

Parameter	Extraction items
	• Authors' comments likelihood and magnitude of publciation bias • Authors' comment on how publication bias was dealt with • Only low ROB RCTs included in review Yes/No • Only low ROB RCTs included in meta-analysis Yesn/No • If RCTs with moderate or high ROB or non-randomised studies of interventions were included in the review, discussion of likely impact of ROB on results and quality of evidence or limitations included in conculsions or summary
Method of analysis The type of research synthesis as stated by the authors of the included review should be detailed. The method of analysis or synthesis used by the included research synthesis should be reported. For example, this may include narrative synthesis, vote counting, random effects meta-analysis, fixed effect meta-analysis, network meta-analysis, thematic synthesis, meta-aggregative synthesis, or meta-ethnography.	• Description of method of analysis as per authors • Justification for narrative synthesis or meta-analysis • Justification for combining data in meta-analysis
Outcome assessed Included here should be the outcomes of interest to the umbrella review question reported on by the research synthesis, i.e. the names or labels of the outcomes (see below for presentation of results).	• List of outcomes assessed and intended time frames • Actual timeframes
Results/findings The relevant findings or results presented by the included research syntheses must be extracted. For quantitative reviews, this will ideally be an effect estimate with 95% Cis or measure from a presented meta- analysis. Measures of heterogeneity should also be extracted where applicable. In the absence of this a statement indicating the key result relevant to an outcome may be inserted in the required field. For qualitative syntheses, the key synthesized finding should be extracted.	• Findings by outcome • GRADE by outcome • Meta-analysis results if available (relative risk, odds ratio, standardised mean difference, 95% confidence intervals, I^2, number of trials or studies, number of participants or teeth, random or fixed effects) • Relative risk, odds ratio, standardised mean difference, 95% confidence intervals and p-value for individual studies

Parameter	Extraction items
	where meta-analysis is not available • Appropriate weighted technique used, adjusted for heterogeneity where necessary • Separate summaries reported for RCTs and prospective cohort studies when included in the same review Yes/No For prospective cohort studies: • Combined effect estimates adjusted for confounding, rather than combining raw data • Justification for combining raw data provided, where adjusted effect estimates unavailable
Significance/direction	See above if results listed by outcome
Heterogeneity	See above if I^2 available • Authors' comment on potential impact of heterogeneity on results and quality of evidence • Causes of heterogeneity investigated
Comments There should be provision to extract and present in the table of included study characteristics any relevant details or comments on the included research synthesis by the authors of the Umbrella Review. These comments may be relevant details regarding the included research synthesis, for example, the congruence between the review results and conclusions, and for highlighting any potential methodological differences between the individual included reviews.	

Appendix B Quality assessment tool for systematic reviews: AMSTAR 2

HRB-adapted AMSTAR 2 instrument

Having piloted the AMSTAR 2 tool and used it in a previous HRB evidence review, we have made a number of adjustments in order to ensure that all reviewers are making decisions using the same parameters:

- The scoring of Items 1, 4, and 8 has been adjusted to provide consistent and more stringent judgement of the parameters being scrutinised.
- For items 1-4, 8, 9, and 11-16, we have added text to further explain and clarify what is required for each parameter.
- References to non-randomised studies of interventions have been replaced by references to prospective cohort studies, as these are the only non-randomised studies included in our eligibility criteria.

The adapted instrument appears in Table 5. The notation for the HRB adapted version of AMSTAR 2 is as follows:

- **An asterisk** * following a number denotes a critical factor.
- **Text in red** indicates an exclusion factor.
- **Text in purple** indicates agreed adaptions and interpretation

These factors will be included in the screening criteria. Any systematic review that searched only one bibliographic database or has not completed any quality assessment or risk of bias assessment will be excluded.

Table 5 HRB-adapted AMSTAR 2 instrument

Item		Scoring
1*	**Did the research questions and inclusion criteria for the review include the components of PICO?** Four of the five components must be in the Introduction or Methods to be awarded Yes: For Yes to PICO: ☐ Population ☐ Intervention ☐ Comparator ☐ Outcome ☐ Timeframe for follow-up	☐ Yes ☐ No
2*	**Did the report of the review contain an explicit statement that the review methods were established prior to the conduct of the review and did the report justify any significant deviations from the protocol? The protocol must be accessible to check that the parameters below are covered.** For Partial Yes: The protocol must be reported as prepared and accessible	☐ Yes ☐ Partial Yes ☐ No

	The authors state that they had a written protocol or guide that included ALL the following: ☐ review question(s) ☐ a search strategy ☐ inclusion/exclusion criteria ☐ a risk of bias assessment For 'full' Yes: Protocol must be registered and accessible As for partial yes, plus the protocol should be registered and should also have specified: ☐ a meta-analysis/synthesis plan, if appropriate, and ☐ a plan for investigating causes of heterogeneity ☐ justification for any deviations from the protocol	
3	**Did the review authors explain their selection of the study designs for inclusion in the review?** Authors must have justified their rationale for selecting the study design to be awarded Yes If study design is provided a-priori but without an explanation, score No For Yes, the review should satisfy ONE of the following: ☐ Explanation for including only RCTs ☐ OR Explanation for including only prospective cohort studies ☐ OR Explanation for including both RCTs and prospective cohort studies	☐ Yes ☐ No
4*	**Did the review authors use a comprehensive literature search strategy?** For Partial Yes (all of the following): ☐ searched at least two databases (relevant to research question) (Exclude if only one database was searched – fatal flaw) ☐ provided key word and/or search strategy ☐ justified publication restrictions (e.g., language and/or duration of search) For 'full' Yes (two or more of the following): ☐ searched the reference lists/bibliographies of included studies ☐ searched trial/study registries ☐ where relevant, searched for grey literature ☐ conducted search within 24 months of completion of the review ☐ included/consulted experts in the field	☐ Yes ☐ Partial Yes ☐ No
5	**Did the review authors perform study selection in duplicate?** For Yes, either ONE of the following: ☐ at least two reviewers independently agreed on selection of eligible studies and achieved consensus on which studies to include ☐ OR two reviewers selected a sample of eligible studies AND achieved good agreement (at least 80	☐ Yes ☐ No

	per cent), with the remainder selected by one reviewer	
6	**Did the review authors perform data extraction in duplicate?** For Yes, either ONE of the following: ☐ at least two reviewers achieved consensus on which data to extract from included studies ☐ OR two reviewers extracted data from a sample of eligible studies AND achieved good agreement (at least 80 per cent), with the remainder extracted by one reviewer	☐ Yes ☐ No
7	**Did the review authors provide a list of excluded studies and justify the exclusions?** For Partial Yes: ☐ provided a list of all potentially relevant studies that were read in full text form but excluded from the review For 'full' Yes, must also have: ☐ justified the exclusion from the review of each potentially relevant study	☐ Yes ☐ Partial Yes ☐ No
8	**Did the review authors describe the included studies in adequate detail?** For Partial Yes (ALL the following): ☐ adequately described populations, including condition/clinical indication, age, gender where relevant ☐ adequately described interventions, including dosing regimen, cannabinoid profile, administration route ☐ described comparators ☐ described outcomes ☐ described research designs For 'full' Yes, should also have ALL the following: ☐ described study's setting ☐ timeframe for follow-up (Removed points on detailed description due to overlap with criteria above)	☐ Yes ☐ Partial Yes ☐ No
9*	**Did the review authors use a satisfactory technique for assessing the risk of bias (RoB) in individual studies that were included in the review?**	**Randomised controlled trials or clinical trials:** ☐ Yes ☐ Partial Yes ☐ No ☐ Includes only prospective cohort studies **Non-randomised prospective cohort studies** ☐ Yes ☐ Partial Yes ☐ No ☐ Includes only randomised controlled trials / clinical trials

	Authors must complete quality or risk of bias assessment on primary studies using the correct instrument for the included study design (risk of bias assessment for RCTs and purposely designed tool for prospective cohort studies) (Exclude if absent – fatal flaw) Did the authors assess the relevant points (see below)? **Randomised controlled trials or clinical trials:** For Partial Yes, must have assessed RoB from ☐ unconcealed allocation (randomization and blinding combined when allocating the intervention), AND ☐ lack of blinding assessors when assessing outcomes (unnecessary for objective outcomes such as all-cause mortality or admission to hospital) For 'full' Yes, must have assessed RoB from: ☐ allocation sequence that was not truly random (individual randomisation versus group randomization), AND ☐ selection of the reported result from among multiple measurements or analyses of a specified outcome, known as selective reporting (using only the outcomes or measurements that provide the researchers with their desired answer and ignoring other outcomes that may contradict the desired findings) **Non-randomised epidemiological studies:** For Partial Yes, must have assessed RoB: ☐ from confounding, AND ☐ from selection bias For Yes, must also have assessed RoB: ☐ methods used to ascertain exposures and outcomes, AND ☐ selection of the reported result from among multiple measurements or analyses of a specified outcome, known as selective reporting (using only the outcomes or measurements that provide the researchers with their desired answer and ignoring other outcomes that may contradict the desired findings)	
10	**Did the review authors report on the sources of funding for the studies included in the review?** For Yes, ☐ Must have reported on the sources of funding for individual studies included in the review (Note: Reporting that the reviewers looked for this information, but it was not reported by study authors also qualifies)	☐ Yes ☐ No

11*	**If meta-analysis was performed did the review authors use appropriate methods for statistical combination of results?**	☐ Yes ☐ No ☐ No meta-analysis

Randomised controlled trials or clinical trials:

For Yes:

- ☐ The authors justified combining the data in a meta-analysis
- ☐ AND they used an appropriate weighted technique to combine study results and adjusted for heterogeneity if present
- ☐ AND investigated the causes of any heterogeneity conducted

If heterogeneity present, appropriate investigations may include: completed feasibility analysis to decide what studies to include (PICO for clinical heterogeneity) and what type of meta-analysis to use (pairwise [2 arm trials and two competing interventions] versus network [three or more arm trials and more than two competing interventions]), used a random effects model if statistical heterogeneity is greater than an pre-agreed level (25%, 50% or 75%), estimated statistical heterogeneity (Q or I^2 test), determined influence of highly weighted studies (any one study influencing the outcome), high risk or unclear risk of bias studies (removed from analysis), or studies with different populations, comparators and intervention formats through sensitivity or sub-group analysis

Non-randomised epidemiological studies:

For Yes:

- ☐ The authors justified combining the data in a meta-analysis
- ☐ AND they used an appropriate weighted technique to combine study results, adjusting for heterogeneity if present
- ☐ AND they statistically combined effect estimates from prospective cohort studies that were adjusted for confounding, rather than combining raw data, or justified combining raw data when adjusted effect estimates were not available
- ☐ AND they reported separate summary estimates for RCTs and prospective cohort studies separately when both were included in the review

If heterogeneity present, appropriate investigations may include: completed feasibility analysis to decide what studies to include (PICO for clinical heterogeneity) and used pairwise meta-analysis, used confounding adjusted risk or odds ratios, used a random effects model if statistical heterogeneity is greater than a pre-agreed level (25%, 50% or 75%), estimated statistical heterogeneity (Q or I2 test), determined influence of highly weighted studies (any one study influencing the outcome),

	determined influence if low quality studies removed from analysis, determined influence if studies with low levels of control for confounding removed from analysis, and/or determined influence of studies with different populations, comparators and intervention formats. The influence should be determined through sensitivity or sub-group analysis	
12*	**If meta-analysis was performed, did the review authors assess the potential impact of RoB in individual studies on the results of the meta-analysis or other evidence synthesis?**	☐ Yes ☐ No ☐ No meta-analysis
	For Yes: ☐ included only low risk of bias RCTs (sensitivity analysis) Note: It is not good practice to combine RCT and prospective cohort studies; therefore, separate results should be provided and their similarities or differences discussed	
13*	**Did the review authors account for RoB in individual studies when interpreting/discussing the results of the review?**	☐ Yes ☐ No
	For Yes: ☐ included only low risk of bias RCTs in the review ☐ included only low risk of bias RCTs (in meta-analysis or a sensitivity analysis and discuss differences) ☐ OR, if RCTs with moderate or high RoB, or prospective cohort studies were included the review provided a discussion of the likely impact of RoB on the results and quality of evidence or limitations in conclusions or summary Note: Generally, non-randomised studies of interventions have more positive results that RCTs because of self-selection bias and lack of randomization and readers should be reminded of this. Confounding should be controlled for in the meta-analysis by using adjusted odds ratios. Loss to follow-up should be controlled for in the inclusion criteria. Loss to follow-up of over 20% introduces a serious bias to longitudinal studies. Risk of bias should also be discussed for narrative analysis Risk of bias should concentrate on the areas that were scored high risk or unclear risk of bias and its effect on the direction of the results.	
14*	**Did the review authors provide a satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review?**	☐ Yes ☐ No
	For Yes: ☐ There was no significant heterogeneity in the results ☐ OR if heterogeneity was present the authors performed an investigation of sources of any heterogeneity in the results (feasibility assessment, random effects model, sensitivity and sub-group analysis) AND discussed the impact of this	

	heterogeneity on the results of the review and the quality of evidence If narrative analysis completed, the effects of clinical heterogeneity on the results and quality of evidence must be discussed.	
15	**If they performed quantitative synthesis did the review authors carry out an adequate investigation of publication bias (small study bias) and discuss its likely impact on the results of the review?**	☐ Yes ☐ No ☐ No meta-analysis
	For Yes: ☐ performed graphical or statistical tests for publication bias and discussed the likelihood and magnitude of impact of publication bias Publication bias occurs when results of published studies are systematically different from unpublished or grey literature studies. Publication bias is trying to estimate the influence of unpublished studies on the results of the systematic review. Publication bias can be controlled for through a good comprehensive search strategy that includes unpublished studies, yet to be published studies, or studies published in grey literature and a wide selection of databases. Publication bias can be measured using a funnel plot and its p-value. A funnel plot is a scatter plot of estimates of the treatment effects of each study against the measure of its precision (1/Standard Error). In the absence of publication bias, plot will look like symmetric inverted funnel. A minimum of 10 studies are required to run the funnel plot analysis. The effect of publication bias should be considered in the GRADE quality of evidence.	
16	**Did the review authors report any potential sources of conflict of interest, including any funding they received for conducting the review?**	☐ Yes ☐ No
	For Yes: ☐ The authors reported no competing interests OR ☐ The authors described their funding sources and how they managed potential conflicts of interest In this case, the industry producing cannabis-based medicinal products is the main source of potential conflicts of interest.	

HRB-adapted AMSTAR 2 critical domains

We have selected eight rather than seven critical domains. Table 6 displays the critical domains selected by us and the original AMSTAR 2 authors, along with justifications for selection of critical domains.

Table 6 HRB-adapted AMSTAR 2 critical domains

Domain	Pollock *et al.* [22] AMSTAR critical domains	Shea *et al.* [19] AMSTAR 2 critical domains	HRB authors critical domains	Agreement or justification for selection of critical domains
Did the research questions and inclusion criteria for the review include the components of PICO (item 1)?	Yes	No	Yes	We regard this item as critical, as overviews indicate that clarity in the PICO leads to a better research objective, search strategy, clear inclusion and exclusion criteria, and a planned approach to analysis.
Protocol registered before commencement of the review (item 2)	No	Yes	Yes	We agree that this item is critical.
Adequacy of the literature search (item 4)	Yes	Yes	Yes	We agree that this item is critical. In addition, the inclusion of this item may help deal with excluding items 7 (excluded primary studies) and 15 (publication bias) as critical, and we agree that trials or cohort studies excluded at full text screening should be listed with a reason for exclusion.
Was there duplicate study selection and data extraction (item 5)?	Yes	No	No	We believe that this item is standard practice nowadays.
Justification for excluding individual studies (item 7)	Yes	Yes	No	We believe that this item overlaps with items 1 (PICO), 4 (search strategy), and 9 (risk of bias), and therefore does not need to be included as a critical domain.
Risk of bias and publication bias based on primary studies being	No	Yes	Yes	We agree that this item is critical.

included in the systematic review (item 9)				
If meta-analysis was performed, did the review authors assess the potential impact of risk of bias in individual studies on the results of the meta-analysis or other evidence synthesis (item 12)?	No	No	Yes	We believe that item 12 (risk of bias in doing meta-analysis) is critical. We think dealing with bias openly is key to avoiding misleading results.
Appropriateness of meta-analytical methods (item 11)	No	Yes	Yes	We agree that this item is critical.
Consideration of risk of bias when interpreting the results of the review (item 13)	No	Yes	Yes	We agree that this item is critical.
Did the review authors provide a satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review (item 14)?	No	No	Yes	We believe that clinical and statistical homogeneity or consistency (item 14) are key to a trustworthy analysis and must be dealt with the authors before and after meta-analysis.
Assessment of presence and likely impact of publication bias (item 15)	No	Yes	No	We regarded other items as more critical, and that this issue may be included under item 9.

Rating overall confidence in the results of individual systematic reviews

We will allocate each included systematic review a confidence rating using the schema from Shea *et al.* [19], shown in Table 7.

Table 7 Rating overall confidence in the results of individual systematic reviews

Score	Criteria
High	No or one non-critical weakness: the systematic review provides an accurate and comprehensive summary of the results of the available studies that address the question of interest
Moderate	More than one non-critical weakness*: the systematic review has more than one weakness but no critical flaws. It may provide an accurate summary of the results of the available studies that were included in the review
Low	One critical flaw with or without non-critical weaknesses: the review has a critical flaw and may not provide an accurate and comprehensive summary of the available studies that address the question of interest
Critically low	More than one critical flaw with or without non-critical weaknesses: the review has more than one critical flaw and should not be relied on to provide an accurate and comprehensive summary of the available studies
*Downgrade	*Multiple non-critical weaknesses may diminish confidence in the review, and it may be appropriate to move the overall appraisal down from moderate to low confidence.

Source: Shea et al. 2017 [19]