NNT : 0000AIXM0000

THÈSE DE DOCTORAT

Soutenue à Aix-Marseille Université
le 1er Mars 2024 par

Beata SMELA

Méthodologie des revues de la littérature : état de l'art, place dans les politiques de santé et perspectives à l'ère des nouvelles technologies

Discipline
Biologie santé

Spécialité
Recherche clinique et Santé Publique

École doctorale
ED 62 - SCIENCES DE LA VIE ET DE LA SANTE

Laboratoire/Partenaires de recherche
EA 3279
CEReSS - Centre d'Etudes et de Recherche sur les Services de Santé et la Qualité de Vie

Composition du jury

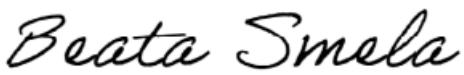
Stéphanie GENTILE Université d'Aix Marseille	Présidente du jury
Mondher TOUMI Université d'Aix Marseille	Directeur de thèse
Bastien BOUSSAT Université Grenoble Alpes	Rapporteur
Hans-Martin SPÄTH Université Claude Bernard Lyon 1	Rapporteur

Affidavit

Je soussigné, Beata Smela, déclare par la présente que le travail présenté dans ce manuscrit est mon propre travail, réalisé sous la direction scientifique de Mondher Toumi et Clément François, dans le respect des principes d'honnêteté, d'intégrité et de responsabilité inhérents à la mission de recherche. Les travaux de recherche et la rédaction de ce manuscrit ont été réalisés dans le respect à la fois de la charte nationale de déontologie des métiers de la recherche et de la charte d'Aix-Marseille Université relative à la lutte contre le plagiat.

Ce travail n'a pas été précédemment soumis en France ou à l'étranger dans une version identique ou similaire à un organisme examinateur.

Fait à Cracovie, Pologne, le 22 décembre 2023

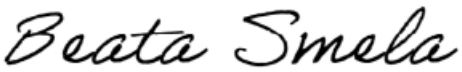


Affidavit

I, undersigned, Beata Smela, hereby declare that the work presented in this manuscript is my own work, carried out under the scientific direction of Mondher Toumi and Clément François, in accordance with the principles of honesty, integrity and responsibility inherent to the research mission. The research work and the writing of this manuscript have been carried out in compliance with both the French national charter for Research Integrity and the Aix-Marseille University charter on the fight against plagiarism.

This work has not been submitted previously either in this country or in another country in the same or in a similar version to any other examination body.

Place Kraków, Poland, date 22 décembre 2023



List of the publications and posters

List of publications produced as part of the PhD thesis project or related to the PhD thesis in general:

01. **Smela B,** Toumi M, Świerk K, Francois C, Biernikiewicz M, Clay E, Boyer L. **Rapid literature review: definition and methodology.** J Mark Access Health Policy. 2023 Jul 28;11(1):2241234. doi: 10.1080/20016689.2023.2241234. PMID: 37533549; PMCID: PMC10392303.

02. **Smela B,** Toumi M, Świerk K, Gawlik K, Clay E, Boyer L. **Systematic literature reviews over the years.** J Mark Access Health Policy. 2023 Aug 21;11(1):2244305. doi: 10.1080/20016689.2023.2244305. PMID: 37614556; PMCID: PMC10443963.

03. **Smela B**, Toumi M, Świerk K, Tusińska A, Klimonczyk K, Clay E, Boyer L. **Assessing the validity of a rapid review against a systematic literature review: A comparison of systematic literature reviews done by Cochrane with rapid reviews and the impact on meta-analyses results.** Journal of Epidemiology and Population Health *(submitted)*

04. Wojciechowski P, Wilson K, Nazir J, Pustułka I, Tytuła A, **Smela B**, Pochopien M, Vredenburg M, McCrae KR, Jurczak W. Efficacy and Safety of Avatrombopag in Patients with Chronic Immune Thrombocytopenia: **A Systematic Literature Review** and Network Meta-Analysis. Adv Ther. 2021 Jun;38(6):3113-3128. doi: 10.1007/s12325-021-01752-4. Epub 2021 May 1. PMID: 33934279; PMCID: PMC8189936.

05. Kemp R, Pustulka I, Boerner G, **Smela B,** Hofstetter E, Sabeva Y, François C. Relationship between FEV1 decline and mortality in patients with bronchiolitis obliterans syndrome-**a systematic literature review**. Respir Med. 2021 Nov;188:106608. doi: 10.1016/j.rmed.2021.106608. Epub 2021 Sep 10. PMID: 34517199.

06. Beck E, Biundo E, Devlin N, Doherty TM, Garcia-Ruiz AJ, Postma M, Sheikh S, **Smela B**, Toumi M, Wasem J, Nolan T, Salisbury D. Capturing the value of vaccination within health technology assessment and health economics: **Literature review** and novel conceptual framework. Vaccine. 2022 Jun 26;40(30):4008-4016. doi: 10.1016/j.vaccine.2022.04.050. Epub 2022 May 23. PMID: 35618559.

07. Folkerts K, Millier A, **Smela B**, Olewinska E, Schmedt N, Mernagh P, Kovesdy CP. **Real-world evidence** for steroidal mineralocorticoid receptor antagonists in patients with chronic kidney disease. J Nephrol. 2023 May;36(4):1135-1167. doi: 10.1007/s40620-022-01492-w. Epub 2022 Nov 23. PMID: 36422853; PMCID: PMC10227157.

An 8th manuscript, presenting the process of two real-world evidence (RWE) filters development and testing together with a comparison to the Cochrane's group results, is at the final stage of development and will be submitted soon.

Participation in conferences and summer schools during the thesis period:

01. Systematic literature reviews over the years. Smela B, Toumi M, Świerk K, Gawlik K, Clay E, Boyer L. IMAS 2023, Miami, US
02. Differences in the requirements for clinical evidence for HTA submissions in different countries. Smela B, Toumi M, Świerk K, Millier A, Francois C, Panek M, Boyer L. IMAS 2023, Miami, US
03. Different approaches to economic evaluations for HTA submissions. Smela B, Toumi M, Świerk K, Millier A, Boyer L. IMAS 2023, Miami, US
04. Differences in the requirements for economic evaluations for HTA submissions among different countries. Smela B, Toumi M, Świerk K, Millier A, Francois C, Panek M, Boyer L IMAS 2023, Miami, US
05. The Relevance of Literature Reviews in the Process of Decision-making by French Health Technology Assessment Agency. Smela B, Toumi M, Belgaied W, Świerk K, Millier A, Boyer L
06. Rapid literature review: definition and methodology. Smela B, Toumi M, Świerk K, Francois C, Biernikiewicz M, Clay E, Boyer L. IMAS 2023, Miami, US
07. Cochrane Systematic Reviews versus Rapid Reviews: the impact of different methodologies on the reviews' conclusions. Smela B, Toumi M, Świerk K, Tusińska A, Clay E, Boyer L. IMAS 2023, Miami, US
08. Cochrane Systematic Review versus Rapid Review: the impact of database coverage and single studies selection on the review's conclusions. Smela B, Toumi M, Świerk K, Tusińska A, Clay E, Boyer L. IMAS 2023, Miami, US
09. Wojciechowski P, Wilson K, Pochopień M, Pustulka I, Smela B, Tytuła A, Vredenburg M, Nazir J. Comparative Efficacy and Safety of Avatrombopag Versus Lusutrombopag in Patients with Chronic LIVER Disease and Severe Thrombocytopenia Undergoing Invasive Procedures: A Systematic Literature Review and Network Meta-Analysis. 2021-05, ISPOR 2021, Montreal, Canada. Value in Health, Volume 24, Issue 5, S1 (May 2021)

10. Use of Artificial Intelligence with Distillersr Software in Selected Systematic Literature Reviews. Smela B, Pustulka I, O'Blenis P, Millier A. 2020-09, ISPOR Asia Pacific 2020, Seoul, South Korea. Value in Health Regional, Volume 22S (September 2020)

11. Use of Artificial Intelligence with Distillersr Software for a Systematic Literature Review of the Economic Burden Related with Selected Headache Disorders. Smela B, Pustulka I, Millier A, O'Blenis P. 2020-11, ISPOR Europe 2020, Milan, Italy. Value in Health, Volume 23, Issue S2 (December 2020)

12. Use of Artificial Intelligence with Distillersr Software for a Systematic Literature Review of the Humanistic Burden Related with Selected Headache Disorders. Smela B, Pustulka I, Millier A, O'Blenis P. 2020-11, ISPOR Europe 2020, Milan, Italy. Value in Health, Volume 23, Issue S2 (December 2020)

13. Differences in Evidence Reviews Requirements for Health Technology Assessment (HTA) Document Development between Selected Countries. Olewinska E, Okolus D, Nomoto M, Millier A, Smela B, Ueyama M. 2020-09, ISPOR Asia Pacific 2020, Seoul, South Korea. Value in Health Regional, Volume 22S (September 2020)

14. Use of artificial intelligence with distillersr software as a reviewer for a systematic literature review of randomized controlled trials. Smela-Lipinska B, Taieb V, Szawara P, Tetzlaff J, O'Blenis P, François C. 2019-11, ISPOR Europe 2019, Copenhagen, Denmark

15. Use of artificial intelligence with distillersr software as a reviewer for a systematic literature review of cost-effectiveness models. Smela-Lipinska B, Taieb V, Myjak I, Tetzlaff J, O'Blenis P, François C. 2019-11, ISPOR Europe 2019, Copenhagen, Denmark

16. Use of artificial intelligence with distillersr software for a systematic literature review of utilities in infectious disease. Taieb V, Smela-Lipińska B, O'blenis P, François C. 2018-11, ISPOR Europe 2018, Barcelona, Spain. Value in Health, Vol. 21, S3 (October 2018)

17. Health-related quality of life and patient reported outcomes measures in idiopathic multicentric castleman's disease- a systematic review. Smela-Lipinska B, Marre C, Myjak I, Szawara P, Abdennadher E, Marsh S, Okhuoya P, Zinzani PL, Toumi M. 2019-11, ISPOR Europe 2019, Copenhagen, Denmark

18. A systematic review of the clinical burden of idiopathic multicentric castleman's disease. Smela-Lipinska B, Marre C, Szawara P, Abdennadher E, Myjak I, Marsh S, Okhuoya P, Zinzani PL, Toumi M. 2019-11, ISPOR Europe 2019, Copenhagen, Denmark

19. The completeness of clinicaltrials.gov. - what is the rate of published clinical trials registered on the website? Myjak I, Bogoeva N, Smela-Lipinska B, François C, Taieb V. 2019-11, ISPOR Europe 2019, Copenhagen, Denmark

20. Comparison of sickness benefit and sick leave duration in oncology across countries. Contente M, Singh P, Smela-Lipińska B, van Dijk B, François C. 2018-11, ISPOR Europe 2018, Barcelona, Spain. Value in Health, Vol. 21, S3 (October 2018)

21. Burden of illness associated with influenza virus infection in japan- a literature review. Fukushima A, Smela-Lipińska B, Tone K, Ikeoka H. 2018-09, ISPOR Asia Pacific 2018, Tokyo, Japan. Value in Health, Vol. 21, S2 (September 2018)

22. Patient-reported outcomes in randomized controlled trials for overactive bladder - a systematic literature review. Germain N, Aballea S, Smela-Lipinska B, Pisarczyk K, Keskes M, Toumi M. 2018-09, ISPOR Asia Pacific 2018, Tokyo, Japan. Value in Health, Vol. 21, S2 (September 2018)

23. Do utility values used in health technology assessment mainly come from widely cited references? Smela-Lipińska B, Szawara P, Taieb V, François C. 2018-11, ISPOR Europe 2018, Barcelona, Spain. Value in Health, Vol. 21, S3 (October 2018)

Résumé

Cette thèse visait à examiner la place des revues systématiques de la littérature (SLR pour Systematic literature review) et des revues rapides (RR) dans les politiques de santé et à évaluer les perspectives à l'ère des nouvelles technologies.

La première partie du travail consistait à évaluer les exigences des évaluations des technologies de santé (ETS) pour les SLR, les tendances des SLRs publiées et le statut des connaissances sur les RR.

Les recommandations pour les ETS concernant les SLR sont hétérogènes et manquent de rigueur; pour les preuves économiques, une approche systématique n'est pas nécessaire.

La volume des publications scientifiques explose, nécessitant des méthodologies efficaces pour les SLRs.

Les RR ont été proposées comme solution. Les SLR et les RR sont souvent utilisées de manière interchangeable ; il est nécessaire d'établir une définition unifiée et de spécifier l'indication pour chaque type de revue.

L'objectif de la deuxième partie de mon travail était d'évaluer trois méthodologies de RR, le logiciel qui soutient le processus des LRs et de créer et tester des filtres sur les designs d'étude.

Les méthodologies testées pour les RR ont des résultats d'efficacité de 84%, 89% et 100% ; ce dernier résultat a été obtenu en limitant l'intervention humaine et en effectuant une recherche dans une seule base de données.

Au total, 6 des 56 essais n'ont pas été récupérés dans les 3 RR : 1) cinq étaient de mauvaise qualité, un avec un risque de biais non clair, 2) quatre études n'étaient pas disponibles dans MEDLINE et Embase et deux étaient mal indexés, 3) Cochrane a inclus des études manquées dans 28 méta-analyses ; cependant, leur absence dans les RR n'a affecté que la conclusion de 4 méta-analyses.

Créer et tester des filtres pour des études observationnelles a montré que ce type d'études est indexé avec des termes non pertinents ou recherché avec des termes trop larges.

En utilisant notre filtre SLR, 84 à 100% des études identifiées par le groupe Cochrane ont été capturées.

L'indexation améliorée et un bon filtrage devraient améliorer les performances des RR. Les LR soutenus par un logiciel IA sont efficaces pour le criblage. Des développements prometteurs de la technologie d'apprentissage automatique soutiendront la recherche documentaire et l'extraction de données.

Abstract

This thesis aimed to review the place of SLRs and RRs in healthcare policy and assess perspectives in the era of new technologies.

The first part of the work was done to assess HTA requirements for LRs, published SLRs' trends, and the knowledge status on RRs.
HTAs' recommendations regarding LRs are heterogeneous and lack stringency; for the economic evidence, a systematic approach is not required. The scientific publications volume is exploding requiring efficient methodologies for LRs. RRs have been proposed as the solution. SLRs and RRs both are often used interchangeably; there is a need for a unified definition and specifying the indication for each reviews' type.

The objective of the second part of this work was to assess three RRs' methodologies, software supporting the LRs process, and create and test study design filters.
The tested methodologies for RRs resulted in 84, 89, and 100% of efficacy; the last was obtained by limiting human intervention and one database searching. In total, 6 out of 56 trials were not retrieved in all 3 RRs: 1) five were of poor quality, one with a not clear risk of bias, 2) four studies were not available in MEDLINE and Embase, and two were inappropriately indexed, 3) Cochrane included missed studies in 28 meta-analyses; however, their lack in RRs did impact significantly only conclusion of 4 meta-analyses.
Creating and testing filters for observational studies showed that this type of studies is being indexed with irrelevant terms or searched with too broad terms. Using our SLR filter, 84 to 100% of studies identified by the Cochrane group were captured. Improved indexing and good filtering should improve the performance of RRs.
LRs' supported by AI software are effective for screening. Promising developments in ML technology will support literature searching and data extraction.

Keywords: systematic literature reviews, rapid reviews, health technology assessment, critical appraisal, Ovid indexing, searching filters, AI tools in the future

Table of contents

Introduction

Systematic literature reviews and evidence synthesis

The need to introduce evidence-based medicine was first expressed in the late 20th century when an issue of subjective and inefficient approaches in research arose. Much data of inconsistent quality was available, making healthcare decision-making difficult [1,2]. This led to an increased emphasis on randomized controlled trials (RCTs) as the most reliable study design in testing treatment efficacy, followed by observational studies and expert opinions[3,4]. Then, the hierarchy of evidence has been changed to include systematic literature reviews (SLRs) as the most valid source of knowledge[5]. The process of conducting an SLR includes merging information from numerous sources published in the same field to comprehend the findings, which allows responsible parties to make informed decisions. Cochrane Collaboration was started to inform clinical practice by producing trusted evidence[6]. The organization provided a manual for conducting an SLR with the lowest possible bias (Cochrane Handbook for Systematic Reviews of Intervention)[7]. The Cochrane Handbook contains specific instructions on performing all elements of SLR, which are clearly stated objectives, pre-defined eligibility criteria for studies, reproducible methodology, a systematic search, an assessment and a comprehensive presentation of the characteristics, findings, and the quality of the findings from the included studies. The proposed guidelines are highly complex, allowing for a comprehensive and reliable review of the evidence[7].

SLRs of high quality are of great importance since the amount of published evidence is exploding, massive and constantly increasing amounts of data need to be analyzed, and, at the same time, the pharmaceutical industry and its' value are continuously growing and progressing. The estimated value of the total global pharmaceutical market for 2022 was 1.48 trillion US dollars. The total worldwide expenditure on research and development in the pharmaceutical industry reached approximately 238 billion US dollars in 2021[8]. The increase in the market for pharmaceuticals results in higher demand for high-quality SLRs to be conducted, as they can support each step of the drug development and then product management, including health technology assessment (HTA) processes leading to a potential refund of the treatment.

Health Technology Assessment

According to the European Network for Health Technology Assessment (EUnetHTA): 'Health technology assessment is a multidisciplinary process that summarizes information about the medical, social, economic and ethical issues related to the use of health technology in a systematic, transparent, unbiased, robust manner. It aims to inform the formulation of safe, effective health policies that are patient-focused and seek to achieve the best value'. Currently, the Health Technology Assessment (HTA) procedure is

considered a standard policy tool for informing decision-makers who manage the entry of new medical technologies into health systems through pricing and reimbursement[9, 10]. Following marketing authorization, authorities in different countries undertake a review of the clinical and economic evidence. They have different requirements for the submitted documentation and the processes of data obtention. The most demanding guidelines for clinical evidence are the ones presented by The National Institute for Health and Care Excellence in the UK (NICE)[11, 12], Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen in Germany (IQWIG)[13], Institute for Clinical and Economic Review in the US (ICER)[14] and the Pharmaceutical Benefits Advisory Committee in Australia (PBAC)[15]. The agencies present specific recommendations about defining inclusion criteria (PICOS – population, intervention, comparators, and study design), preparing search strategies, selection process, managing discrepancies, and critical appraisal. All of the HTA agencies recommend conducting a review of clinical evidence in a systematic manner. The four most demanding agencies underline the necessity of defining PICOS criteria and searches running in three major databases: MEDLINE, Embase, and CENTRAL. If appropriate, discipline-specific databases should be searched too. Additional scanning of reference lists of relevant studies, hand-searching key journals, contacting experts, citation searching, and searching of grey literature and clinical trial registries are advised by the agencies to avoid selection bias. Specific guidelines for developing search strategies are provided; agencies recommend using medical subject headings combined with free text[13-15] , and some of them underline that the search strategy should be reproducible[11,12,15]. Some agencies allow the restriction to a specific language[13, 14]. For the selection of studies, the UK and US agencies mention the necessity of double-screening by two independent reviewers[11,12,14]. Agencies from Australia, Germany, the UK and the US require a representation of a flow of studies in the review and proper documentation of the review process. Quality assessment of included studies is mandatory for all four agencies[11,12, 14 ,13,15].

As for economic evaluations reviews, only the French agency Haute Autorité de Santé (HAS) requires an SLR[16]. NICE states that reviews may not be exhaustive if additional studies identified would merely provide further support that is consistent with the already-identified evidence[11,12,] . IQWiG suggests a focused search of previous evaluations may be performed, at least in MEDLINE[13]. Australian HTA agency requires that the literature review for economic evaluations needs to be conducted, but does not specify whether it should be systematic[15]. To obtain data on utilities, two European agencies, NICE and HAS, require that literature is reviewed systematically; the rationale for terms used in the search strategy and any inclusion and exclusion criteria used should be provided, and the search strategies should be listed in the submission. Exact databases are not specified, but the search should be exhaustive[11,12,16]. PBAC recommends that when utility weights cannot be directly estimated from data collected in the clinical studies, the estimates may be derived from literature; the details of search strategies and inclusion and exclusion criteria used to identify relevant utility studies should be presented.

However, it is not specified if the search needs to be systematic[15]. NICE and HAS mention the necessity for critical appraisal of identified cost-effectiveness studies[11,12, 16].

Taking into consideration all the requirements mentioned, a significant workload is anticipated to conduct all the literature reviews needed. Involving multiple reviewers for a period of even two years is necessary to develop documents for HTA submission[17].

Rapid reviews

The rigorous and transparent approach used in SLRs make them the preferred ones[18]. A disadvantage of SLRs is that they are time-consuming and demand a lot of human resources[19, 20]. Considering the significant increase in newly published data and the unchanged timelines for HTA submissions, or even manuscript submissions, there is an escalating demand for rapid information analysis. In this situation, rapid reviews (RRs) can be an alternative under some circumstances. Running RRs is a quite new area, and still, no agreement on optimal methods can be found[21]. There are several guidelines on the methodology of RRs[21-27]; however, only recently, one publication from 2021 attempted to construct a unified definition[24]. The definition of RR proposed by the Cochrane Rapid Review Methods Group is as follows: 'A rapid review is a form of knowledge synthesis that accelerates the process of conducting a traditional systematic review through streamlining or omitting various methods to produce evidence for stakeholders in a resource-efficient manner'[24]. Generally, by RRs, researchers understand evidence synthesis during which some of the components of the systematic approach are being used to facilitate answering a focused research question; however, scope restrictions, narrower search strategies, simplified approach to studies selection, data extraction, and potential support of AI tools help to make the project manageable in a shorter time and to get the key conclusions faster [19, 25, 28]. It has been suggested that most RRs are conducted within 12 weeks; some of the resources suggest timeframes between a few weeks to no more than six months[21, 25].

Objectives of the thesis

This thesis aimed to support placing SLRs and RRs in the area of healthcare policy and assess their perspectives in the era of new technologies. The important goal of this work is to facilitate an understanding of how crucial it should be for reviewers working on SLRs and RRs to make sure the chosen methodology is relevant, as well as for policymakers to critically evaluate literature reviews' (LRs') quality and consider potential biases when analyzing the results and making recommendations.

1. History of systematic literature reviews

1.1. Key findings

The questions being answered by this section:

- **What are the trends in publication of SLRs, particularly focusing on RCTs and epidemiological studies?**
- **What was the impact of the COVID-19 pandemic on the process of SLRs publication?**
- **What are the other factors impacting the process?**
- **What is the future of SLRs?**

- An RR was run and 6,892 titles and abstracts of SLRs were analyzed in-depth when it comes to disease of interest
- Publication rate of RCT SLRs is growing at an exponential rate
- The increase observed during the last 3 years is partially attributed to the emergence of the COVID-19 pandemic
- Based on the presented analysis of the published SLRs, the most frequently evaluated diseases are aligned with leading causes of death and disability worldwide indicated in the reports published by WHO and IHME[29, 30]; the most frequently analysed areas among the identified articles were diseases of the circulatory system (I00-I99, 10.9% of included articles), such as heart failure or stroke, closely followed by endocrine, nutritional, and metabolic diseases (E00-E90, 10.7%), mainly diabetes
- 7.8% of SLRs were assessing the impact of various supplementations or diets on health; this observed increase in such SLRs is anticipated to persist, driven by factors including the expanding geriatric population, emphasis on preventive healthcare, growing societal focus on well-being, and a shift from conventional pharmaceuticals to supplements and dietary interventions
- The majority of SLRs incorporated data from a moderate number of RCTs, ranging between 11 and 50 trials, accounting for 46.4% of the sample. Approximately one-third of the analyzed reviews included a smaller number of studies, between 1 and 10 trials. Notably, only 87 out of 6,892 identified RCT

SLRs analyzed more than 200 trials. This distribution remained consistent across different disease areas and over time

- The analysis showed the increasing trend also for SLRs of epidemiological studies; it may be explained based on the recently observed intensified trend of RWE analysis, increase in digitisation of health records facilitating data analysis, and escalating focus on the importance of patient-reported outcomes

➢ **As the volume of RCTs and epidemiological studies continues to expand, there will be increasing need for unbiased evidence summary provided in a timely manner way, supporting healthcare policy decision making**

➢ **In light of these considerations, novel solutions will be imperative, necessitating the development of RRs with refined and tailored methodologies, the integration of AI tools to enhance the efficiency of LRs, and the optimization of processes involved in LR execution, including using well-founded, specific and effective study design filters**

1.2. Introduction

By the late 20th century, recognizing the latest evidence to support clinical practice took a lot of work. The reason for this was an overwhelming amount of information available in the literature and the inconsistent quality of reviewing the knowledge. Experts were accused of a subjective, inefficient manner in preparing the synthesis[1, 2].

In 1972 Archie Cochrane set out that medical practice should be evidence-based[3]. He underlined the importance of RCTs in research. He also introduced the terms: effectiveness, efficiency, and equality. Another prominent scientist in the area of evidence synthesis was Gene Glass, who synthesized knowledge in the field of psychotherapy and class size and proposed the term' meta-analysis' for the first time[31]. In 1989, Sackett proposed a grading of recommendations – he set out that RCTs are the most reliable source, superior to cohort, case-control studies, case series, and expert opinion[4]. Then, Greenhalgh updated the hierarchy by adding SLRs or meta-analyses of RCTs to the top of the list[5]. CONSORT (Consolidated Standards of Reporting Trials) was established in 1993 to guide the authors on essential information to be included for readers to assess the reliability of presented results[32]. In 1993 Cochrane Collaboration was founded as well[6]. The non-profit organization's purpose is to inform clinical practice by conducting SLRs and providing a manual for systematic reviewing (Cochrane Handbook for Systematic Reviews of Intervention)[7]. In 1996 QUORUM (Quality of Reporting of Meta-Analyses) checklist was established to ensure transparency and reliability of the reviews, and in 2009 the guidelines were updated and renamed as Preferred Reporting Items for

Systematic Reviews and Meta-Analyses (PRISMA)[33]. Its key feature is the PRISMA flowchart, which shows the flow of papers in the screening process. Nowadays, a methodical approach is available to support healthcare decision-making, and HTA agencies worldwide have their own specific requirements to ensure the quality of evidence.

SLRs and meta-analyses are placed at the top of the study hierarchy of evidence. The main objective of this work was to evaluate the trends in SLRs of RCTs throughout the years, including number of published records, their distribution among different disease areas and based on the number of included trials. Additionally, trends were compared to the ones observed for SLRs on epidemiological studies.

1.3. Methods

The search in PubMed was performed in May 2023, using a highly focused search strategy: "randomised controlled" OR "randomised clinical" OR "randomized controlled" OR "randomized clinical" OR RCT*. Later, an appropriate filter was applied in order to retrieve only studies with an SLR design, and then the ones focusing only on RCTs. Numbers of retrieved studies stratified by publication years were exported into an Excel file. It was then analyzed to get overall trends in the amount of RCTs SLRs published per year. Additionally, number of RCTs published annually was searched by using the following search queries in PubMed: "randomised controlled" OR "randomised clinical" OR "randomized controlled" OR "randomized clinical" OR RCT*, and applying filter for RCTs. Results were exported into an Excel file and analyzed.

To run a detailed analysis of the trends in RCT SLRs, an RR was conducted in MEDLINE database; a very restricted approach was used, with search queries being searched only in titles of records. The objective was to get a reasonable sample of hits to be analyzed, and the approach was backed up by the fact that SLRs on RCTs of good quality, when published, usually contain both terms ''systematic literature review'' and ''randomized controlled trial'' in title. MEDLINE database was used instead of PubMed due to its capacity for conducting more refined searches and facilitating the restriction of searches, such as limiting them to specific fields like the title field. The results from MEDLINE were compared to the ones from PubMed, to analyse consistency between them (check if the same trend in the number of published SLRs would be observed). MEDLINE database was searched via Ovid in May 2023 using the following search terms combination: (systematic review* or systematic literature review*).ti AND randomi?ed controlled trial*.ti. The search results were then imported to EndNote 20 (Clarivate) program and analysed using the Eppi-Reviewer Web software[34, 35]. A single screening of titles and abstracts was performed. Additionally, based on information provided in titles and abstracts, each paper was coded according to a specific ICD-10 code[36] (depending on the analysed disease area or procedure, Table 1) or, if no ICD-10 code was applicable (for example, the SLR analysed healthy subjects), to "Treatments" code, consisting of "Pain treatment", "Anaesthesia", "Supplements/diet" and "Other treatments/interventions" subcategories. When appropriate, two or more codes were selected. Data regarding the number of RCTs included in each analysed SLR was also retrieved and divided into six categories: 1-10, 11-50, 51-100, 101-200, >200, and not reported (in abstract/title, NR).

All SLRs analysing RCTs were included, without any restrictions on population, interventions, or outcomes. Protocols, commentaries, or errata were excluded. No restrictions on date or language were applied.

The last step was running a focused search for SLRs of epidemiological studies. The following search queries were used in PubMed: "epidemiologic stud*" OR "epidemiology" OR "epidemiologic" OR "epidemiological" OR "epidemiol*", and the filter for SLRs was applied. Again, results were exported into an Excel file.

Table 1 Disease area according to ICD-10 classification[36]

Code	Title	Code	Title
A00–B99	Certain infectious and parasitic diseases	**L00–L99**	Diseases of the skin and subcutaneous tissue
C00–D48	Neoplasms	**M00–M99**	Diseases of the musculoskeletal system and connective tissue
D50–D89	Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism	**N00–N99**	Diseases of the genitourinary system
E00–E90	Endocrine, nutritional and metabolic diseases	**O00–O99**	Pregnancy, childbirth and the puerperium
F00–F99	Mental and behavioral disorders	**P00–P96**	Certain conditions originating in the perinatal period
G00–G99	Diseases of the nervous system	**Q00–Q99**	Congenital malformations, deformations and chromosomal abnormalities
H00–H59	Diseases of the eye and adnexa	**R00–R99**	Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
H60–H95	Diseases of the ear and mastoid process	**S00–T98**	Injury, poisoning and certain other consequences of external causes
I00–I99	Diseases of the circulatory system	**V01–Y98**	External causes of morbidity and mortality
J00–J99	Diseases of the respiratory system	**Z00–Z99**	Factors influencing health status and contact with health services
K00–K93	Diseases of the digestive system	**U00–U99**	Codes for special purposes

1.4. Results

1.4.1. Database analysis

The PubMed search for the SLR of RCTs yielded 86,765 results (Figure 1). The first identified article was published in 1990 and compared the effects of corticosteroid to no corticosteroid treatment before preterm delivery based on data from 12 RCTs[37]. Later, another 10 articles were published in 1994, and since that time, new articles were being issued every year. The number of SLRs published has been increasing within time: in 1999

100 articles were published, and 1000 in 2005. Recently, a couple of thousands of new RCT SLRs were being published annually; almost 40% of all identified records were published since the year 2020. An exponential growth of published RCT SLRs is being observed. The highest number was observed in the year 2022, with more than 10,000 publications issued. As presented on Figure 2, the number of published RCTs was also growing, reaching a maximum in year 2014, and has been staying on a similar level by nowadays (between 27,000 – 30,000 records published per year).

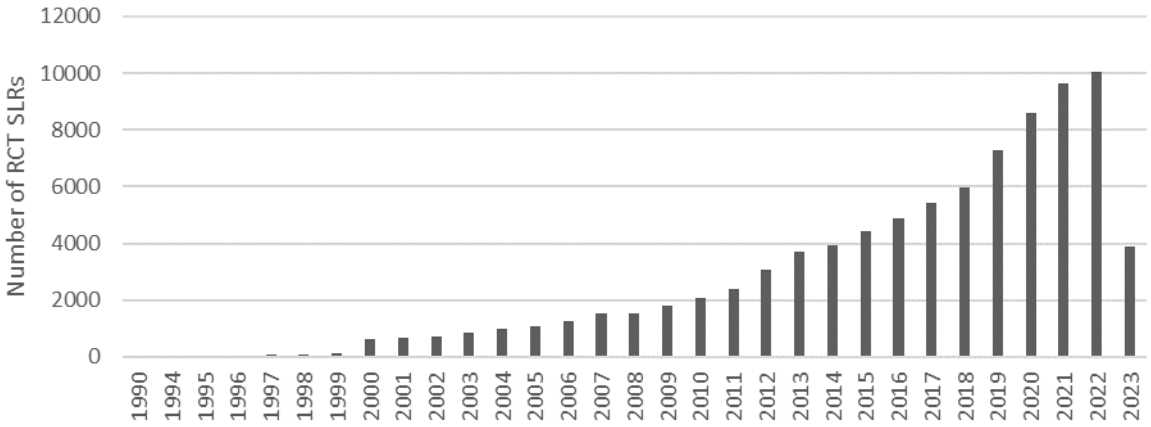


Figure 1 The number of RCT SLRs published over the years. Source: PubMed (search run in May 2023)

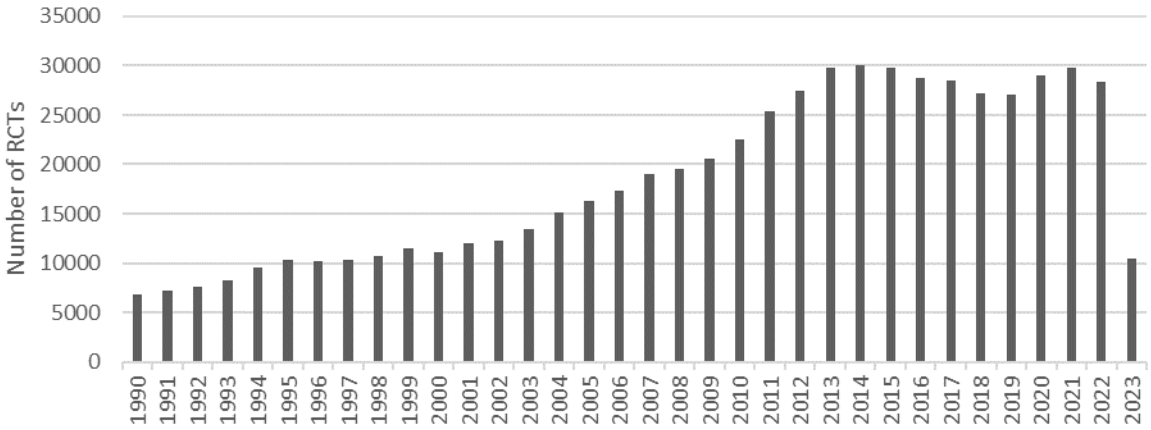


Figure 2 The number of RCT published over the years. Source: PubMed (search run in May 2023)

1.4.1. Rapid review

The search run in MEDLINE yielded 7,534 records. After deduplication, 7,465 titles and abstracts were analysed, and 6,892 were included to be further analyzed (Figure 3). The oldest retrieved publications were published in 1994. The increase in the number of published RCT SLRs was consistent with the one observed in analysis of data from PubMed; over the past few years, there has been a noticeable escalation observed.

Therefore, it was assumed that the identified records provided a representative sample for further analysis.

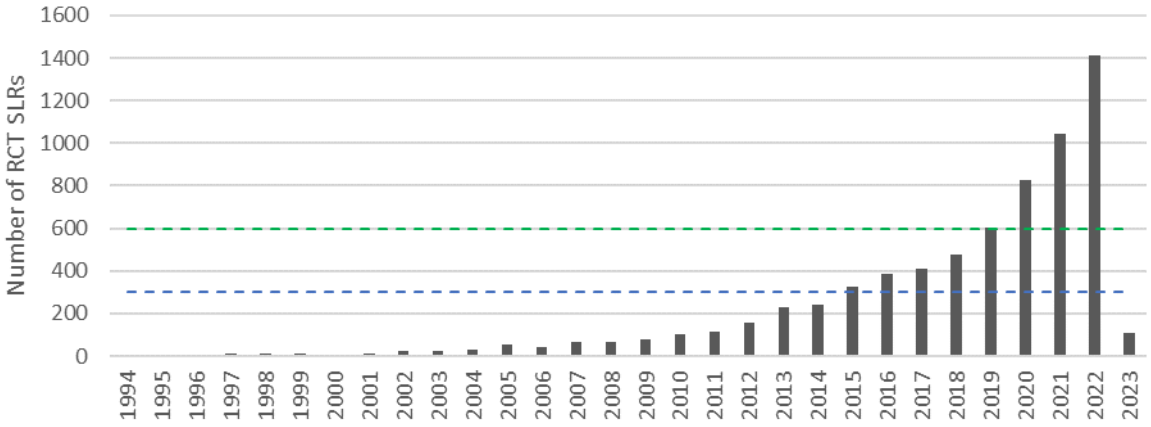


Figure 3 Distribution of RCT SLRs over the years. Source: MEDLINE (search run in May 2023)

The distribution of all identified RCT SLRs by the disease area, and the number of included RCTs, is presented in Figure 4. The most frequently analysed among the identified articles were diseases of the circulatory system (I00-I99, 10.9% of included articles), such as heart failure or stroke, closely followed by endocrine, nutritional, and metabolic diseases (E00-E90, 10.7%), mainly diabetes. Additionally, 7.8% of SLRs were assessing the impact of various supplementations or diets on health. The relatively low number of SLRs was assessing the following disease areas: congenital malformations, deformations and chromosomal abnormalities (Q00-Q99, n=10), diseases of the ear and mastoid process (H60-H95, n=21) and external causes of morbidity and mortality (V01-Y98, n=26).

The majority of SLRs included data from a moderate number of RCTs (between 11 and 50, 46.4%); and a one-third of analysed reviews included a lower number of studies (between 1 and 10). Among the rest of SLRs, 5.2% included between 51 and 100 trials, 1.9% between 101 and 200; and only 87 out of 6,892 identified RCT SLRs analysed more than 200 trials. For 6.9% of articles, abstracts did not report the number of included trials. In general, this proportion was consistent throughout the years and among various disease areas.

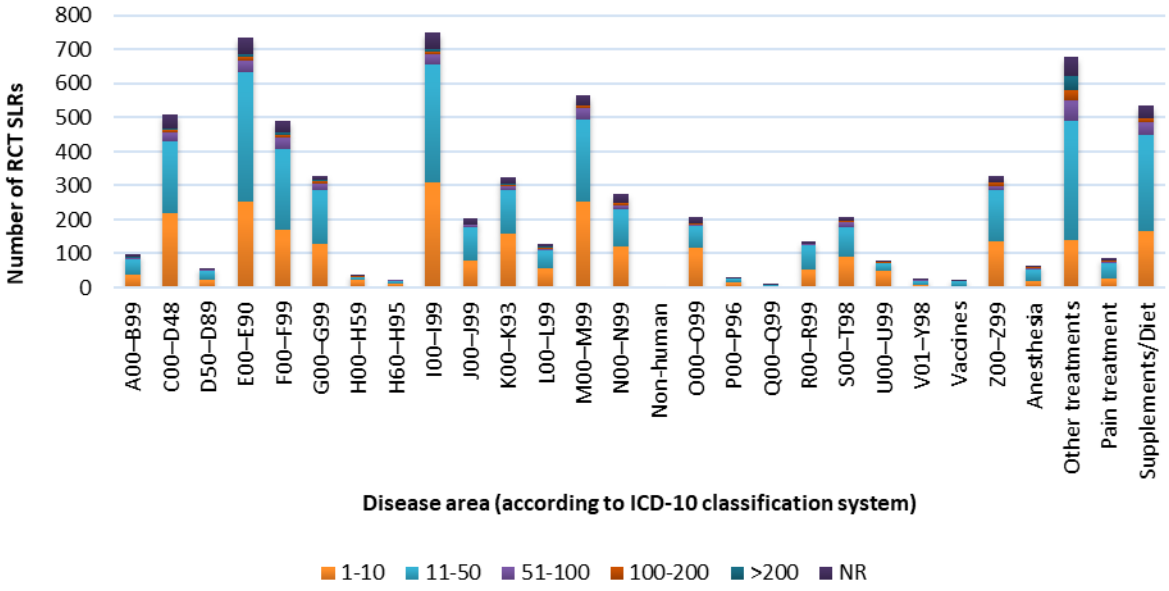


Figure 4 Number of RCT SLRs stratified by disease area and by the number of included RCTs

The distribution of studies by disease area across three distinct periods is illustrated in Figure 5: from 1994 to 2015 (A), from 2016 to 2019 (B) and from 2020 onward (C).

A total of 1,624 analyzed RCT SLRs were published between 1994 and 2015. The main areas of focus were diseases affecting the circulatory system (I00-I99; 10.8%) and the musculoskeletal system and connective tissue (M00-M99, 8.5%). Neoplasms (C00-D48, 7.5%) and mental and behavioural disorders (F00-F99, 7.4%) were also ones of the most commonly studied topics. 13.7% of studies was assigned to the group of records that could not be assigned to any specific ICD-10 code.

During the period spanning from 2016 to 2019, a total of 1,880 SLRs focusing on RCTs were published. One significant change in comparison to the 1994-2015 period is that SLRs on endocrine, nutritional, and metabolic diseases (E00-E90, 12.3%) outnumbered the reviews focusing on circulatory system diseases (I00-I99, n=11.3%). Additionally, the number of studies analysing the impact of supplementations or diets on health increased twofold.

Half of all the identified RCT SLRs were published in recent years (2020 to May 2023; n=3,416). Similarly to the period 2016 - 2019, the most frequently analysed disease area was endocrine, nutritional, and metabolic diseases (11.5%), followed by the diseases of the circulatory system (n=10.6%) and of the musculoskeletal system and connective tissue (8.9%). The number of studies analysing the impact of supplementations or diets on health further increased by twofold in comparison to the previously analyzed period. Among the recent years, there was a threefold increase observed in the number of trials concerning nervous system diseases (G00-G99) such as Alzheimer's disease. The first appearance of codes for special purposes (U00-U99; n=80), related to the COVID-19 pandemic, is also noteworthy.

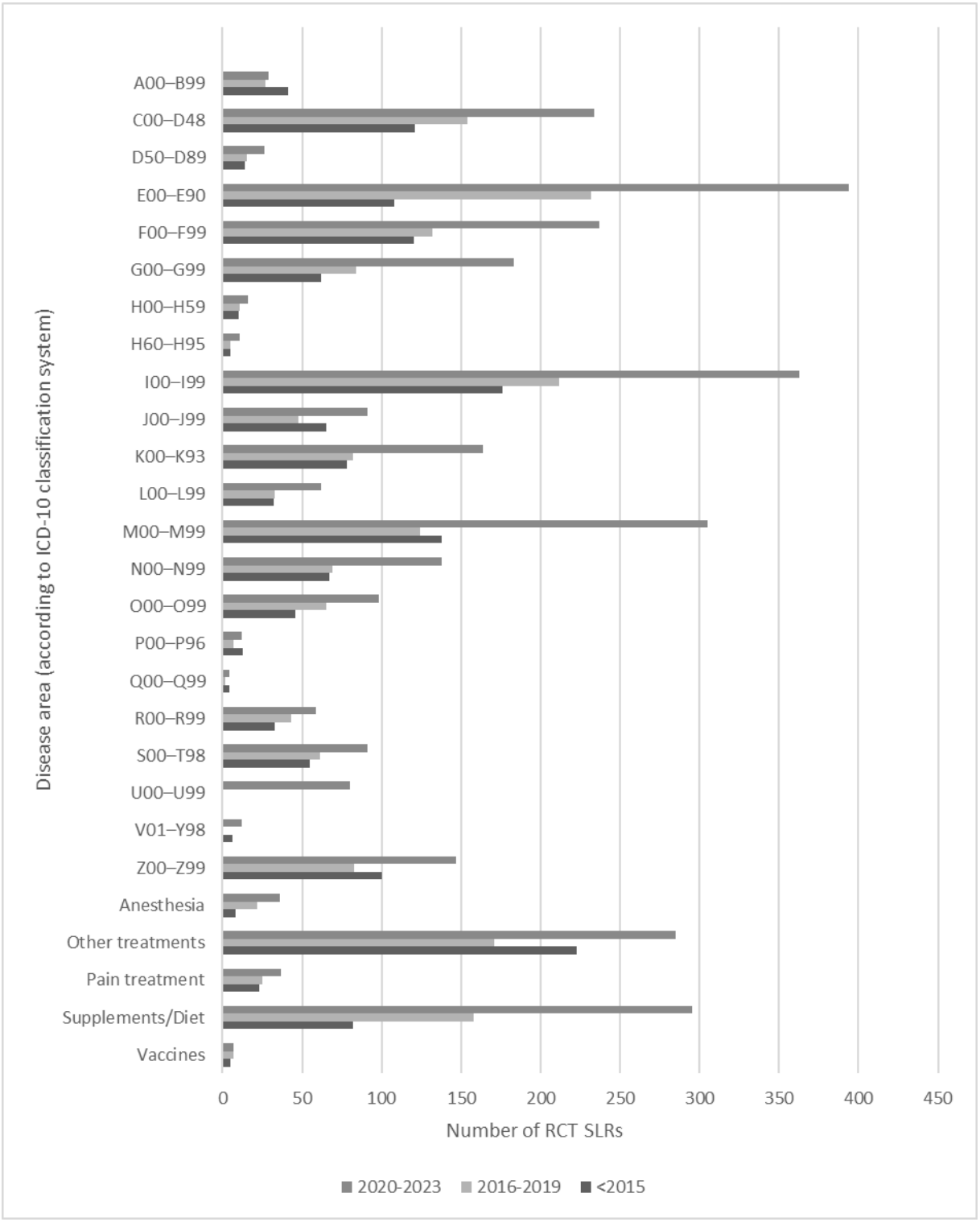


Figure 5 The number of RCT SLRs stratified by disease area, published in 3 distinct periods: from 1994 to 2015, from 2016 to 2019 and from 2020 to May of 2023

1.4.2. SLRs on epidemiological studies

According to the performed analysis, the overall trend in number of the SLRs of epidemiological studies is similar to the RCTs SLRs: the amount of published evidence for both study designs consistently grows, although SLRs of RCTs are more numerous (e.g.: RCTs SLRs: n= 10,061; SLRs of epidemiological studies: n= 8,947, in 2022) (Figure 6).

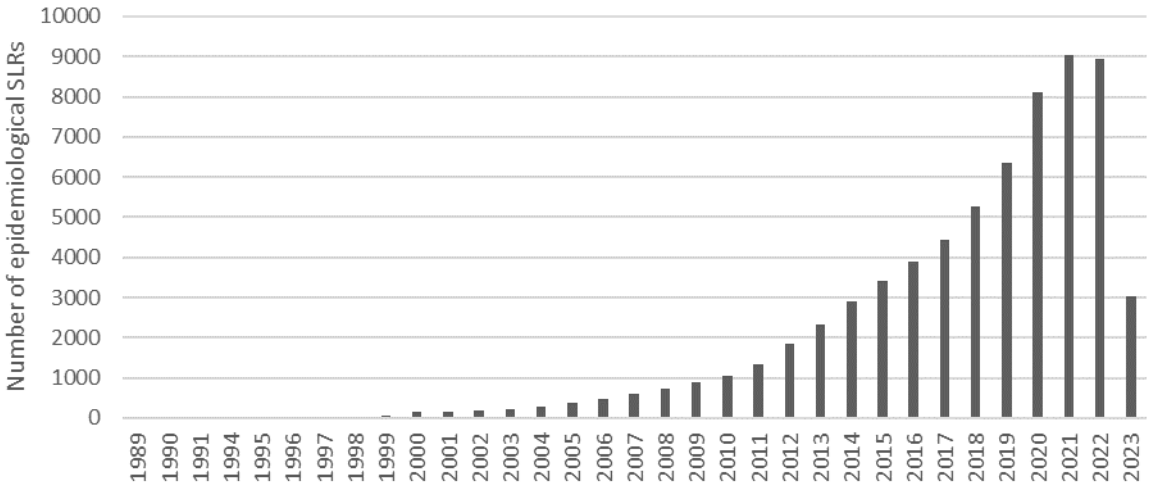


Figure 6. The number of epidemiological SLRs published over the years. Source: PubMed (search run in May 2023)

1.5. Discussion

SLRs are becoming more crucial than individual studies, as they help with comparing studies, combining findings, and making evidence more accessible. They also support the identification of the most cost-effective treatments and help future research to be better designed.

Publication rate of RCT SLRs is growing with an exponential rate. The increase observed during the last 3 years is partly driven by the emergence of the COVID-19 pandemic. Nevertheless, according to the analysis presented, the diseases most frequently assessed through SLRs focusing on RCTs align with the primary causes of mortality and morbidity globally, as outlined in reports published by the World Health Organization (WHO) and Institute for Health Metrics and Evaluation (IHME)[29, 30].

The Global Health Estimates published by the WHO, and presenting data for the period 2019-2020, point that non-communicable diseases (e.g.: chronic diseases such as heart disease, diabetes, chronic respiratory disease, or cancer) cover 7 of the top 10 causes of death worldwide[29, 38]. The leading were ischemic heart diseases, in 2019 accounting for 8.9 million deaths, with stroke being at the second place, causing 11% of deaths[29]. Those numbers are consistent with the number of RCT SLRs in the cardiovascular area and expounds the interests that physicians and trialists worldwide take in cardiovascular diseases. With such a high mortality caused by abovementioned diseases, it is crucial to study all possible interventions that can help to ease the suffering of many patients and extend their life expectancy and quality. Among last years, deaths caused by diabetes increased by 70% globally. It represented the highest percentage increase of all WHO regions[29, 38], and was also observed in the published SLRs' trends. Another change observed in the recent years, and visible as well in the conducted analysis, was the appearance of Alzheimer's disease and other types of dementia among the top 10 death causes worldwide[38].

It is noteworthy to emphasize the increase in the quantity of SLRs focusing on food supplements and diets, underscoring the heightened interest in these interventions. Considering the expanding geriatric population, the emphasis on preventive healthcare, the overarching trend towards well-being, and the inclination towards transitioning from conventional pharmaceuticals to supplements and diets, this pattern is likely to persist.

The next interesting point is that not only the number of RCT SLRs is rapidly increasing. The same trend is observed for SLRs of epidemiological studies; their amount is about to catch up the number of RCT SLRs. The recently observed intensified trend of RWE analysis, increase in digitisation of health records facilitating data analysis and escalating focus on the importance of patient-reported outcomes, may have played a role in the development of epidemiological studies[39].

When it comes to strengths of this work, the methodology used was well-thought-out and exhaustive. The first search was performed in PubMed and results were analyzed to get overall trends in the amount of RCTs and SLRs of RCTs published per year. Then, to complement general findings with a comprehensive analysis, an RR was conducted in MEDLINE database. The results were compared to analyse consistency between them. Additionally, based on information provided in titles and abstracts, each paper was coded according to a specific ICD-10 code and when no ICD-10 code was applicable, to "Treatments" code, consisting of several subcategories. When appropriate, two or more codes were selected. Moreover, data regarding the number of RCTs included in each analysed SLR was also retrieved and divided into several categories. What is also important, all SLRs analysing RCTs were included into this analysis, without any restrictions on population, interventions, or outcomes, and no restrictions on date or language were applied.

However, despite the strengths of the project, it is important to acknowledge several limitations. The first of them should be the fact that searches were conducted using a highly focused search strategy and study design filters. When running search in MEDLINE database, queries were searched only in titles of records. The objective was to get a reasonable sample of hits to be analyzed, and the approach was backed up by the fact that SLRs on RCTs of good quality, when published, usually contain study design terms in title. Despite the rationale, this is still a limitation that should be kept in mind. Additionally, the possibility of mistakes made by analysts during screening and assigning studies to ICD-10 codes should be considered.

As the number of RCTs and epidemiological studies will grow, there will be increasing need for unbiased evidence summary provided in a timely manner, supporting healthcare policy decision making.

In light of these considerations, novel solutions will be imperative, necessitating the development of RRs with refined and tailored methodologies, the integration of AI tools to enhance the efficiency of LRs, and the optimization of processes involved in LR execution, including using well-founded, specific and effective study design filters.

2. LRs and HTA agencies' requirements

2.1. Key findings

The questions being answered by this section:

- **What are the current recommendations regarding LRs supporting the refund process among different countries?**
- **What impact do they have on methodologies being used?**
- **What will be the future of HTA submissions? Is there a place for RRs?**

- The clinical and economic evidence criteria stipulated by HTA agencies in the UK, France, Germany, Canada, the US, Australia, and Japan, alongside the latest specifications delineated by EUnetHTA, underwent comprehensive review. From this analysis, pivotal methodological data pertinent LRs were systematically extracted and compared
- Among all the analyzed agencies there is a common recommendation regarding clinical data:
    - They have to be collected in a systematic way, from multiple sources, and based on the prespecified research question
    - The quality of identified evidence should be critically appraised to ensure maximum bias reduction
- Those recommendations are, however, very general and leave a space for adjustments regarding e.g. : search strategies development, usage of software supporting the process, the way of data extraction and quality-check
- Constructing economic evaluation depends on the specifics of a country's healthcare system; each agency provides guidelines adapted for country-specific realities
- Only HAS requires an SLR for economic evaluations
- Regarding review on utilities, NICE and HAS require them to be systematic; both agencies require the rationale for search terms used and presentation of the strategies and inclusion and exclusion criteria applied. They do not specify databases to be searched, however, require searches to be exhaustive

- Recommendations regarding LRs on the economic burden are less strict among the analysed agencies. Only the French agency requires an SLR. However, NICE, HAS, and IQWiG highlight the need of providing detailed rationale for further inclusion and exclusion of costs data

➢ **When constructing an HTA submission, the manufacturers must consider the specifics of the given market**

➢ **There is a place for RRs when it comes to economic reviews, however, a specific approach to the methodology should be developed and tested so it can reduce the risk of bias and make submission consisted and comparable**

➢ **Regarding clinical evidence, there is a room for new solutions such as support of software and optimization of SLRs' processes, including using well-founded, specific and effective study design filters.**

2.2. Introduction

HTA is defined by WHO as "systematic and multidisciplinary evaluation of the properties of health technologies and interventions covering both their direct and indirect consequences"[40], and as such, offers an approach for improving the knowledge base for healthcare decision making across a broad range of different technologies[41, 42]. Health technologies includes medicinal products, devices, and procedures, as well as specific measures for disease prevention, diagnosis or treatment[43]. There are many HTA bodies – as of 2021, in 32 settings in European Union, the UK, Canada and Australia, 63 active HTA bodies/institutions have been functioning[44]. Many of HTA agencies exist as autonomous governmental bodies, having either an advisory or regulatory function[45]. The topic selection process is typically not transparent, with the belief that most agencies predominantly assess new medical technologies that are expensive or with uncertain benefits[45].

HTA can be characterized by three phases:

- Assessment (collection and critical review of evidence)
- Appraisal (review of the assessment and providing recommendation)
- Implementation.

However, it is essential to note that not all HTA bodies conduct all of them; in particular, other organizations usually implement decisions[46].

For medicines, the process usually starts when a pharmaceutical company submits documents to an HTA body to market their medicine or technology – the manufacturer makes a detailed case (supported by clinical trial evidence) as to why their intervention is suitable for reimbursement within the given country health system[47].

HTA bodies usually rely on SLRs and meta-analyses based on existing relevant primary data studies to formulate conclusions[46]. SLRs' objectives are identification, evaluation and summary the data from all relevant individual primary studies, and making the evidence more accessible to decision makers[48]. Another significant aspect of introducing new technologies are economic considerations and their relation to effectiveness. When assessing new health technology, relevant comparators are being considered and economic evaluation is conducted in order to determine the dominance or non-dominance of introduced intervention[49].

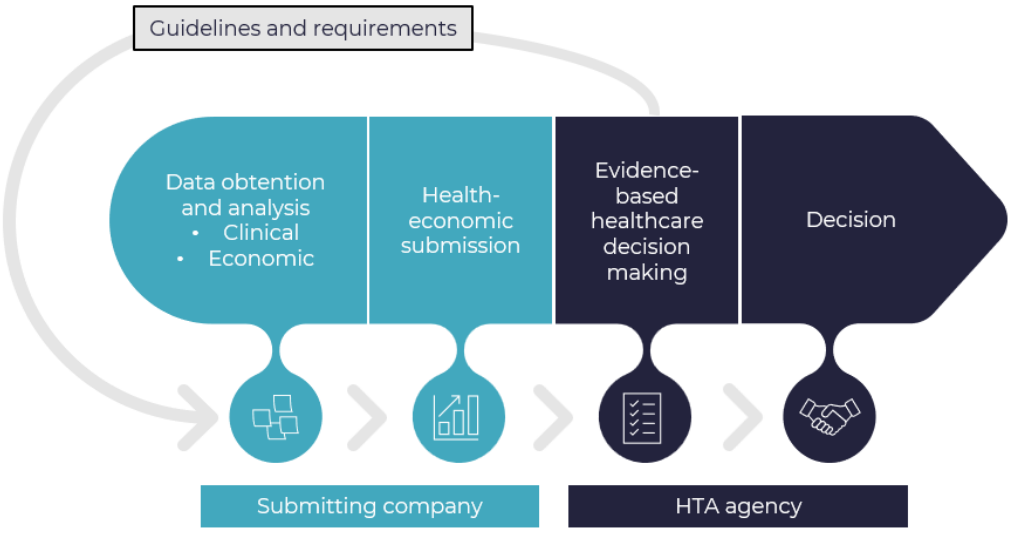


Figure 7 Simplified process of healthcare policy decision making

2.3. Methods

HTA is a key tool in healthcare decision making and HTA requirements regarding SLRs vary globally. The objective was to overview, compare and make key conclusions on clinical and economic evidence requirements of HTA agencies in the UK, France, Germany, Canada, the US, Australia and Japan, and the new requirements published by EUnetHTA. Comprehensive searches of selected HTA agencies' websites were performed in April 2022 and updated in May 2023. The most up-to-date guidelines for SLRs development used in products' submissions were identified, and the most methodologically important data were extracted and compared. Following agencies were considered:

- The UK: The National Institute for Clinical Excellence (NICE)[11,12]
- France: Haute Autorité de Santé (HAS)[16]
- Germany: Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG)[13]
- Canada: Canada's Drug and Health Technology Agency (CADTH)[50-52]
- The US: Institute for Clinical and Economic Review (ICER)[14]
- Australia: Pharmaceutical Benefits Advisory Committee (PBAC)[15]
- Japan: Core2 Health (C2H)[53]
- Europe: European Network for Health Technology Assessment (EUnetHTA)[54,55].

2.3.1. The analysis of the selected documents

The selected documents has been analyzed and information regarding the following domains were extracted and compared for guidelines on the clinical evidence LRs:

- SLR needed: if SLR is required by the agency; if yes, how SLR is defined, on which aspect guideline is focused most
- PICOS: if the need of PICOS defining is mentioned by the agency; if yes, to what extend it should be precise; if there is any information regarding possibility to modify the PICOS criteria during the review process
- Databases and sources: if data sources to be searches are listed; if there are any requirements regarding number of databases to be searched; if there is any information regarding recommended platforms to be used
- Search strategy: if information regarding search terms to be applied is presented; if there are recommendation on how broad should the strategy be; if there are any rules regarding searches restricting, or any rules for specific filters and limits to be used
- Selection process: if document presents requirements regarding selection process; should the process be done by two reviewers, what about their seniority level, what should be the process of discrepancies management
- Critical appraisal: What tool is recommended for assessing the quality of the studies included.

When it comes to guidelines on the economic reviews, as the recommendations were less precise than the ones for the clinical part, it has been decided to create and assess two main domains: approach to search strategy and studies selection, and critical appraisal.
Results are presented in the following sections.

2.4. Results

2.4.1. Requirements for clinical LRs

The findings are summarized below (Table 2). Depending on the requirements the following assessments were assigned:

- ‘’-‘’ no requirements specified
- ‘’+’’, ‘’++’’ or ‘’+++’’ for the least to the most demanding requirements.

Table 2 HTA's level of requirements regarding clinical LRs

Agency	SLR needed	PICOS	Databases and sources	Search strategy	Selection process	Critical appraisal	Overall score
NICE[11,12]	+++	+++	+++	+++	+++	+++	18/18
HAS[16]	+++	+++	++	++	+	+++	14/18
IQWIG[13]	+++	+++	+++	+++	++	+++	17/18
ICER[14]	+++	+++	+++	++	+++	+++	17/18
CADTH[50-52]	+++	+++	++	++	+	+++	14/18
PBAC[15]	+++	+++	+++	+++	++	+++	17/18
C2H[53]	+++	++	-	+	+++	+++	12/18
EUnetHTA[54,55]	+++	+++	+++	+++	+++	+++	18/18

The most demanding requirements were presented in guidelines of NICE and EUnetHTA[11, 12 54, 55]. The recommendations include details regarding rules to be followed when conducting search strategies, studies selection process, and inform how to manage discrepancies and critically appraise studies.

All analysed HTA agencies recommend that review of the available clinical evidence should be conducted in a systematic manner[11, 12 13-16, 50-55].

The need of defining PICOS criteria is underlined by agencies in the UK, the US, Canada, Germany and Australia, as well as by EUnetHTA[11, 12 14 50-52 13 15 54, 55].

At the same time, Japanese agency only loosely mentions the clinical questions should be clearly presented, using PICOS as an example of question structure[53]. HAS does not mention PICOS directly, however, it invokes to both, Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) guidelines[56] and Cochrane Handbook[3], which require the research question to be structured around PICOS[16].

2.4.1.1. Databases and sources

HTA agencies from the UK[11, 12], Germany[13], Australia[15], the US[14], and EUnetHTA[54] have the most specific recommendations regarding data sources. To cover the majority of available data, those agencies suggest using three databases: MEDLINE, Embase and CENTRAL. The discipline-specific databases (e.g., PsychINFO or CINAHL) can be searched when appropriate. The restriction to electronic databases only may introduce selection bias. Consequently, the aforementioned agencies describe other methods, such as scanning reference lists of included studies, searching relevant Internet resources, handsearching key journals, citation searching, experts contacting, etc. The other recommended step is also clinical trials registries search.

HAS only briefly suggests that all clinical data sources for the intervention and comparators should be identified by using a reproducible, systematic methodology, and presented in accordance with HAS guidelines and international standards; however, it is specified that they recommend the Cochrane Handbook, which is in line with recommendations described above[16, 3].

CADTH does not present transparent guideline on clinical evidence synthesis, only mentions that it considers a variety of information, e.g.: clinical practice guidelines, clinical trial protocols, availability of comparators, and stakeholder input (i.e., information from clinical experts, patient groups, expert review committee members and drug plans). It should be noted that CADTH provide readers with variety of searching solutions, such as a tool for health-related grey literature searches[50-52].

C2H agency does not present any recommendations on data sources[53].

2.4.1.2. Search strategy

NICE underlines the necessity for search strategy to be reproducible. Review's authors should present a full list of all data sources and databases' full electronic search strategies, including any restrictions applied[11, 12]. CADTH also highlights the search strategy should be available in the clinical review[50-52]. Using the combination of MeSH terms and free-text when running search is recommended by IQWiG[13], ICER[14], EUnetHTA[54, 55], and PBAC[15]; additionally, PBAC underlines that filters should initially be set to restrict study design to

only randomised trials (Cochrane Highly Sensitive Search Strategies), while regarding population, search has to be broad and include synonyms for the target patients/disease/condition[15]. For intervention and comparators, PBAC suggests including MeSH terms, known proprietary and non-proprietary names, and developmental/provisional medicine names. IQWiG suggests the use of the PRESS checklist to make sure the quality of the search is relevant[13]; though not clearly recommended in the guidance, CADTH also provides access to PRESS tool on their website. Canadian agency also created the Search Filters Database which presents an assortment of search filters developed and maintained by CADTH's Research Information Services Filters Working Group[50-52].

Regarding reporting, PBAC specifies that search strategy run in MEDLINE should be presented; any key differences from this search strategy should be presented for other sources, e.g. Embase[15]. Language restriction is allowed by both: German[13] and US[14] agencies; they highlight that evaluations have shown the exclusion of non-English publications does not change the conclusion of most SLRs. It is worth to note that CRD guidance (followed and recommended by NICE) underlines that restriction to English language only may introduce language bias[48]. ICER allows for the exclusion of articles indexed as guidelines, letters, editorials, narrative reviews, case reports, or news items, and animal studies[14]. HAS invokes Cochrane Handbook and PRISMA-P guidelines[16].

The least detailed requirements are provided by C2H and CADTH: C2H states only that information about databases used, search algorithm and research selection process is required[53], while CADTH conducts one or more independent systematic searches according to the protocol to supplement the submission provided by the sponsor. The search strategy used and the identified literature are included in the clinical evidence review[50-52].

2.4.1.3. Selection process

Two-step screening phase (firstly by titles/abstracts, and then by full texts) is described in guidelines from the UK[11, 12], Germany[13], EUnetHTA[54, 55], the US[14], Australia[15], and Japan[53]. Only NICE[11, 12], IQWiG[13], EUnetHTA[54, 55], and ICER[14] highlight the need of performing double screening by the independent analysts. NICE and ICER require that the reasons for exclusion should be noted for both phases: titles/abstracts screening and full text review, while IQWiG requires reasons tracking only for the second phase[11, 12, 14 13]. PBAC, C2H and HAS do not specify if exclusion reasons for titles/abstract screening are needed[15, 16, 53].

PRISMA flowchart is required by most agencies[11-13, 16 14, 15, 53]. All agencies highlight the necessity for documentation of the selection process.

PBAC and C2H suggest expanding the scope of the clinical evidence SLR to include non-RCTs when no sufficient evidence was identified in RCTs[15, 53].

2.4.1.4. Critical appraisal

All analysed agencies highlight the necessity of performing a quality assessment of the identified studies using the appropriate tools.

2.4.2. Requirements for economic LRs

The findings are summarized below (Table 3). The most important fields were selected, and depending on the requirements the following assessments were assigned:

- ''-'' no requirements specified
- ''+'', ''++'' or ''+++'' for the least to the most demanding requirements.

Table 3 HTA's level of requirements regarding economic reviews

Agency	Approach to search strategy and studies selection	Critical appraisal	Overall score
NICE [11,12]	++	+++	5/6
HAS[16]	+++	+++	6/6
IQWiG[13]	++	-	2/6
ICER[14]	+	-	1/6
CADTH[50-52]	+	-	1/6
PBAC[15]	++	-	2/6
C2H[53]	++	-	2/6

Since EUnetHTA did not present de novo guideline regarding economic reviews, but rather summarised recommendations from European countries and made conclusions, it has been not included in the further analysis[54,55].

Overall, NICE and HAS present the most demanding requirements for economic evaluations reviews, presenting detailed recommendations on methodology for studies selection and critical appraisal[11, 12, 16]. Japanese HTA agency does not provide detailed instructions, just recommends conducting SLR of Japanese and overseas clinical research to obtain effectiveness, safety, and QoL data[53].

2.4.2.1. LR of economic evaluations

Among European HTA agencies, only HAS requires an SLR of previous economic evaluations. This recommendation is not clearly specified in the agency's guidelines, however, it is based on what EUnetHTA reports[16,54,55]. NICE's approach is less precise; reviews on economic evaluations may not be systematic if newly identified studies would merely provide additional support consistent with the already-identified evidence[11, 12]. Interestingly, EUnetHTA interpretation of NICE's recommendation is different; in their opinion, both, France and the UK's guidelines request the SLR over previous economic evaluations is conducted[54,55]. IQWiG allows a focused search of previous evaluations, run at least in MEDLINE, and specifies that analyses of secondary data should be in line with guidelines on the good practice of secondary data analysis. If guideline is used, should originate from the German or comparable healthcare system, and should be evidence-based[13]. Australian HTA agency does not specify whether economic evaluations LRs should be systematic, however, highlight the need for conducting them[15]. Other agencies

do not specify any rules related to searching for previously published economic evaluations.
Only NICE and HAS require the critical appraisal of identified economic evaluation studies[11,12 16]. They both expect validated instrument to be used, e.g. Drummond checklist. NICE highlights that studies' results should be interpreted with reference to critical appraisal of their methodology, and quality assessment process should be completed for each identified cost-effectiveness study[11, 12]. HAS suggests that only studies with the highest level of evidence should be included and justified, clear discussion should support estimation of the robustness of the conclusions, and define the conditions under which the conclusions would be altered[16]. Other agencies do not present any details regarding the quality assessment of the included studies.

2.4.2.2. LRs on utilities

Two European agencies, NICE and HAS, require LR on utilities to be systematic. They require the rationale for search terms used and presentation of the strategies and inclusion and exclusion criteria. Databases that should be searched are not specified, however, agencies require searches to be exhaustive[11,12 16].
PBAC recommends gaining utilities from literature when they cannot be estimated from clinical studies data; the details of searches, and the PICOS used to identify utility studies should be presented. Any data and references supporting the selection of the multi-attribute utility instruments are required to be included in a technical document or an attachment. However, PBAC does not specify if the search should be systematic[15].
IQWiG states that a focused literature review can be conducted to obtain utility values, if considered necessary[13]. The Canadian agency does not specify the methods for review running but require health state utility values to be based on their fitness for purpose, credibility, and consistency, and that the trade-offs among these criteria should be described and decision on the selected sources should be justified[50-52].

2.4.2.3. LRs on economic burden

Recommendations regarding LRs on the economic burden are less strict among the analysed agencies. Only the French agency requires an SLR[16]. NICE specifies that when the systematic search results present limited data for England, the search may be extended to capture data from other countries[11,12]. IQWiG states only that data for costs (also for epidemiology) may be collected from various sources, without moving into details[13]. Australian agency advises that when possible, the source of costs recommended by the Manual of resource items should be used, and C2H requires claims databases established in Japan to be used[15, 53]. CADTH states that during resources valuation, data sources that most closely reflect the opportunity cost should be selected, given the perspective of the analysis. Fees and prices listed in formularies and schedules of

Canadian ministries of health are recommended as unit-cost measures when the perspective of the public payer is considered, as long as they reflect actual payments[50-52].

NICE, HAS, and IQWiG highlight the need of providing detailed rationale for inclusion and exclusion of costs data[11, 12,16,13]. C2H states that Japanese data with high level of evidence should be used preferentially[53].

2.5. Discussion

HTA agencies, both regional and national, provide recommendations on drugs and medical devices that can be reimbursed by the healthcare system. Those agencies make decisions based on a complex submissions from manufacturers, and each organization has unique guidelines or set of requirements and templates to guide companies in preparing documentation needed. Data from clinical trials (mainly RCTs) serve as a foundation for health economic evaluations, which then allow the HTA bodies to provide evidence-based decisions.

This analysis shows that among all the analyzed agencies, there is a common recommendation that clinical data has to be collected in a systematic way, from multiple sources, based on the prespecified research question, and the quality of identified evidence should be critically appraised to ensure maximum bias reduction. However, constructing economic evaluation depends on the specifics of a country's health care system. For instance, a recent paper by Thokala et al. that compared two HTA agencies – ICER and NICE – underlines that crucial differences stem from the UK's single payer health care system and the US's pluralistic system[57]. Therefore, despite several consistent recommendations regarding economic evaluations (such as the need for a sufficient time horizon or the importance of costs and outcomes discounting), each agency provides guidelines adapted for country-specific realities.

The primary strength of this analysis lies in its comprehensive overview and key conclusions on the HTA requirements pertaining to LRs, presented by seven distinct agencies, juxtaposed against each other, and against the latest guidelines released by EUnetHTA. This work involved comprehensive searches of designated agencies websites and thorough analysis of recommendations, prioritizing methodological significance. Data extraction from guidelines underwent rigorous quality control by two independent researchers to mitigate errors.

However, like any research endeavor, this analysis is not immune to limitations. Potential errors in data extraction and the temporal scope of the analysis, current only as of May 2023, warrant consideration. HTA agencies constantly refine guidelines and requirements, so it is important to keep in mind those have evolved by the time of this thesis publication.

In terms of final conclusions and implications, this analysis underscores the imperative for manufacturers to intricately consider market-specific nuances and adhere to the latest guidelines when constructing HTA submissions. While there is room for utilizing RRs in economic reviews, a tailored and tested methodological approach is essential to mitigate bias and ensure consistency in submissions. Regarding clinical evidence reviews, a systematic approach is always required, however, to support the process, the discussions are ongoing and guidelines regarding software and AI tools usage are under development.

Nevertheless, opportunities exist for novel solutions and optimization of SLRs, including the implementation of robust study design filters.

3. Rapid review in the era of exploding volume of information

3.1. Key findings

The questions being answered by this section:

- **What is the current state of knowledge regarding RRs?**
- **Which methodologies are most frequently recommended by guidelines?**
- **What is the future of RRs; to what extent can they replace SLRs?**
- To identify the current state of knowledge regarding RRs, 3,813 titles and abstracts were screened, and an analysis was prepared based on 12 finally included publications: 7 guidelines, 2 papers summarizing the results of the Delphi consensus on RRs' methodology, and 3 publications assessing different approaches to RRs
- The general objectives of conducting RR are to speed up and simplify the review process, but there are different methodologies being used and the research shows they depend on authors' subjective opinions
- Based on the results from the modified Delphi method including inputs from 113 stakeholders:
    - The most feasible approach to RRs is the one combining inclusion of published literature only, searching of more than one database and restricting by date and language, with single study selection, double data extraction and quality assessment (72%, 81/113 responses)
    - Considering timelines, the best approach assumes updating the search from a previously published LR, with no newly applied limits, single studies selection and data extraction, and no quality assessment
    - The most comprehensive approach assumes searching more than one database and grey literature, using date limits, and one reviewer working on screening, data extraction, and risk of bias assessment
- Authors of reviews often use the terms SLR, RR, targeted literature review, and desk research interchangeably, which can result in significant consequences, especially in the healthcare area

- There is no final consensus on the formal definiton and the best approach to RRs. Only in 2021, the unified definition of RR was published by the Cochrane group. However, it is a very general definition, matching only slightly over 50% of papers analysed by the group, which highlight the lack of homogeneity in this area
- When it comes to the methodological approach's details, the Cochrane group's requirements presented the most demanding ones, with some of rules being exactly the same as for the comprehensive SLRs

➢ **When resources and time are limited, RRs can be an efficient tool to provide the summary of evidence critical for making rapid policy-related decisions**

➢ **There is more work needed to ultimately distinguish between SLRs and RRs, and to define the most robust, justified methodological approaches to RRs depending on specific objectives.**

3.2. Introduction

The methodology and transparency of SLRs make them the most reliable type among LRs [58], providing an objective and comprehensive, summary of the evidence for a given topic[19, 20]. On the other hand, the SLR conducting often requires a lot of time and human resources. A high increase of new publications and stable timelines for activities such as HTA submission result in a growing need to analyze data and synthesize evidence faster. Here, RRs can be a solution and can partially replace SLRs. However, there are still some limitations as lack of guidelines distinguishing between SLRs and RRs and presenting methodologies of RRs adjusted to the specific objectives.

Without new approach, running SLRs supporting different activities can be unaffordable very soon. A good example is the COVID-19 pandemic, characterized by an outburst of demands for evidence, when reliable syntheses of data were essential due to the large amount of research that has been and continues to be produced. There were some studies conducted to assess the quality of LRs developed during the pandemic. One of them was study conducted by Baumeister et al. assessing the quality of SLRs published in the first 5 months of the COVID-19 pandemic. Authors analyzed 584 records labelled as SLRs with/without meta-analysis, RRs, scoping review or umbrella review, and 145 of them were assessed as of poor quality, due to issues with methodology: critical steps were often omitted without a proper explanation of the shortcuts taken, the reasoning behind them, and an explanation on how they affected the review outcomes[59]. Eighty-one studies did not precise the study designs included, 15% did not report excluded studies and only 5% of reviews listed the reasons for exclusions. Only in 104 reviews double screening was reported, and in 118 data extraction was performed by two independent reviewers. Less than 4% of studies

provided an information on the funding. Critical appraisal of studies was addressed in 48.7% of the reviews, and heterogeneity analysis in 67%. In another study, the quality of SLRs on the COVID-19, severe acute respiratory syndrome, and Middle East respiratory syndrome outbreaks, was assesses. Among 49 reviews included, according to AMSTAR-2 tool, only 6 were of moderate quality, the rest were of low (12 studies) to critically low quality (31 studies). The researchers suggested that the factors behind the poor quality of the reviews might be: urgency, lack of experienced reviewers, absence of funding, and lack of primary studies in the subject[60]. It shows that performing LRs of good quality within short timelines, and without firmly established rules, is not feasible, leads to errors and can have a significant negative impact on healthcare decision making.

RRs are a quite new area, and still no agreement on optimal methods can be found[61]. They are completed within shorter timeframes than SLRs, and the lack of time is one of the main reasons they are being conducted. It has been suggested that most rapid reviews are conducted within 12 weeks, however, some of the resources suggest time between a few weeks to no more than 6 months[21, 25]. Despite the fact that RRs are run by many scientists around the world and play an important role in both, health economics and outcomes research, as well as Market Access fields, the first definition of RR hasn't been published until 2021[24].

The objective was to run RR and to identify and summarize available information regarding different approaches to RR defining and the methodology applied when conducting such reviews.

3.3. Methods

Searches were run in the MEDLINE and Embase databases in November 2022. The following keywords were searched in titles and abstracts fields: 'targeted adj2 review' OR 'focused adj2 review' OR 'rapid adj2 review', and 'methodology' OR 'design' OR 'scheme' OR 'approach'. Additional search was run in Google Scholar with keywords including 'targeted review methodology' OR 'focused review methodology' OR 'rapid review methodology'. Only publications in English were included, and the date of publication was restricted to year 2016 onward in order to identify the most up-to-date literature. The reference lists of each included article were searched manually to capture the other potentially eligible articles. Titles and abstracts were screened by one reviewer to exclude evidently irrelevant articles, then, the reviewer analyzed full texts of potentially relevant papers to examine their eligibility.

A pre-defined Excel grid was developed, and the extraction of the following information related to the methodology of RR from guidelines was conducted:

1) Definition
2) PICOS development and search strategy
3) Studies selection
4) Data extraction and quality assessment
5) Data presentation
6) Any additional comments.

There was no restriction on the study types to be analyzed; any study reporting on the methodology of RRs could be included: reviews, practice guidelines, commentaries, and expert opinions on RR relevant to healthcare policymakers or practitioners. All of the other records that did not focus on RRs' methodology, were excluded. As this was an RR, a single screening, as well as single analysis of full texts, were run, with no recording of exclusion criteria. The data extraction and evidence summary were conducted by one analyst and further examined by a senior analyst to ensure the quality. Disagreements were resolved by consensus.

3.4. Results

A total of 3,864 records have been identified through MEDLINE and Embase searches, and Google Scholar search yielded 34 articles. After deduplication 3,813 titles and abstracts were screened. Then, 43 full articles were analyzed resulting in 12 publications finally included: 7 guidelines[22-25, 27, 61, 62] on the methodology, 2 papers summarizing the results of the Delphi consensus on RRs' methodology[63, 64], and 3 publications assessing different approaches to RRs[19, 28, 65].

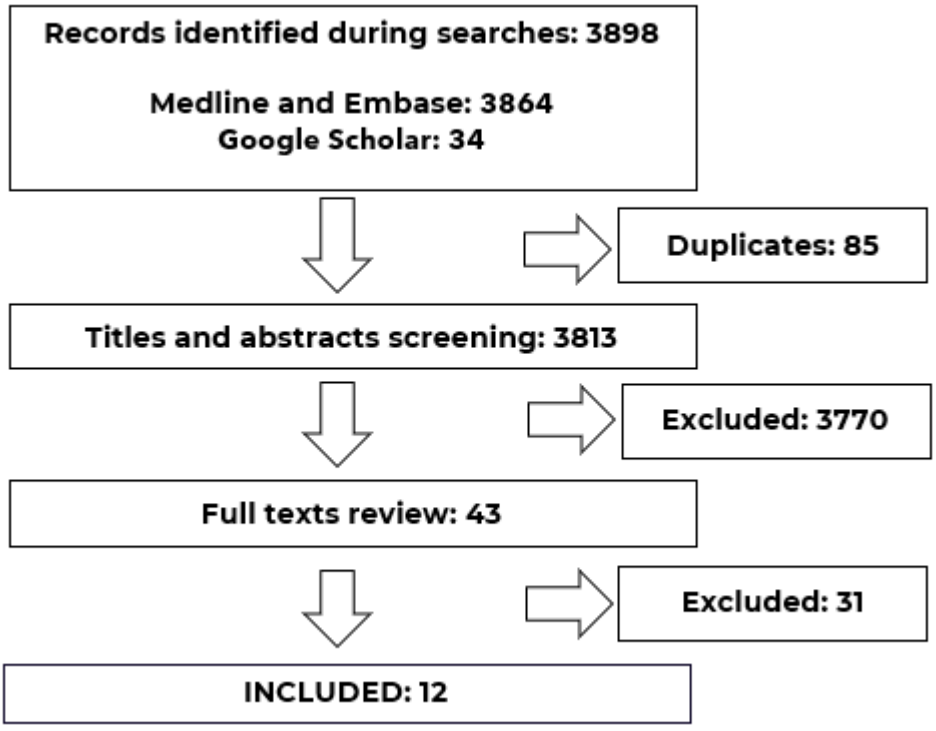


Figure 8 PRISMA diagram

3.4.1. Guidelines

Information regarding definitions the authors used for RRs, approach to the inclusion/exclusion criteria and search strategy, studies selection and data extractions, quality assessment, and reporting were collected from all the identified guidelines.

3.4.1.1. Definition

Cochrane Rapid Reviews Methods group used scoping review of the evidence, primary methods studies conducted, the results of surveys sent to Cochrane representative and discussion between them, to develop the RR guideline[24]. They analyzed over 300 RRs and based on the methodology used among the studies, constructed a broad definition of RR: "*A rapid review is a form of knowledge synthesis that accelerates the process of conducting a traditional systematic review through streamlining or omitting a variety of methods to produce evidence in a resource-efficient manner*." This interpretation is in line with slightly more than 50% of RRs identified by the Cochrane group. Authors highlighted that RRs should be driven by the necessity of timely evidence generation for healthcare decision-making[24].

Except this one broad definition of RR, other so called ''definitions'' were rather a more or less detailed methodologies descriptions, different between each other, with no final statement. Authors were underlying that RR vary in their format and methods used, and that there is still a lack of agreement on optimal solutions to be used[61]. All the identified guidelines alluded to the fact that the main reason they are conducted is searching for time-saving solution. It has been stated that most RRs are being conducted within 12 weeks, or, as suggested by other sources, running them took between a few weeks to no more than 6 months[25, 61]. Different mechanisms of enhancing the timeliness of review were presented: increase of the intensity of work by parallelization of tasks, reducing one or more steps normally carried out in a review, automatizing RRs by using new technologies[61]. The UK government report[22] presented two different RRs: quick scoping reviews (QSR) and rapid evidence assessments (REA). While being less resource and time-consuming compared to standard SLRs, QSRs and REAs are designed to be transparent and to minimize bias, and QSRs are being applied more to open-ended questions[22].

3.4.1.2. Research questions and searches

As for SLRs, also for RR a clear research question and search protocol development is a step recommended for the beginning of the review by all the analyzed guidelines. Additionally, the Cochrane Rapid Reviews Methods Group recommend involving main review users to set the review question, inclusion criteria, and outcomes, and consult them thorough the entire process, and, moreover, suggest registering protocols with the international prospective register of reviews – PROSPERO (however, at this moment in most cases RRs are not formally registered)[25, 61]. In some authors's opinion, a search protocol should contain a description of how the following review steps will be carried out, including a full strategy together with a list of databases and other data sources, precise PICOS criteria, a detailed presentation of data extraction and critical appraisal processes, and a plan for data synthesis[22].

This approach mirrors exactly the methodologies proposed for systematic literature reviews (SLRs), thus failing to yield any savings in terms of time and human resources.

Regarding research questions, it is suggested to create them in a neutral way than focus on a specific direction for the outcome[23]. Researchers can add a supportive research question when needed[22], and it is recommended to limit the number of interventions and outcomes, just to focus on the most important ones from the decision-making

perspective[24]. In SLRs researchers should pay attention if the scope of the review is broad enough, and they cannot use many restrictions of PICOS. In contrast, when working on RRs, a researcher may spend more time defining each of the components of the PICOS, as the crucial step is ensuring that the research question addresses the needs of RR's recipients, and, at the same time, is feasible to be completed within the specified timelines[23].

Some of authors recommended doing a quick research or reading some SLRs on the topic at the begging of the work on RR to get a better understanding of the area[23].

In RRs, an exhaustive process of searches running will not be possible. The focus should be on effectiveness and efficiency of the search. Again, Cochrane Rapid Reviews Methods Group suggests more systematic approach, which is involving an information specialist and conducting peer review of at least one search strategy[24]. The rapid review guidebook by authors from McMaster University [23] recommend RRs being fed by the body of literature rather than single studies. In their opinion the effect would be more generalizable conclusions applied at the level of a population, and providing more realistic findings for program decisions[23]. Guidelines published by Collins et al. as well as Dobbins et al. recommend documentation of the search strategy with the date of searching and information regarding the applied date limits, to make sure it is repeatable. Furthermore, they suggest inclusion of the list of individual databases included in platform services to be reported for full transparency[22, 23]. Searching of a limited number of databases and other sources is recommended by some authors [23] with information that usually the PubMed/MEDLINE, Cochrane Library, and Embase databases are being used and commonly used limits are language, study design, date, and geographical area. Additionally, some authors suggest the grey literature searching, but direct contact with authors is rather uncommon[22, 61]. According to guideline of the University of Oxford, the search should be conducted in at least one major database such as PubMed, and one additional source e.g. Google Scholar[62]. In some authors' opinion, if there is any type of evidence that will not be included into the RR, it should be mentioned in the protocol and justified[22]. Additionally, authors of a guideline published by WHO suggest running a staged search to identify existing SLRs firstly, and then focusing on studies with other designs[61]. If there is a low number of records retrieved, reviewers may expand searches, remove some of the restrictions, and search additional sources, as the searching in RRs is an iterative process and usually some modifications and adjustments are needed [23]; and some organizations require for all of them to be tracked and reflected in the final report[61].

3.4.1.3. Selection of studies

The selection of studies consists of two phases: screening of titles and abstracts, and analysis of full texts. Some of authors suggest to conduct a pilot exercise using the same sample of records (30-50 abstracts and 5-10 full texts) for the entire team working on the RR, to calibrate and test the review form[24]. In RR, the studies selection can be conducted by one reviewer with or without quality-check by a second one, in contract to a standard SLR. When verification is performed, the second reviewer often checks only a subset of records and compares the results. The Cochrane Group recommends a specific, stricter approach: at least 20% of titles and abstracts should be double-screened, and the rest can be screened by one researcher, but the excluded ones need to be re-examined by the

second person; a similar approach is suggested to be applied for full texts review[24]. This helps to make sure the bias is reduced and the inclusion criteria are applied in a relevant way[22, 24, 61, 62]. According to Dobbins et al., there is no need to report reasons for exclusion for abstracts, however, they should be monitored for all excluded full texts[23].

3.4.1.4. Data extraction and quality assessment

According to the WHO recommendation, data extraction in RR is most commonly done by a single reviewer, with or without a partial quality-check. The authors recommend that the second reviewer verify a minimum of 10% of the extracted data for accuracy. Double-extraction is more needed for quantitative results than for descriptive study information[61]. In contrast, Cochrane group states the second reviewer should review the correctness and completeness of all the extracted data[24]. Authors from McMaster University highlight the extraction process should be conducted ideally by two researchers independently and if there are discrepancies, consensus should always be reached[23]. Plüddemann et al. present the least restrictive approach to the extractions and suggest the decision on the applied methodology should depend on the time available, and should reflect the complexity of the review[62].

According to several identified guidelines, extractions should be limited to crucial characteristics and outcomes. The same approach to extraction of information is also recommended for a quality assessment process within RRs[24, 61, 62].

According to the guide published by WHO, for the studies selection, data extractions, and quality assessments, reviewers may use supporting technologies making those steps of RRs more efficient[61].

3.4.1.5. Reporting

The guideline of the McMaster University recommends developing the report structured in a 1:2:20 format, with one page for key messages, two pages for an executive summary, and the rest of the report consisting of up to 20 pages including the background, status of the current knowledge, project question, synthesis of findings, conclusions, and recommendations[23]. Recommendations on reporting presented by other authors are related more to the methodology and potential risk of bias. They highlight that all the limitations of the review should be considered and conclusions should be made with caution[61]; it is worth to assess the quality of the evidence and strength of recommendations by using e.g.: the GRADE system[61]. When quoting the reference, researchers should use a primary source[23].

Ideally, all the standard phases of SLR should be also completed when running RRs; however, some steps may not be conducted as thoroughly as should be because of timelines, or may need skipping. Authors highlight how crucial it is to decide which steps are the key ones, and which can be skipped within the specific RR, and to make sure it will be described sufficiently in the report[23]. Some authors recommended inviting researchers with experience in SLRs to join the RR development and reporting process when possible[61, 62]. Also, using some software tools and templates can support its development and enhance the transparency[61].

3.4.1. Studies analyzing RRs' methodology

During the review, two interesting studies, presenting and summarizing the results of Delphi consensus on the RR methods, were identified[63, 64].

Tricco et al.[64] first run an international survey and scoping review to gather information on the various methodologies employed in the conduct of RR, based on which, authors employed a modified Delphi method including inputs from 113 stakeholders to search for the most optimized approach. Among the six most frequent approaches to RRs, the one combining inclusion of published literature only, searching of more than one database and restricting by date and language, single study selection, double data extraction and quality assessment, was ranked as the most feasible approach (72%, 81/113 responses) with the potentially lowest risk of bias (12%, 12/103) (Table 4, approach 1). When considering timelines, the best approach assumes updating the search from a previously published LR, no newly applied limits on search, studies selection and data extraction done by one reviewer, and no quality assessment (Table 4, approach 2). Finally, the most comprehensive RRs can be developed by searching more than one database and grey literature, using date limits, and assigning one reviewer working on screening, data extraction, and risk of bias assessment (Table 4, approach 5).

Table 4 Six most frequent approaches to RRs (adapted from Tricco et al.[64])

Domain	***Approach 1***	***Approach 2***	***Approach 3***	***Approach 4***	***Approach 5***	***Approach 6***
Literature search	More than one database, only published data	Update of previous literature search	More than one database, grey literature included			
Search limit	Date and language	None	Date and language	Date or language	Date	Date and language
Screening	One reviewer					Two reviewers
Data extraction	One reviewer and one verifier	One reviewer				
Risk of bias appraisal	One reviewer and one verifier	Not performed			One reviewer	Not performed
Overall assessment	Highly feasible and timely, having the lowest potential risk of bias	Highly feasible and timely	Moderate	Moderate	Highly comprehensive, but low-ranking in terms of feasibility and timeliness	Highly comprehensive, but low-ranking in terms of feasibility and timeliness

Pandor et al. [63] developed a decision tool for SelecTing Approaches for Rapid Reviews, produced through the Delphi consensus of international experts in a rigorous, iterative process. Experts were asked to provide their assessment of the importance of predefined items in four domains related to the RR process: interaction with commissioners, understanding the evidence base, data extraction and synthesis methods, and reporting of RR methods. All items assigned to those four domains achieved >70% of consensus, and the first consensus-driven tool supporting authors of RRs in planning and deciding on approaches, has been developed.

Haby et al.[28] searched eleven databases and two websites, and based on five SLRs and one RCT being finally included, they developed a comprehensive summary of the methodologies used in RRs. All the included SLRs were consistent regarding the fact there is no final, unified definition of RR, and no agreement on the best methodological rules to be followed. Among the approaches used by RRs' authors to make running them faster than SLRs, authors listed: limiting the scope of research questions, searching lower number of databases, restrictions on language and date (especially limits to English only, most recent publications), limited searching of grey literature, running updates of the existing SLRs, eliminating or restricting hand searches of references, noniterative search strategies, limiting double study selection, data extraction and quality assessment, eliminating consultation with experts, restricted data synthesis with concise summary of key findings and recommendations.

Gordon et al.[19] presented the advantages of conducting an RR and provided twelve tips for its performing. They define focused reviews as "a form of knowledge synthesis in which the components of the systematic process are applied to facilitate the analysis of a focused research question". The first advice is related to decision if an RR is a good solution for the considered project, as in authors' opinion RRs suit emerging topics for which early data synthesis can support policymakers, doctors, etc., but also they can direct future research. The following three tips highlight how crucial it is to run preliminary searches and narrow the number of records by using reasonable restrictions considering the local issues, context, and available time. Further recommendations include creating a team consisted of experienced reviewers, registering the protocol in the PROSPERO and considering the target journal from the beginning, contacting authors of studies when data available in publications are incongruent or missing. The last three suggestions are related to the decision on evidence synthesis method, usage of visual presentation of data, and description of all the limitations of the RR.

In the end, it is worth to mention the publication by Tricco et al. from 2022, describing JBI position statement[65] and highlighting that, so far, there is no specific tool that can be used for critical appraisal of the RR's methodology. Researchers can use tools being used by default for SLRs, like ROBIS, the JBI critical appraisal tools, or AMSTAR.

3.5. Discussion

Different terms are being used to describe the RR, including 'rapid review', 'quick scoping review', 'rapid evidence assessment', and 'focused review'. Although the general objectives of conducting RR are to speed up and simplify the process, there are different methodologies being used to achieve the goals, and, as presented in results from this review, they depend on authors' subjective opinions, and there is no final consensus on the formal definition and the best approach to RRs. Only in 2021, a study presenting the unified definition of RR was published. However, it is worth to note that the definition was matching only slightly over 50% of papers analysed by the Cochrane group, which highlight the lack of homogeneity in this area[24]. Additionally, when it comes to the methodological approach's details, the Cochrane group's ones were the most demanding, presenting some of rules being exactly the same as for the comprehensive SLRs[24].

The primary strength of this analysis lies in its exhaustive approach. A comprehensive search strategy was employed, utilizing both the MEDLINE and Embase databases to ensure the retrieval of all pertinent information related to RRs. Additionally, Google Scholar was utilized, and reference lists of identified articles were searched manually to capture any potentially relevant publications. An extensive examination of the gathered materials was undertaken using a predefined Excel grid, which systematically captured key domains including the definition of RRs, approach to developing the PICOS framework and search strategy, criteria for study selection, methods for data extraction and quality assessment, data presentation techniques, and any supplementary information influencing methodology. Furthermore, a notable strength of this work is the absence of restrictions on the types of studies included; any study delineating the methodology of RRs was considered eligible for analysis. Data extraction and evidence synthesis were independently conducted by two researchers, ensuring meticulous attention to detail and minimizing the potential for errors.

In light of the project's strengths, it is imperative to recognize certain limitations. Firstly, it's crucial to acknowledge that the results are reflective only up to November 2022, and given the evolving nature of the approach to RRs, some changes and new approaches could have appeared by the time of the thesis publication. Furthermore, owing to the rapid methodology employed in this research, only English-language publications were included, and the publication date was restricted to 2016 onwards to ensure the incorporation of the most current literature. While this selective approach is justified, it warrants consideration when interpreting the findings. Lastly, the single-researcher-based process for selecting studies introduces the potential for errors, which cannot be disregarded. All these factors should be considered when interpreting the study results.

Nevertheless, the identified references highlight the fact that when resources and time are limited, RRs can be an efficient tool to provide the summary of evidence critical for making rapid policy-related decisions[61]. However, there is more work needed to ultimately distinguish between SLRs and RRs, and to define the most robust, justified methodological approaches to RRs depending on specific objectives.

4. 3 RRs methodologies vs Cochrane's SLRs

4.1. Key findings

The questions being answered by this section:

- **Is the simplified methodology of RRs applicable for conducting clinical evidence reviews?**
- **To what extent can simplification be employed while maintaining effectiveness?**
- **What enhancements can be implemented to make RRs more effective in the future?**

- Three different approaches to RRs' methodology were created, tested, and compared; the impact of the simplified methodologies on workload reduction, as well as the impact on final results of meta-analyses, were assessed in comparison to SLRs performed by the Cochrane group
- Methodologies differed between each other in terms of: data source searched, single/double studies selection, supporting software tool usage, and the researchers' awareness of the expected outcome
- In this analysis, access to Cochrane's SLR did not demonstrate any influence on the quality of the study selection process. Regardless of the pre-existing awareness of the SLR's results, no studies were omitted due to analyst error
- Using 3 different approaches to RRs resulted in 84, 89 and 100% of efficacy:
    - Surprisingly, the 100% corresponded to the methodology requiring the least human resources, related to work reduction by 18 analyst work days, and based on searching of only one database
    - The most complex approach to RR, including searches of several sources and the AI tool support, resulted in 89% of efficacy and 11 work days saved
- In total, 6 out of 56 trials have been lost as a result of modified methodology in all 3 RRs: 3 in RR_1 and 3 in RR_2. However:
    - Among those six studies, five have been assessed as having high RoB, and for 1 the RoB was not clear

- Four studies were not available at all in MEDLINE and Embase, and two were indexed inappropriately
- Missed studies were included by the Cochrane group in meta-analyses for 28 outcomes in total, however, their lack in analyses run for RRs impacted only 4 meta-analyses

- **The research shows the simplified methodology can be considered even for reviews in the clinical area, also when the objective is not only qualitative, but also quantitative analysis**
- **The findings highlight the importance of relevant indexing of records in databases, as well as using by authors of studies relevant keywords in titles and abstracts; researchers working on the review do not have impact on any of the two aforemented points, however, they can reduce mistakes being results of those shortcomings by using filters comprehensive enough to capture all important studies, and, at the same time, specific enough to make running LR possible within reasonable timelines.**

4.2. Introduction

Since 1993, a multitude of SLRs have been created to enlighten clinical practices and provide backing for healthcare decision-makers and HTA agencies on a global scale[6]. The methodological rigor and transparency intrinsic to SLRs establish them as the preeminent form of LR, providing an objective and comprehensive overview of evidence relevant to a given subject[58,19, 20]. Nevertheless, the execution of SLRs often necessitates substantial temporal and human resource investments.

Consequently, RRs have emerged as a complementary approach to address the time and resource constraints associated with SLRs. However, despite their utility, RRs encounter limitations, notably the lack of standardized guidelines delineating methodologies customized to address specific research objectives[61].

The Cochrane Rapid Reviews Methods group examined more than 300 RRs and, from the diverse methodologies employed across these studies, formulated a definition for RR. However, this definition is broad and lacks specificity[24].

Tricco et al.[64] first run an international survey and scoping review to get information on the different approaches to the RRs conducting, based on which, authors employed a modified Delphi method including inputs from 113 stakeholders to search for the most optimized approach. The method that includes searching for published literature exclusively, searching multiple databases while restricting by date and language, adopting single study selection, conducting double data extraction and quality assessment, was

identified as the most feasible approach with potentially the lowest risk of bias. Conversely, the most comprehensive RRs may be achieved by searching multiple databases and including grey literature, applying date restrictions only, and having a single reviewer responsible for screening, data extraction, and risk of bias assessment.

The main objective of this work was to create and test three different approaches to RRs and compare their results with SLRs done by Cochrane. The modified approaches to data sources selection, screening of title and abstracts, and full texts review were tested. Moreover, in one of the methodologies, the AI ranking of references was used aiming to evaluate its impact on time efficiency.

4.3. Methods

The objective of this study was to create and test three different RRs' methodologies and check to what extent the results overlap with the ones from SLRs with exactly the same scopes, conducted by the Cochrane group. The results of RRs' literature searches and the studies selection process were compared to the ones from SLRs. Then, the results of meta-analyses run for the studies identified in RRs, were compared with the ones conducted for SLRs, and the impact of not capturing some of studies by RRs was assessed.

Importantly, given the absence of access to full extractions and quality assessments conducted by Cochrane's reviewers, a fundamental assumption in this analysis was that performing these steps at an equivalent level of quality would require a similar amount of time from any experienced analyst. Consequently, these stages of the review were not tested in further analysis.

When creating RRs' methodologies to be applied, different approaches were selected for searches running, studies selection, usage of supporting software and AI tools, and analysts' awareness of the expected outcome (presented in Figure 9). Scopes of RRs were as follows:

- To evaluate the effects of fluoxetine for overweight or obese adults (RR_1)[66]
- To assess the effects of melatonin on preoperative and postoperative anxiety compared to placebo or benzodiazepines (RR_2)[67]
- To evaluate the effects of immunonutrition compared to standard non-immunonutrition formula feeding on mechanically ventilated adults (aged 18 years or older) with acute respiratory distress syndrome (ARDS) (RR_3)[68].

The approaches to RR's methodologies were created based on: self-experience, knowledge of databases specifics, and Tricco et al.[64] analyses. Based on the last one, the most feasible approach to RR with potentially lowest risk of bias includes searching for only published data in several databases, and adopting single study selection. Based on that, searching of many published and unpublished sources used by Cochrane, was compared to searching of two main databases (MEDLINE and Embase) together with main clinical trials registry, and versus searching of only one of the two main databases –

MEDLINE. Double studies selection, characteristic for SLRs, was compared with single one, as well as with double titles and abstracts screening (which based on self-experience is more prone to errors) combined with single full-text analysis (less prone to errors). Finally, in one of RRs, the AI tool was used. Different methodologies were tested against different SLRs. The objective of this analysis was not to compare RRs methodologies between them, but to assess the modified methodology versus standard systematic approach used by Cochrane (please note that analyses comparing different methodologies tested on the same SLR have been conducted as well and will be published soon, however, are not part of this PhD thesis).

In RR_1 only MEDLINE database was searched, omitting other databases searched by Cochrane's authors: Embase, CENTRAL, Latin American and Caribbean Health Science Information database (LILACS), ClinicalTrials.gov and WHO International Clinical Trials Registry Platform (ICTRP). There was no geographical or language restriction used, and search was run in October 2022; studies published after 2018 were excluded, since search strategies for SLR were run in December 2018. The check how awareness of review's results can impact work on RR, analysts got access to Cochrane's SLR[66] before the screening phase. Then, a double screening of titles and abstracts, and single review of full texts were done in MS Excel with macros embedded to facilitate the analysis. Disagreements were solved by consensus or by a third reviewer. The used selection criteria were the same as in Cochrane's SLR[66]: reviewers were including only RCTs comparing the administration of fluoxetine vs placebo, other anti-obesity agents, non-pharmacological therapy or no treatment in overweight or obese adults. Studies on people with depression, mental illness or abnormal eating patterns were out of scope.

In RR_2, the used approach was more complex in comparison to RR_1. Two most recognizable databases, MEDLINE and Embase, as well as ClinicalTrials.gov (here, only studies with published results were considered) were searched,. The following data sources, searched by the Cochrane group, were omitted: CENTRAL, Cumulative Index to Nursing and Allied Health Literature (CINAHL), Web of Science and the WHO ICTRP. The screening of reference lists from the eligible reviews and trials was omitted. Searches were conducted in October 2022; studies published after 2021 were excluded to be consistent with the SLR. No geographical and language restrictions were applied. To minimize the risk of being biased by recognizing some of the studies included to SLR, analysts were not provided with access to Cochrane's review[67].

As in RR_1, a double screening of titles and abstracts and single review of full texts were conducted, but this time EPPI-Reviewer 4 software[34] was used to simplify the process. Titles and abstracts screening was conducted with use of the AI tool called priority screening. The tool is learning as the user is making decisions, and prioritizes publications that might be relevant[34]. Again, disagreements between analysts were solved by consensus or by a third reviewer. The used PICOS criteria were in line with the ones used by the Cochrane group[67]: RCTs or standard treatment-controlled studies, assessing the effects of preoperatively administered melatonin on preoperative or postoperative anxiety, were included. Adult patients undergoing any type of surgery for which it was needed to use general, regional, or topical anaesthesia, were the population of interest.

Finally, in RR_3, only MEDLINE database was searched, omitting all of other steps done by the Cochrane group, i.e.: searching of Embase, CENTRAL, WHO ICTRP and ClinicalTrials.gov, reviewing reference lists of included reviews and trials, and manually searching for relevant citations from previously published SLRs, studies, and conference proceedings. The same ARDS definition and selection criteria as described in Cochrane's SLR were adapted: RCTs and quasi-randomized controlled trials comparing immunonutrition versus a control or placebo nutritional formula in the target population were included. Records published after 2019 were excluded to be in line with Cochrane's authors, and no language restrictions were applied. Analysts did not have access to the Cochrane's SLR to avoid being biased[68]. Single screening, as well as review of full texts, were conducted, both in EPPI-reviewer software[34]. Uncertainties were solved by a senior reviewer.

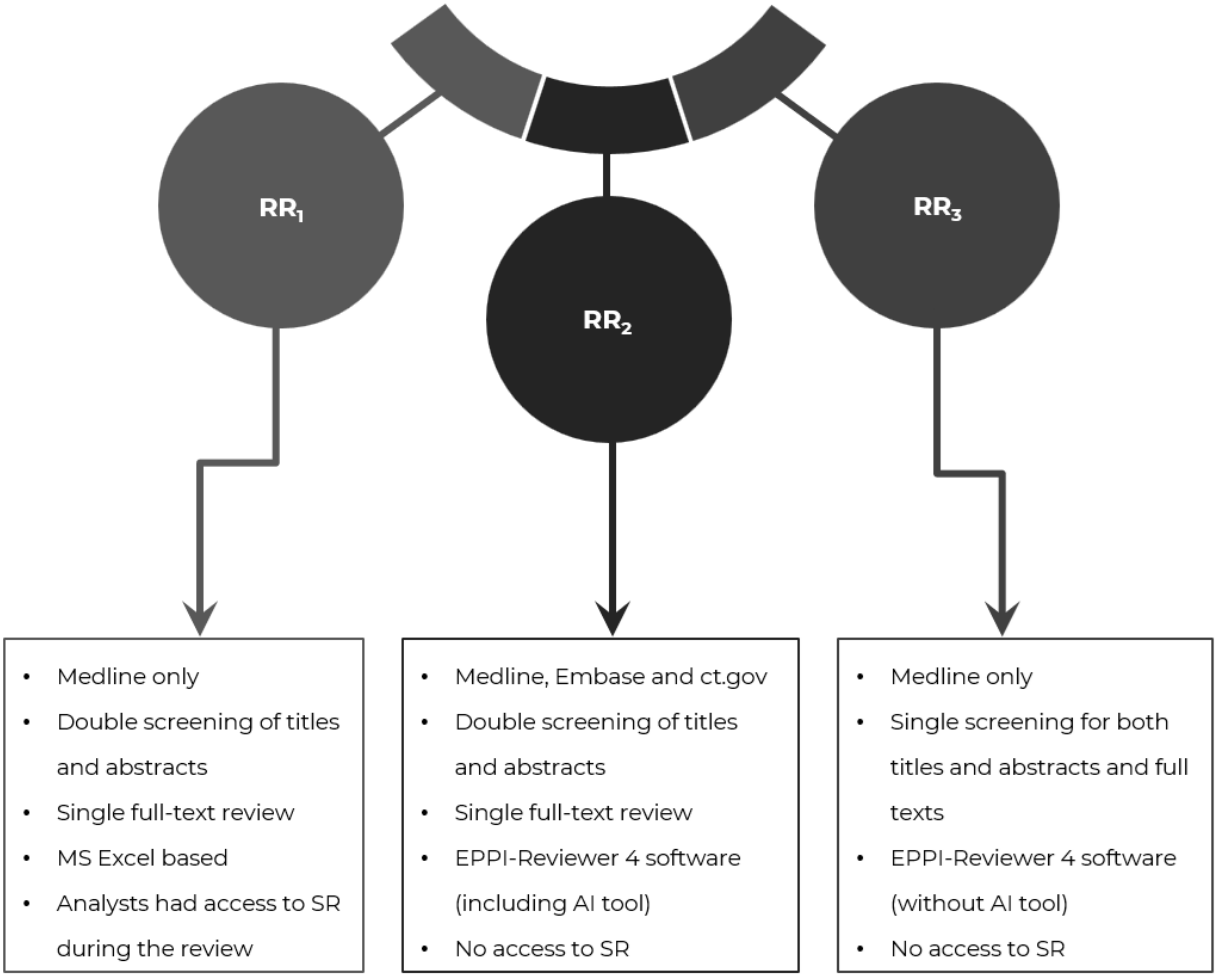


Figure 9 Various methodological approaches to RRs

Abbreviations: ct.gov – clinicaltrials.gov; MS – Microsoft; RR – rapid review; SLR – systematic review

Once the studies selection was completed, the lists of included records were compared against the ones included in SLRs and efficacy, defined as ratio of those two numbers, was

counted. When considering the workload reduction, the difference in the number of hits to be screened, as well as of full articles to be analyzed, were considered and calculated to analysts' workdays saved. Calculations were based on the average rate of analysts work in RRs, which was around 400 hits screened and 40 articles analyzed per day.
Quality assessment of all studies: included in RRs, as well as the ones missing in RRs but captured by SLRs, was taken from Cochrane's review – Cochrane's authors based it on the RoB assessment conducted according to criteria presented in the Cochrane Handbook for Systematic Reviews of Interventions[3, 66, 67, 68]. Each study was being evaluated across various domains as having low, uncertain, or high RoB. For the purpose of the presented research, overall RoB was set to (1) low when all domains had low RoB, (2) uncertain, if at least one domain had uncertain RoB, or (3) high, if at least one domain had high RoB.
The last step of this analysis was meta-analyses running and comparing results with those of the Cochrane group. Assuming that not all of the studies identified in SLRs would be also captured in RRs, the plan was to run meta-analyses for the same outcomes as considered in Cochrane's reviews, but with using only the studies identified during RRs, and to check the impact on the final results.
When analyzing the results, the quality of studies not captured by RRs was also considered.
Lack of inclusion of the specific studies on meta-analysis was considered as significantly impacting the results when it altered the direction of the findings, affected statistical significance, or made it impossible to conduct meta-analysis for the specific outcome.
Meta-analyses were run according to the inverse variance method with 'DerSimonian-Laird' algorithm to estimate the heterogeneity between studies[69]. For continuous outcomes the input data were number of observations, mean and standard deviation in experimental and control groups; for binary outcomes the input data included number of events and observations in experimental and control treatment arms. As the output, Weighted Mean Difference or Risk Ratio together with confidence interval were generated. All computations and forest plots were created by using 'package in R'[70].

4.4. Results

When running RRs, the inclusion lists were recreated with 84, 89 and 100% of efficacy in comparison to SLRs (Table 5). The reduction in the number of abstracts to screen was 54, 56 and 88%, and for full texts 2, 54 and 58%; those reductions in workload corresponded to around 3, 11 and 18 workdays saved.

Table 5 Impact of RRs' methodologies on efficacy and workload

REVIEW	EFFICACY (%)	REDUCTION: REFERENCES SCREENING (%)	REDUCTION: FULL-TEXTS REVIEW	IMPACT ON WORKLOAD (ANALYST'S WORK DAYS SAVED)*
RR1	84%	56%	2%	Around 3
RR2	89%	54%	54%	Around 11
RR3	100%	88%	58%	Around 18

*Analyst's work days saved were estimated based on analysts work rate in all 3 RRs

4.4.1. RR_1

In the Cochrane's SLR 1,036 records were screened, and 19 RCTs finally included[66]. Among the identified studies, eleven had high RoB, while for eight the RoB was unclear. When running the RR_1, 453 titles and abstracts were analyzed, and 55 full texts were reviewed in full. Sixteen RCTs were captured and included, all covered by the SLR (Figure 10).

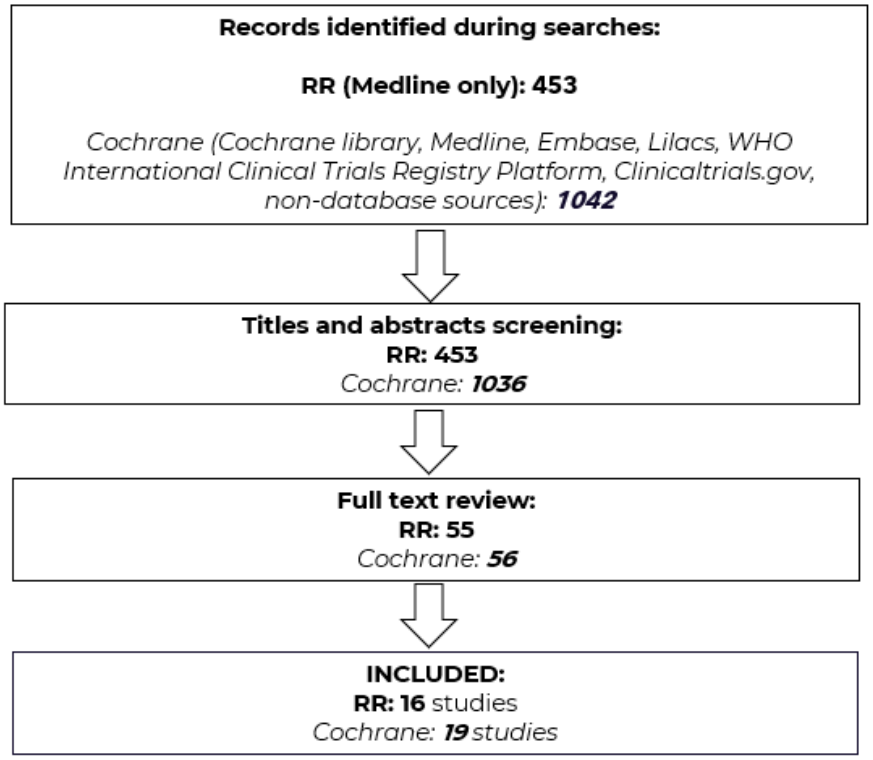


Figure 10 PRISMA chart RR_1

Three papers were not captured by RR: Levine 1987[71], Wurtman 1993[72], and Pedrinola 1993[73]. All 3 of them were assessed as having high RoB[66]. Analysis of reasons for missing the studies was conducted and it revealed that neither Levine 1987[71] nor Wurtman 1993[72] were indexed as RCTs in MEDLINE (no relevant keywords were used in title, abstract, or any other field), therefore, the strategy used in RR could not identify them. The third study, Pedrinola 1993[73] was not available in MEDLINE database at all.

Those three papers were included in 23 separate meta-analyses run by Cochrane[66]: Levine 1987[71] was included in 19, Wurtman 1993[72] in 14, and Pedrinola 1993[73] in 3 meta-analyses (Table 6).

Table 6 RR_1: summary of meta-analyses' results and studies excluded from analyses

Outcome	SLR's results	RR's results	Missing study/ies
Weight loss (end of trial weight; analysis 1.1.1, fluoxetine dose of 60 mg/d)	-2.50 [-3.78; -1.22]	-2.28 [-3.94; -0.61]	Levine 1987[71]

Outcome	SLR's results	RR's results	Missing study/ies
Weight loss (end of trial weight; analysis 1.1.2, fluoxetine dose of 40 mg/d)	**-3.97 [-8.75; 0.81]**	**NA**	**Pedrinola 1993[73]**
Weight loss (end of trial weight; analysis 1.1.3, fluoxetine dose of 20 mg/d)	-1.5 [-3.53; 0.54]	-0.65 [-2.53; 1.24]	Wurtman 1993[72]
Weight loss (end of trial weight; analysis 1.1, total score)**	-2.46 [-3.46; -1.47]	-1.82 [-2.84; -0.79]	Wurtman 1993[72], Pedrinola 1993[73], Levine 1987[71]
Weight loss (duration of intervention 0-3 months; analysis 1.2)	-3.34 [-3.93, -2.76]	-3.46 [-4.17; -2.74]	Levine 1987[71]
Weight loss (duration of intervention; analysis 1.2, total score)	-2.72 [-3.73, -1.71]	-2.65 [-3.89; -1.40]	Levine 1987[71]
Any adverse event (analysis 1.4)	1.18 [0.99, 1.42]	1.19 [0.95, 1.48]	Wurtman 1993[72], Levine 1987[71]
Specific adverse event (abdominal pain; analysis 1.5.1)	1.51 [0.58, 3.90]	2.10 [0.74, 6.01]	Levine 1987[71]
Specific adverse event (anxiety; analysis 1.5.5)	1.07 [0.56, 2.03]	1.19 [0.52, 2.73]	Wurtman 1993[72], Levine 1987[71]
Specific adverse event (constipation; analysis 1.5.6)	2.83 [0.58, 13.90]	2.83 [0.58, 13.90]	Wurtman 1993[72]*
Specific adverse event (diarrhoea; analysis 1.5.7)	**1.44 [0.97, 2.13]**	**1.54 [1.04, 2.30]**	**Wurtman 1993[72], Levine 1987[71]**
Specific adverse event (dizziness; analysis 1.5.8)	**2.40 [1.03, 5.60]**	**2.12 [0.88, 5.15]**	**Levine 1987[71]**
Specific adverse event (drowsiness; analysis 1.5.9)	2.67 [1.68, 4.24]	2.52 [1.56, 4.07]	Wurtman 1993[72], Levine 1987[71]
Specific adverse event (dry mouth; analysis 1.5.10)	1.23 [0.66, 2.30]	1.48 [0.74, 2.98]	Wurtman 1993[72]
Specific adverse event (dyspepsia; analysis 1.5.11)	1.99 [0.71, 5.55]	1.40 [0.33, 5.88]	Wurtman 1993[72], Levine 1987[71]
Specific adverse event (fatique; analysis 1.5.12)	2.50 [1.62, 3.85]	2.49 [1.65, 3.75]	Wurtman 1993[72], Levine 1987[71]
Specific adverse event (headache; analysis 1.5.13)	1.17 [0.94, 1.47]	1.24 [0.98, 1.57]	Wurtman 1993[72], Levine 1987[71]
Specific adverse event (insomnia; analysis 1.5.14)	2.23 [1.22, 4.08]	1.79 [1.11, 2.89]	Wurtman 1993[72], Levine 1987[71]
Specific adverse event (irritability; analysis 1.5.15)	1.41 [0.63, 3.15]	1.14 [0.61, 2.13]	Levine 1987[71]
Specific adverse event (nausea; analysis 1.5.17)	1.99 [1.35, 2.91]	1.98 [1.28, 3.06]	Wurtman 1993[72], Levine 1987[71]
Participants with any adverse event (per dose; analysis 1.6.1, fluoxetine dose of 60 mg/d)	1.16 [0.93, 1.44]	1.09 [0.91, 1.32]	Levine 1987[71]
Participants with any adverse event (per dose; analysis 1.6.3, fluoxetine dose of 20 mg/d)	1.10 [0.92, 1.31]	1.16 [0.77, 1.75]	Wurtman 1993[72]
Participants with any adverse event (duration of intervention 0-3 months; analysis 1.7.1)	1.71 [0.87, 3.33]	1.60 [0.70, 3.68]	Levine 1987[71]

Outcome	SLR's results	RR's results	Missing study/ies
Participants discontinuing due to adverse events (analysis 1.8)	1.88 [0.87, 4.06]	1.33 [0.60, 2.92]	Wurtman 1993[72], Pedrinola 1993[73], Levine 1987[71]
BMI (analysis 1.9, total score)	-1.11 [-3.66, 1.43]	-0.99 [-4.25, 2.26]	Pedrinola 1993[73]
Morbidity/depression (analysis 1.10)	1.20 [0.57, 2.52]	1.03 [0.47, 2.26]	Levine 1987[71]
Morbidity/depression (per dose, analysis 1.11.1, fluoxetine dose of 60 mg/d)	**1.41 [0.53, 3.79]**	**NA**	**Levine 1987[71]**
Morbidity/depression (per dose, analysis 1.11, total score)	0.94 [0.52, 1.73]	0.84 [0.46, 1.55]	Levine 1987[71]

**Please note : Total score for analysis A was not calculated by Cochrane, results was counted additionally by authors of this work

In case of main meta-analysis, comparing fluoxetine versus placebo on weight loss (end of trial weight), the absence of Pedrinola 1993[73] made it impossible to run meta-analysis for fluoxetine of dosage 40 mg/d, as it was one of two studies analysing this dose of the drug (Table 6 and Figure 11). When analyzing dosage of 20 mg/d, lack of study by Wurtman et al. [72] in RR's meta-analysis did not impact the final outcome, as well as the lack of study Levine 1987[71] in meta-analysis for fluoxetine 60 mg/d (Table 6 and Figure 11). Total score for analysis of weight loss (end of trial weight) was not calculated by the Cochrane group; however, the additional calculations run in this research showed that the lack of all three studies did not have a significant effect on the final meta-analysis's result (mean difference: -2.46 [-3.46; -1.47] vs -1.82 [-2.84; -0.79], Table 6 and Figure 11).

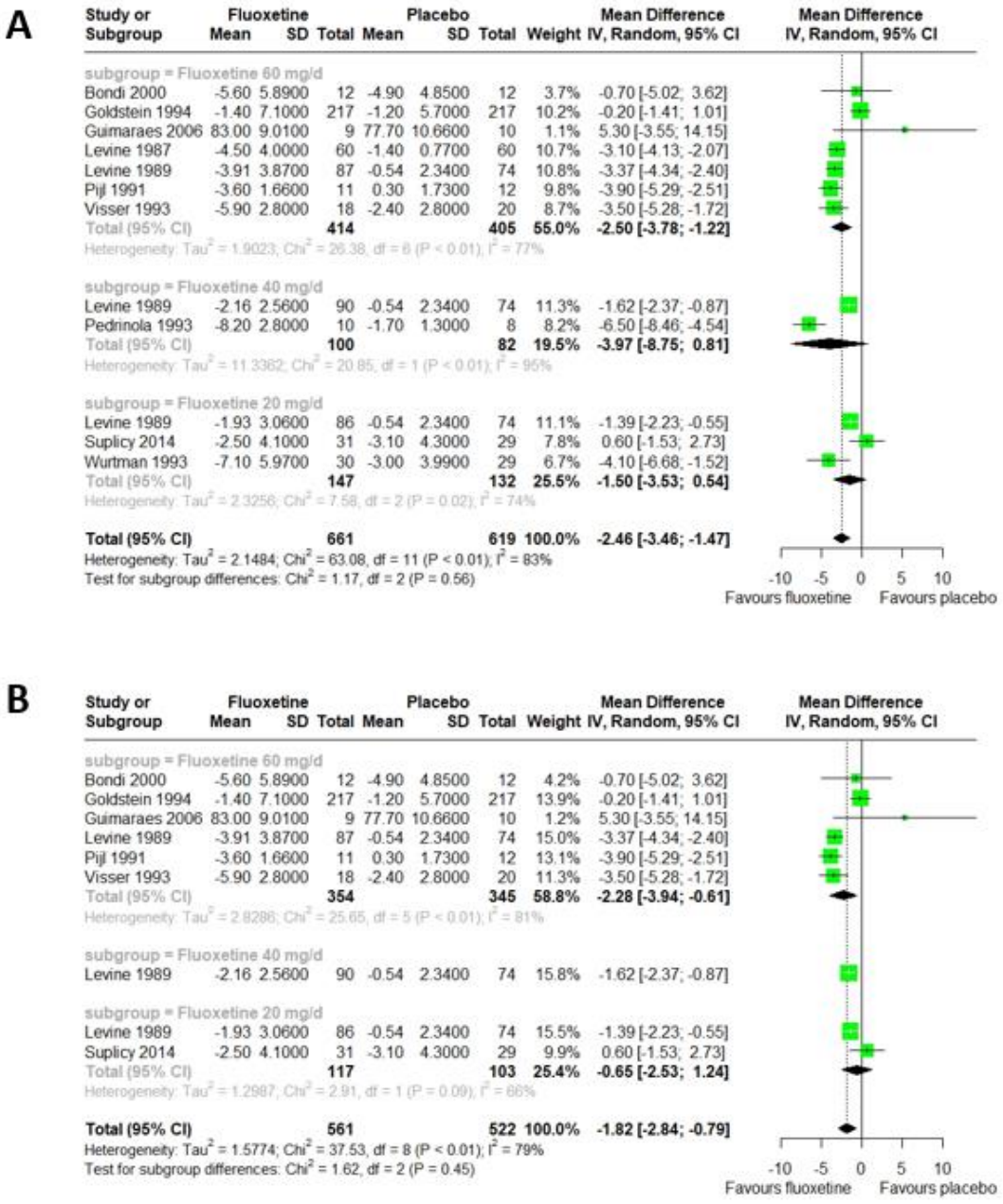


Figure 11 Forest plots comparing fluoxetine versus placebo on weight loss (end of trial weight – analysis 1.1) – with all data (A) and without the missing studies (B). Please note : Total score for analysis A was not calculated by Cochrane, results was counted additionally by authors of this work

In meta-analysis for diarrhoea, the lack of study by Levine et al.[71] resulted in change of the statistical significance of the final outcome (analysis 1.5.7, Table 6 and Figure 12). Levine 1987 was one of two studies shifting the result in direction of fluoxetine, while the final outcome of meta-analysis suggested the opposite direction. The final outcome was close to being statistically significant (risk ratio 1.44 [CI 0.97, 2.13]), therefore, the lack of Levine 1987 shifted result further towards favouring placebo making outcome statistically significant (risk ratio 1.54 [Cl 1.04, 2.30]). Since the outcome was not estimable in Wurtman 1993 study[72], the lack of this records did not influence the meta-analysis (Table 6 and Figure 12).

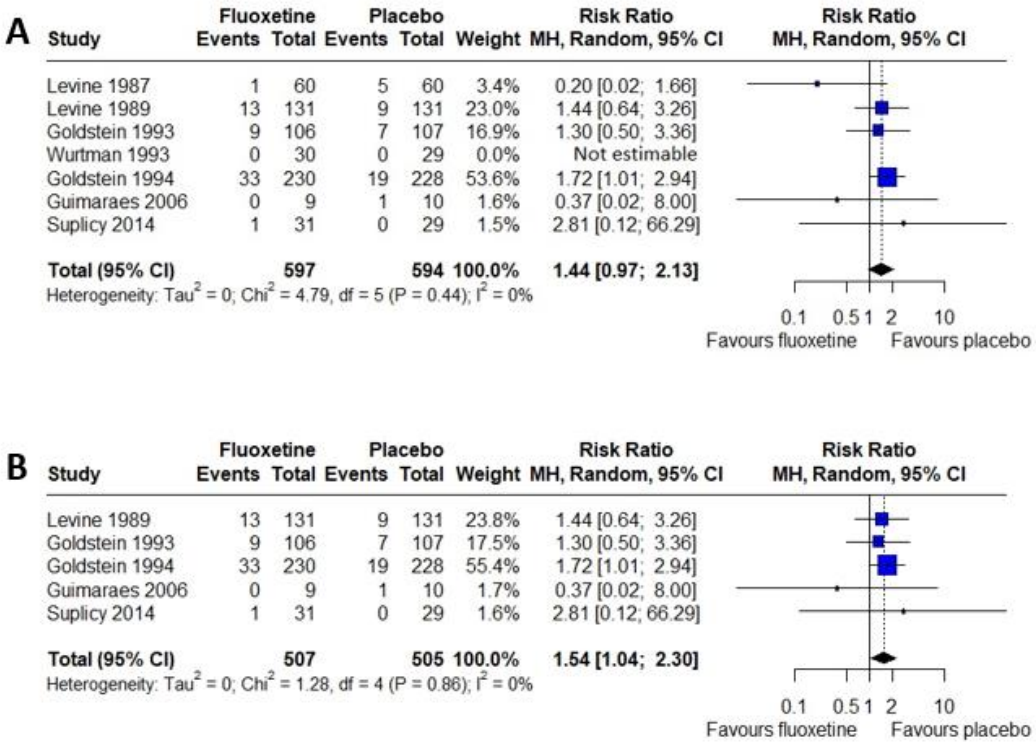


Figure 12 Forest plot comparing fluoxetine versus placebo on outcome 1.5.7 (specific adverse events – diarrhoea) – with all data (A) and without the missing studies (B)

The lack of the study conducted by Levine et al.[71] affected results of the meta-analysis for dizziness (analysis 1.5.8; Table 6 and Figure 13). Removing the study shifted the Cl of the final outcome, making it statistically insignificant, with risk ratio of 2.12 [Cl 0.88, 5.15], in comparison to the Cochrane's results (2.40 [Cl 1.03, 5.60]).

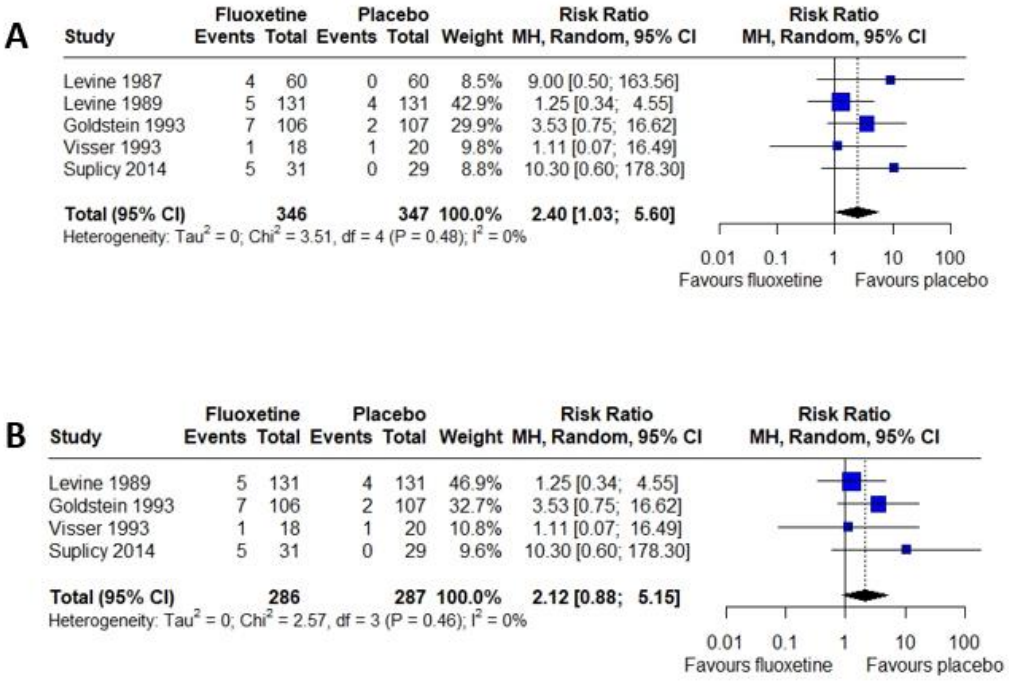


Figure 13 Forest plot comparing fluoxetine versus placebo on outcome 1.5.8 (specific adverse events – dizziness) – with all data (A) and without the missing studies (B)

The three abovementioned studies were also used in Cochrane's meta-analyses for specific adverse events: abdominal pain[71], anxiety[71, 72], constipation[72], drowsiness[71, 72], dry mouth[72], dyspepsia[71, 72], fatigue[71, 72], headache[71, 72], insomnia[71, 72], irritability[71], and nausea[71, 72], however, their absence did not affect the results of RR's meta-analyses in a significant way.

Additionally, they were considered in Cochrane's meta-analyses for the following outcomes: weight loss (duration of intervention)[71], any adverse events[71, 72], number of participants with any adverse events (per dose)[71, 72] and duration of intervention[71, 72], discontinuation due to adverse events[71-73], Body Mass Index (BMI) change[73] and morbidity (depression and depression per dose)[71]. For those outcomes absence of the studies did not impact the final results of our analyses in comparison to Cochrane's ones, except one case; for the morbidity/depression per dose outcome, it was not possible to run meta-analysis for fluoxetine dosage of 60 mg/d, as Levine 1987[71] was one of two studies analysing drug in this dosage. However, it has not impacted the final outcome of the meta-analysis total score (0.94 [0.52, 1.73] in Cochrane review vs. 0.84 [0.46, 1.55] in the RR's analysis, Table 6).

4.4.2. RR_2

In the second SLR, the Cochrane group screened 3,519 records and finally included 27 studies[67]. Melatonin was compared to placebo in 24 studies, while 11 studies compared melatonin to benzodiazepines. In a small number of studies, the other comparator drugs were gabapentin, clonidine and pregabalin. Among the identified studies, none had

overall low RoB, 19 were rated as having an unclear RoB, and 8 as having a high RoB. When running the RR_2, 1,604 titles and abstracts were analyzed, and 44 papers were reviewed in full. Twenty-four RCTs were included, all covered by the SLR (Figure 14).

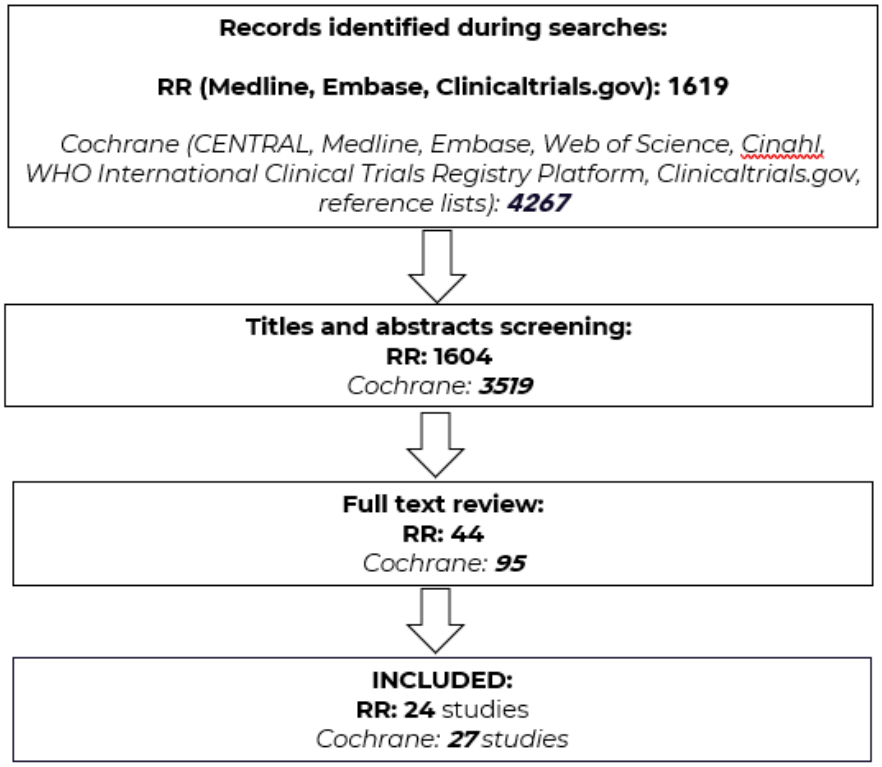


Figure 14 PRISMA chart RR_2

The analysis conducted for three studies not captured by the RR_2: Jain 2019[74], Marzban 2016[75], and Hosseini 2015[76], showed that none of these studies were available in MEDLINE nor Embase databases.

Studies done by Marzban et al.[75] and Hosseini et al.[76] were of low quality with at least one domain assessed as having high RoB. The remaining one, Jain 2019[74], was assessed as having an unclear RoB.

Hosseini 2015 study[76] was not considered in any of the Cochrane's meta-analyses, so no capturing it in RR did not have any impact on final conclusions.

Jain 2019[74] was included by Cochrane authors in meta-analysis for preoperative anxiety, measured by visual analogue scale (VAS). Lack of this study in RR showed no significant impact on the final meta-analysis's outcome, as the study results, favouring melatonin versus placebo, were in line with majority of other studies included in the analysis (-11.69 [Cl -13.80, -9.59] in SLR vs -11.69 [-13.91, -9.47] in RR; Table 7).

Study Marzban 2016[75] was included by the Cochrane group in two meta-analyses: the effect of melatonin versus benzodiazepine on pre- and postoperative anxiety using the VAS scale (analyses 2.1 and 2.2, respectively; Table 7). In case of preoperative anxiety, results in Marzban 2016 were in line with the final outcome of the analysis, slightly in favour of benzodiazepine, and not statistically significant. Four out of six other included

studies also favoured benzodiazepine, and one result was statistically significant, when results of the two studies in favour of melatonin were not statistically significant. As an effect, the lack of the study in the RR's meta-analysis did not influence significantly the final outcome. For postoperative anxiety, only Marzban 2016 reported final VAS scores. Its results were consistent with the ones reported by other included studies presenting change in VAS scores, and therefore, missing the study in RR_2 did not change the final conclusions of the analysis (-2.02 [Cl -5.82, 1.77] in SLR vs -2.12 [Cl -4.61, 0.36] in RR, Table 7).

Table 7 RR_2: summary of meta-analyses' results and studies excluded from analyses

Outcome	SLR's results	RR's results	Missing study
Melatonin vs placebo, preoperative anxiety (VAS, mm, final VAS scores, analysis 1.1.1)	-11.58 [-14.08, -9.08]	-11.57 [-14.23, -8.90]	Jain 2019[74]
Melatonin vs placebo, preoperative anxiety (VAS, mm, analysis 1.1, total score)	-11.69 [-13.80, -9.59]	-11.69 [-13.91, -9.47]	Jain 2019[74]
Melatonin vs benzodiazepine, preoperative anxiety (VAS, mm, final VAS scores, analysis 2.1.1)	0.68 [-2.54, 3.91]	-0.95 [-7.97, 6.07]	Marzban 2016[75]
Melatonin vs benzodiazepine, preoperative anxiety (VAS, mm, analysis 2.1, total score)	0.78 [-2.02, 3.58]	0.80 [-2.86, 4.46]	Marzban 2016[75]
Melatonin vs benzodiazepine, postoperative anxiety (VAS, mm, analysis 2.2, total score)	-2.12 [-4.61, 0.36]	-2.02 [-5.82, 1.78]	Marzban 2016[75]

4.4.3. RR_3

In the last SLR, the Cochrane authors reviewed 3,367 records and identified 10 studies meeting the PICOS criteria[67]. The included studies were heterogenous in many aspects, including type and duration of interventions and reported outcomes[68].

Using the third approach to RR's methodology, 409 abstracts were generated and screened, resulting in 55 full texts being reviewed and 10 RCTs finally included (Figure 15). The list of the selected studies overlap fully with the Cochrane's SLR. All the ten studies were included in quantitative analysis by the Cochrane group and, potential differences in data synthesis aside, the same conclusions would have been done based on RR_3's results, despite applying the simplified methodology.

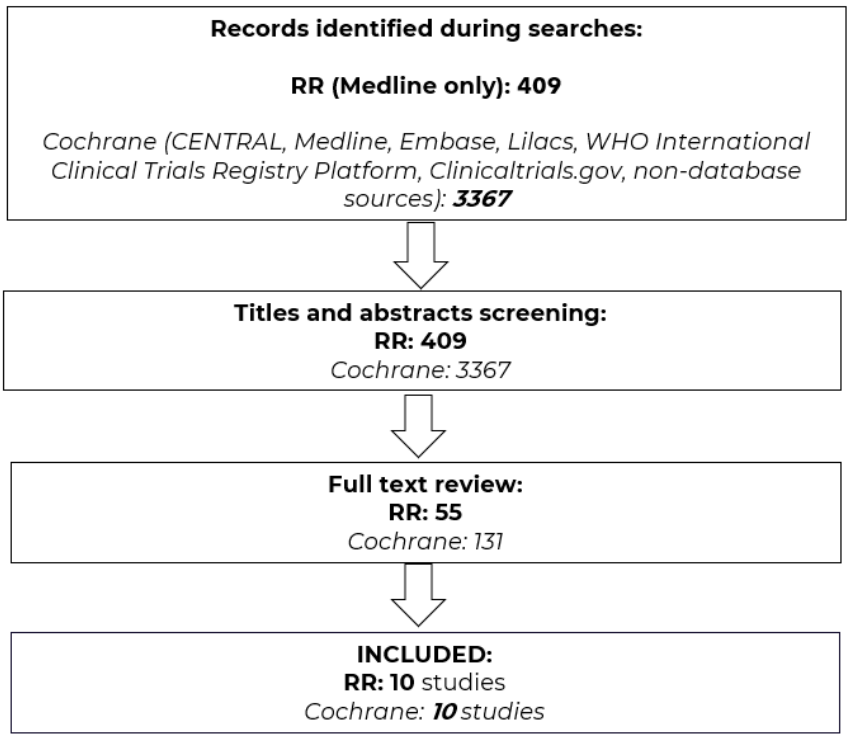


Figure 15 PRISMA chart RR_3

4.5. Discussion

SLRs are the most complex form of reviews[58]. To conduct them, researchers follow a specific rules which makes review process human resources demanding and time-consuming[3]. Taking into account the number of newly published studies, and an increasing need for fast information analysis, RR seems to be a suitable alternative and a practical solution for evidence synthesis[61]. Despite abovementioned facts, there is still no consensus on the best methodology to be used when performing RRs. The three different approaches to RRs' methodology were created, tested, and compared. Methodologies differed between each other in terms of: data source searched, single/double studies selection, software support and AI tool usage. The impact of the simplified methodologies on workload reduction, as well as the impact on final results on meta-analyses, were assessed in comparison to SLRs performed by the Cochrane group.

It has been showed that using 3 different approaches to RRs resulted in 84, 89 and 100% of efficacy (defined as the number of studies captured during the RR divided by the number of studies identified in SLR), and surprisingly, the 100% corresponded to the methodology requiring the least human resources, related to work reduction by 18 analyst workdays, and based on searching of only one database. The most complex approach to RR, including searches of several sources and the AI tool support, resulted in 89% of efficacy and 11 workdays saved, which seems to be still a very good outcome.

In total, 6 out of 56 trials have been lost as a result of modified methodology in all 3 RRs: 3 in RR_1 and 3 in RR_2. However, the three points that should be considered are:

- among those six studies, five have been assessed as having high RoB, and for 1 the RoB was not clear

- four studies were not available in main databases (MEDLINE and Embase) and two were indexed inappropriately
- missed studies were included by the Cochrane group in 28 meta-analyses in total, however, their lack in analyses run based on RRs' results impacted in a significant way only 4 of them.

Those observations shows that the simplified methodology can be considered even for reviews in the clinical area when the objective is not only qualitative, but also quantitative analysis, however, factors such as disease area, date when the first studies meeting the PICOS criteria have been published, and the complexity of the topic, should be considered as those are factors impacting the final outcomes. The findings also highlight there is the necessity of relevant indexing of records in databases, as well as using by authors of studies relevant keywords in titles and abstracts. Researchers working on the review do not have impact on indexing nor using relevant keywords in abstracts of already published manuscripts, however, they can reduce mistakes being results of those shortcomings by using filters comprehensive enough to capture all important studies, and, at the same time, specific enough to make running LR possible within reasonable timelines. In this analysis, access to Cochrane's SLR did not demonstrate any influence on the quality of the study selection process. Regardless of the pre-existing awareness of the SLR's results, no studies were omitted due to analyst error.

The analysis has several strengths and limitations. Firstly, it was conducted based on a predefined methodology with clear approach to data source, selection process, assumptions regarding the further phases of the review, and quality assessment of studies. Additionally, it was clearly defined when the effect of exclusion from meta-analyses studies not identified during RRs was considered as significant. When running RRs, an exactly the same search strategies as the ones used by Cochrane group were run, and only those studies published by the date of searches used by SLRs' authors were considered.

However, some limitations arose despite the project' strengths. It was assumed that data extraction and quality assessment processes done by Cochrane's reviewers were of the highest quality and would be conducted within the same speed by analysts working on RRs. Those assumptions impacted calculations regarding workload reduction. Moreover, each of RRs was tested against different SLR. It make it impossible to compare RRs' methodologies directly between them. However, the objective of this analysis was to test the simplified methodologies again SLRs in general, to get an overview on methods possible and factors which can influence RRs processes. It should be treated as an introduction to further analyses and improvement of RRs methods. Indeed, this analysis was a basis and supported further ideas and projects. The research in which different RRs were compared versus the same SLR have been conducted, and will be published soon; however, it is not a part of this PhD thesis.

5. LRs in the era of new technologies

5.1. Key findings

The questions being answered by this section:

- **What is the current status of software supporting LRs available on the market?**
- **Which phases of the LRs can be supported by the AI?**
- **What is the future of new technologies in the LRs' area?**

- Software- and AI-support of LRs is a dynamically developing field
- 63 software tools supporting searching, screening, data extraction, critical appraisal, meta-analysis, reporting, reference management and visualizing of the results, have been identified, and information regarding them was collected and summarized
- 6 software tools supporting at least several phases of LRs have been tested in depth

- **The most commonly AI-supported stage of LRs is screening; the conducted research show the usage of the AI-supported tools during titles and abstracts reviewing can be a time-saving solution**
- **The most difficult step of LRs to be software- and AI-supported, is the data extraction process. It needs a lot of aspects to be considered, and it is highly probable that extraction of objective and some basic characteristics of a study, such as a sample size, will be supported in the future, but, based on the conducted research, there is a low probability for automatic extraction of more complex data**
- **Following an extensive analysis, it can be inferred that the promises of favorable outcomes made on manufacturers' websites regarding the tools may often be mere marketing strategies**
- **None of the software evaluated demonstrated satisfactory performance in terms of automated or semi-automated data extraction**
- **The costs of certain tools are elevated, despite their functionalities being comparable to those offered for free or at a lower price.**

5.2. Introduction

There is a growing interest in the use of artificial intelligence (AI) technologies in diverse disciplines, including health sciences[77]. One of the fields that could be supported by the technologies is the area of LRs. LRs, especially systematic ones, are characterized by a complex methodology that provides a transparent and reproducible process but at the same time is very time-consuming[78]. Automating the stages of the process could reduce workload, therefore making LRs more timely.

One of the models that could introduce AI technologies in the process of LRs is the large language model (LLM), which is pre-trained on large datasets in the form of text. It was implemented in ChatGPT proposed and built by OpenAI upon GPT-3 and GPT-4 engines, which are open-access and more advanced paid versions, respectively[79]. LLMs are multi-purpose, not focusing on any specific domain, but learning statistical patterns and structures of language to predict the following words[80] based on billions of parameters[81], which, unfortunately, can result in unintended actions, including fabricating facts or introducing bias[79, 82]. They also have an ability of dialogue, which resembles of human-like conversation[83]. In addition to these functions, LLMs have the ability to learn from the training data, so it can answer specialized questions. Several downsides related to LLMs' usage in the LRs area can be mentioned. Testing the ChatGPT in different LR stages proved that it is not an appropriate tool for this objective yet, i.e., it proposed an incorrect search strategy, developed a code for conducting a meta-analysis with multiple errors, and failed when synthesizing a small amount of data[84, 85]. However, it is worth noting that its' performance was appropriate in some of the other phases, e.g.: formulating a review questions, creating inclusion and exclusion criteria, and titles screening, nevertheless, using it requires expertise knowledge, since the responses are not identical each time a question is being asked[84]. Another issue related to getting prompts from the tools is that no specific approach for developing them exists, and the most suitable one should be tested empirically for each objective. Developing simpler models for specific objectives could be useful for the LR process, therefore, the field should be further analysed.

The objective of this research was to check what tools supporting the LRs process are currently available on the market, to test some of them, and to make conclusions on how the future of LRs looks like and for which steps of the reviews we can expect the highest progress in the software area during the following years.

5.3. Methods

To identify all the tools supporting the LR's process, an RR in PubMed was run by using the following combination of words: (machine learning OR computational tools OR artificial intelligence) AND (literature reviews), and restricting it by the filter for systematic reviews. The most up-to-date records were screened (search restricted to the last year and run on 05/06/2022 and 13/06/2023).

The interest was in SLRs summarizing the current knowledge status on the available tools. To answer the research question, the list of tools was extracted and per each of them data regarding stage supported, machine learning (ML) technique used and price, were collected. Articles focusing on tools supporting other domains of the health economic and outcomes research area or market access field were excluded as not of interest for this analysis.

For tools supporting more than one phase of LRs, information regarding keywords highlighting, PDFs searching, studies selection and data extraction support solutions were extracted. If not available directly from the RR, the additional search of the Internet was conducted, however, for some of the analysed tools no websites have been identified.

Additionally, desk research was run in Google to complement the list of the most recognizable tools supporting the full process of LRs.

For the selected tools an in-depth analysis has been conducted, focusing on how they can support the LRs process, including usage of the AI tools.

For the most promising ones, when no sufficient data were available, developers from supportive team and/or CEOs of the companies producing software, have been contacted. There was no specific survey conducted, support team and/or CEOs were contacted only when some information was not possible to identify/verify.

It is imperative to recognize that this summary was accurate as of June 2023, and it is plausible that tools and functionalities may have been updated by the time of this thesis publication. Furthermore, the analysis was conducted solely for academic purposes, and confidential commercial information was not disclosed by the companies. Hence, in case of any uncertainties, the analysis presented in the thesis should not be regarded as the definitive source of knowledge, and it is advisable to consult the software's manufacturers.

5.4. Results

5.4.1. Overall market analysis

In total, 1144 abstracts were reviewed, resulting in identifying one comprehensive review providing analysis and summary on numerous tools created to assist reviewers in the process of the LR[86]. The purpose of the paper was to map the computational tools that use ML techniques to assist SLR process. Authors performed comprehensive, tailored searches in several databases and software repositories. Any ML software to support SLR process that have been developed or updated within the last 10 years was included in the overview. In general, 63 tools were identified. For the identified tools, data regarding stage supported, ML technique used and price, were collected. The tools are provided as Web-based applications, desktop applications, or coding packages to use in specific programming tools, which usually require specialized knowledge to use them. Some of the identified tools provide assistance for multiple stages of LRs, and for them, a detailed analysis on how they can support the process, is presented in Table 19 (Appendix). However, most of the identified software solutions were dedicated to specific phases of the process, i.e.: search (Table 20), screening (Table 21), data extraction (Table 22),

critical appraisal (Table 23), meta-analysis (Table 24), reporting (Table 25), reference management (Table 26) and visualizing of the results (Table 27). Table 28 in the Appendix provides an overview of databases that register, collect and disseminate SLRs in order to support evidence-based healthcare.

The tools are developed in order to save reviewers time, yet using separate tools for different LRs' steps creates new time-consuming activities, such as combining the results or adjusting the formatting.

The most often AI-supported functionality is screening; the tools that use ML techniques to aid screening are: AbstracR, Active_learning_document_screening, Active-learning-for-systematic-review, ASReview, ASReview-covid, Cochrane RCT Classifier, Colandr, Concept Encoder, DAE-FF, DoCTER, FASTREAD, GAPscreener, Inclusion criteria, Machine Learning Functions, PICO Portal, PubmedClassifier, Pvtopic, RapidMiner, RAx, Rayyan, Research Screener, RevTools, RobotAnalyst, RobotSearch, rules_cochranereviews, Screen4Me, SLR_SearchStrings, SWIFT-Active Screener, SWIFT-Review, SyRF, Sysrev, Sys_review_ml and Systematic Review Accelerator. In one tool (CADIMA) text-mining capabilities are in development. Authors of the programmes use natural language processing (NLP), supervised learning (SL), semi-supervised learning (SSL) and unsupervised learning (USL) technologies, yet the most incorporated ones are SL (11 out of 33 tools) and NLP (8 tools) or both the technologies together (12 tools). More than a half of the listed tools are open-source[86].

Seventy-three of the identified tools were developed to support the searching process, and 32 of them are AI-supported, including 26 programmes dedicated only to this stage of the review (i.e., 2dsearch, Aggregator, BIBOT, Carrot2, Chilibot, Cochrane Register of Studies, Coremine Medical, Costumer, Doctor Evidence, Epistemonikos, FACTA+, Heoro, Leximancer, Linguamatics, Litesearchr, RCT Tagger, RoBotAnalyst, Sample Size Search (SSS) Tool for PubMed, SLR_SearchStrings, Swift-Review, TextRank, Thalia, Trial2rev, TrialStreamer, Voyant Tools and WordStat 9.0), and 6 supporting multiple phases of review (i.e., CADIMA, Colandr, RAx, SySRev, IRIS.AI and Systematic Review Accelerator). Majority of the tools are open-access (including 1 requiring Cochrane Account)[86].

There are 35 tools supporting data extraction process presented in the analyzed article[86], and in 9 of them AI technologies are implemented (NLP in 4 of them, SL in 2 of them, and NLP with SL combined in 3 of them). Six of the AI-driven programmes are intended to support only the data extraction process (A Keyword-Based Literature Review Data Generating Alhorithm, ExaCT, Raptor, Rnatlp, RoBotReviewer, rules_cochranreviews), and two were created to assist in several stages of LR (i.e., Colandr and RAx). Again, more than 50% of tools are freely available, however, one of them only with institutional access to Cochrane Library.

Critical appraisal process is being assisted by 10 of the identified tools overall[86], however, only 2 of them, i.e., RobotReviewer and Rax, are AI-supported (NLP and SL technologies are involved), and in CADIMA text-mining capabilities are in development. Half of the tools are open-source.

Among 20 tools developed to facilitate a meta-analysis process[86], 15 are open-access, but yet none of them is supported by AI.

Reporting phase of SLR is aided by 8 programmes, with 2 of them (RAx and SyRF) being AI-supported. Additionally, there are 13 tools facilitating reference management and 7 tools for figures development, however, neither rely on AI solutions[86].

5.4.2. Results of the selected tools' tests

After the desk research, the following multi-stages supporting tools have been selected for an in-depth analysis to check on how they can support the LRs process in reality: EPPI Reviewer 4[87], SysRev[88], Rayyan QCRI[89], SUMARI[90], Silvi.ai[91], Evidence Prime[92].

Table 8 presents the final results of the analysis. They are in line with what is presented in literature on the topic. The screening phase is the easiest step of the LRs to be supported with some AI solutions. The data extraction is a complex process requiring experience and relevant knowledge. Therefore, even if suggested on the producents' websites, at this moment there is no good AI-supported tool that could help and automatize the extraction process.

Table 8 Summary of the in-depth analysis of the selected tools

Tool name	**KWs highlighting**	**PDFs searching**	**AI support for screening**	**AI support for extraction**	**Comments** **Please note: It is crucial to acknowledge that this summary was current as of June 2023, and it is possible that tools and functionalities may have been updated by the time of this thesis publication. Moreover, the analysis was done for the academic purpose only and the confidential commercial information was not shared by the companies.** **Therefore, when there are any doubts, the analysis presented in the thesis should not be treated as the final source of knowledge and the software's producents should be consulted.**
EPPI Reviewer4[87]	-Yes	-Yes*	-Yes	-No	-AI supports titles and abstracts screening: there are four ways how ML can be used to make reviewing more efficient: -- Using pre-built classification for particular study types -- Building reviewer's own bespoke classifiers -- Using 'priority screening' to order records for manual screening -- Automatically clustering records based on the text they contain -Complex extractions possible, templates ready, and possibility to create new ones -Interface seems to be user-friendly; however, based on this assessment not a good tool for complex extractions -Easy uploading of references to the tool; filters adjusting records based on the source database available -Easy deduplication process -Supports creating report, especially tables with results, but also supporting PRISMA creation
SysRev[88]	-Not found	-No	-Yes	-No	-AI support for screening, by ranking, users are able to filter articles based on the model's predictions -Tool supporting extraction, but without AI functionalities -It seems to be possible to do more complex extractions, by adding group labels (not tested); possibility to download extraction results as csv files -By using API analyst can easily work on extracted data in RStudio -Possible to run some direct searches in PubMed, FDA and clinicaltrials.gov -Adding comments possible -Option for deduplication running not found -Possible to create labels - additional questions to analysts, so not only inclusion/exclusion decision can to be added, but also some more precise information; can be useful especially for junior reviewers, and when many factors need to be considered during studies selection -PDFs review is not just the following step of titles and abstracts screening, a new project needs to be created
Rayyan QCRI[89]	-No	-No	-Yes	-No	-No tool for PDFs downloading; analyst has to map each PDF to a specific reference, and then each PDF needs to be open separately -After 50 records are screened AI engine calculates the probability of inclusion for the undecided references -Tool supports only screening and full-text review, no good solution for conflicts resolving; no support of data extraction

Tool name	KWs highlighting	PDFs searching	AI support for screening	AI support for extraction	Comments **Please note: It is crucial to acknowledge that this summary was current as of June 2023, and it is possible that tools and functionalities may have been updated by the time of this thesis publication. Moreover, the analysis was done for the academic purpose only and the confidential commercial information was not shared by the companies.** **Therefore, when there are any doubts, the analysis presented in the thesis should not be treated as the final source of knowledge and the software's producents should be consulted.**
					-Standard way of references uploading, no support for searches in databases, not possible to edit or delete specific record, analyst can delete just the whole import; however, automatic deduplication process can be run by the tool, analyst can decide on duplicates -Management of work possible - analyst can directly send email to people who are assigned to screening; however, assigning specific ranges of references not possible, analyst will have to screen the ones for which there is no decision -To resolve conflict one of reviewers have to change decision -Possible to add label during screening, to choose reason of exclusion or type a specific new one -Review chat available for each ongoing review, making it possible to communicate some key messages to other people -For full-text review analyst should create a new project -Export of results in a format of Endnote or csv, need to be shared by email, cannot be downloaded directly
SUMARI[90]	-No	-No	-No	-No	-A very basic tool -Conflicts resolving only by changing decision of one of analysts -Not possible to do FTR in the software, not possible to save PDF -In-built citation management and automatic generation of appendices, making reporting easier
Silvi.ai[91]	-Yes	-Yes	-No	-No	-No AI supporting screening -No AI supporting extraction; just some data can be read from tables -Possibility to run simple automatic meta-analyses -Possibility to run searches by using tool in PubMed, no information what about other databases -Request for test account was done, however, no further contact from the company's side
Evidence Prime[92]	-Yes	-No	-Yes	-No	-Tool is suggesting studies to be included, it teaches itself constantly, based on each new reference -Probabilities for inclusion per each record are not available to analysts; however, this option can be added per user's request -Tool not very intuitive -Extraction templates can be created but only very simple ones, row after row, unintuitive - Interface semi user-friendly; some options were specific for the previous project; not sure which ones, not clear how to use it, additional support was needed -Models available for extraction relevant for experimental studies; by using them it was not possible to find data of interest in clinical trials. When used for experimental study, it worked for simple data like sex, species, strain, dose, and also for endpoints

*-searching for PDFs possible in Google, Google Scholar - but one by one, when analyzing abstracts, so not the fastest solution

API-application programming interfaces; NA-not available

5.5. Discussion

The general conclusions from the research show that the most commonly AI-supported stage of LRs is screening, where the tools assist the reviewer by suggesting, classifying, or selecting records. The conducted research show the usage of the AI-supported solutions during titles and abstracts reviewing can be a time-saving solution. Promising developments in ML technology are literature searching (47% of identified tools) and data extraction (17% of identified tools)[86]. However, based on the conducted in-depth analysis, it can be concluded that even when good outcomes achieved by the tools are promised on producents' websites, it can be just a marketing ploy. None of the tested software showed good outcomes when it comes to automatic or semi-automatic data extractions. Especially for this step, further collaboration and information sharing seems to be one of the crucial points needed for future development of the area. This observation is in line with other authors' opinions, caused by the fact that data extraction is the process that needs a lot of aspects to be considered, and it is highly probable that extraction of objective and some basic characteristics of a study, such as a sample size, will be supported in the future, but, in authors' opinion, there is a low probability for automatic extraction of complex data that can be interpreted in multiple ways, such as main conclusions[93].

Two additional points that should be considered by researchers making decision on the potential usage of the software are: to draw special attention when deciding for using a free programs (especially when working on commercial projects), as they can be open-source aiming data collection from different reviewers and building database to be used then for further tests, and therefore, may not be secured enough; and to ask for free trial period when considering paid software, as sometimes functionalities are comparable with the ones available for free or at a lower price.

Regarding how other researchers see the future of software supporting LRs, Wagner et al.[93] assessed the potential for AI support of each stage of the LR process separately, taking into consideration that the technologies constantly evolve. They anticipated moderate potential for AI-support in the first stage of review, which is problem formulation and identifying gaps in research, therefore, it is possible that AI will be able to point out promising research areas[93]. Authors mentioned already existing advances, including automated generation of hypotheses in biochemistry[94]. Very high potential was noted for the AI-support in the process of searching, since search methods are often repetitive[93], however, as mentioned before, ChatGPT did not perform adequately when it comes to a search strategy development[84]. Screening of titles and abstracts have a high potential to be efficiently semi-automated by the AI, since it relies on making repetitive decisions based on predefined inclusion and exclusion criteria; reviewers could benefit from such tool, since the task can be tedious and result in fatigue, which makes it difficult to make accurate decisions. As for full-text review, the potential is moderate, since it requires more professional judgement, as well as the quality assessment[93]. The last step, which is data analysis and interpretation, has high potential for use of AI when it comes to descriptive syntheses, however, a moderate in case of theory development and testing[93].

This analysis has several strengths and limitations. Firstly, to make sure the analysis of the tools available is comprehensive, an RR in PubMed and desk research in Google were run, and for the identified tools a complementary searches were done to make sure data regarding stage supported, ML technique used and price, were collected. For tools supporting more than one phase of LRs, information that is crucial from the perspective of researchers working on reviews, i.e.: regarding keywords highlighting, PDFs searching, studies selection and data extraction support solutions, were extracted. Again, if not available directly from the RR, the additional search of the Internet was conducted. Moreover, for the selected tools an in-depth analysis was done. In certain instances where adequate data were lacking, developers from the support team and/or CEOs of the software-producing companies were contacted for further clarification.

However, despite the strengths of the project, it is important to acknowledge several limitations. Foremost among these is the timeframe of the analysis, which was current only as of June 2023, implying that tools and functionalities may have evolved by the time of this thesis publication. Furthermore, it's crucial to note that not all information was readily accessible, thereby impeding a comprehensive assessment. For example, in one instance, access to a test account was not granted despite repeated requests to the producer. Additionally, the possibility of errors made by analysts during testing should be considered. These factors collectively made it challenging to conduct a thorough comparison of all the tools and arrive at a definitive recommendation.

It would be beneficial to conduct regular analyses of this nature and disseminate the findings to researchers with an interest in the LR field. As previously mentioned, collaboration and information sharing are essential for the ongoing advancement of the area. However, it is imperative to acknowledge the significance of competition among producers as a key factor in this landscape.

6. Search strategies and crucial role of filters

6.1. Key findings

The questions being answered by this section:

- **When and why study design filters should be used?**
- **How to create and test filters?**
- **What are the main reason for studies not being identified when applying the filters, and how they should be managed?**
- **What is the rationale for using different filters for SLRs and RRs?**

- Two filters supporting the identification of observational studies were developed: one that could be used when running RRs, hereinafter referred to as ''basic'', and the second one, broader, that could suit well SLRs, named ''comprehensive''. They were created based on guidelines from SIGN (both), BMJ Knowledge Center's (both), OVID expert searches (both), and CADTH search filters database (comprehensive only); both filters were designed to accommodate for syntaxing differences between MEDLINE and Embase databases, and are intended for use in the OVID platform
- Filters were tested and compared against searches used in 8 Cochrane's SLRs among different disease areas, but always including observational studies in their scope
- Usage of filters resulted in a significant reduction in the number of references to be screened: compared to the search strategies without any restriction to study design, search strategies restricted by basic or comprehensive filters generated 72 and 53% fewer hits, respectively
- The usage of comprehensive filter resulted in 84 to 100% of studies coverage; moreover, in case of half of the analyzed reviews, it was higher than 94%
- The application of the basic filter resulted in data coverage between 84 and 96% for all but 2 studies: Iheozor-Ejiofor 2015 and Lansbury 2019 (coverage of 48% and 68%, respectively)

- Using the basic filter can be considered for RRs, when full coverage is not the objective of the highest priority
- The reduction in the number of abstracts to be screened when using this filter, in comparison to the comprehensive one, was from 18 to 63%, and it resulted in workload being reduced by around 2 to 7 analyst's workdays

- The research shows that some of the subject headings being used in Cochrane's searches seem to be not of relevance for observational studies
- A significant number of missed studies tended to have a very broad, generic indexations in the database, instead of specific observational study designs – such as subject headings for major clinical study, comparative study or controlled study, which generate more than 5 million, 2.9 and 10 million hits in the MEDLINE and Embase databases, respectively (as of June 2023)
- For some studies there is no specific study design indicated (i.e., cross-sectional, cohort, case control) in neither title nor abstract, instead providing the general description of methodology

- **Most of the studies not covered in searches run with either basic or comprehensive filters were not captured because of poor indexing and not using relevant keywords in abstracts; the issues with indexing and describing study design properly make creating the filter of high relevance a difficult task**
- **Nowadays, an increasing amount of data is being generated every day, which makes identification of the data relevant in decision-making that much difficult; also, when dealing with overwhelming amount of data, there is a higher probability of the human error and the false exclusion rates**
- **When conducting an RR, which can streamline or omit a variety of methods used in an SLR to produce evidence in a resource-efficient manner, usage of the study design filter, even a basic one, which is less sensitive than a comprehensive one, is justified – the filter is designed to capture the most relevant studies, while allowing for significant reduction in workload.**

6.2. Introduction

Each SLR is a multi-stage, time-consuming and difficult process. The creation process should be properly defined and accurately specified to minimize systematic errors. The review begins with the formulation of PICOS and the execution of searches[95].

A well-constructed search strategy is the core of an SLR. For unbiased judgement and to avoid overlooking any essential data, a rigorous examination of all available material, including non-English sources, is required[95, 96]. The fundamental databases of biomedical scientific literature are MEDLINE and Embase covering abstracts for millions of published papers[97].

The process of developing search strategies begins with the identification of search phrases. It is vital to find appropriate keywords and synonyms that will allow to find reasonable amounts of relevant material - not too many results that are unmanageable and cause information overload, or too few results that provide insufficient information[98]. Then, the search syntax must be adjusted to each registry because every database has a somewhat different search interface and slightly different search options. Individual search strategies must be developed for selected databases or interfaces using both free text terms and, if available, subject headings[99]. Other useful tools are the Boolean operators that help to combine the terms to broaden (OR) or narrow (AND) the results. In order to achieve a particular type of record, search filters can be used[54]. The mix of search filters and content terms impacts the sensitivity, precision and specificity of a search strategy[100, 101].

Search filters are predetermined combinations of phrases that have been developed to obtain a subset of records based on specific topics of interest[102]. The most effective filters are those for RCTs. These are successful because authors usually indicate their randomization procedures in the title or abstract, and controlled vocabulary has been used regularly in both MEDLINE and Embase since the mid-1990s[103]. However, according to the EUnetHTA, the use of study filters is not recommended to identify studies on diagnostic accuracy unless further search techniques, such as screening reference lists, are applied[104]. Institute for Quality and Efficiency in Health Cares takes a similar stance, noting that there is a contentious debate concerning the use of search filters in the search for diagnostic accuracy research[105]. Even though validated filters for various research types are now available, study filters cannot always be used when looking for non-randomized studies[106, 107].

In principle, applying methodological filters to a search strategy could reduce the number of records returned in a search while preventing the exclusion of important papers and could reduce the number of records that must be examined for inclusion in the review. On the other hand, when incorrectly constructed, filters may raise the risk of missing important records[97]. The research question determines the methodologies and data types that are most appropriate for answering the inquiry[108].

RWE is clinical evidence obtained by real-world data (RWD) analysis concerning the use and prospective advantages or dangers of a medical product resulting from routine healthcare delivery. RWD can be gathered and analyzed using several study designs such as cross-sectional, retrospective/prospective cohort, case-control, and case

series/studies, which are considered observational[109]. Observational research cannot be fully controlled for bias and confounding. Adjustment for potential confounders is possible, but only for a limited number of variables and only for those that are known. However, observational studies are often less expensive than RCTs and provide essential descriptive data and information on long-term effectiveness and safety of drugs[110]. Observational studies can be great in size which allows for investigating an uncommon outcome and can be used when a randomized controlled experiment would be unethical. Additionally, they provide data based on real everyday practice and that is why the interest in them is increasing among stakeholders, and the usage of observational studies to support healthcare decision-making is being discussed more often.

Taking into account the abovementioned information, one of the objectives of the thesis was to create and analyze effectiveness of two methodological filters restricting searches to observational studies. One of them to potentially support SLRs, and the second one to be used in RRs. Developing an effective filter to search observational studies would improve the quality of literature reviews focusing on this study design.

6.3. Methods

The objective was to create and test two filters supporting the identification of observational studies: one that could be used when running RRs, containing only the main terms related to this type of studies, hereinafter referred to as ''basic'', and the second one, broader, that could suit well SLRs, named ''comprehensive'' (Table 9 and Table 10). Both filters were designed to accommodate for syntaxing differences between MEDLINE and Embase databases, and are intended for use in the OVID platform.

The basic filter was created by combining keywords used in the observational search filters published by 3 selected sources: Scottish Intercollegiate Guidelines Network's (SIGN)[111], British Medical Journal (BMJ) Knowledge Center's[112], and OVID expert searches[113]. Those sources were chosen because they were listed by the InterTASC Information Specialists' Sub-Group (ISSG) Search Filter Resource[114] and were focusing on filters that can be used in MEDLINE and Embase databases through the Ovid platform.

The comprehensive filter was developed based on a combination of terms used in the basic filter[111-113] with those published in the Canadian Agency for Drugs and Technologies in Health (CADTH)[115] search filters database[115]. The search filter from CADTH was developed and extensively tested by CADTH's Research Information Services Filters Working Group.

The following observational study designs were included: case-control, cohort, cross-sectional studies and case reports or series[116, 117].

When the filters were ready, the 2nd step of the project was conducted. Several Cochrane SLRs among different disease areas, but always including observational studies in their scope, have been selected. They were treated as the 'gold standard' in the following steps of the project. After removing any terms related to the study design from search strategies used by Cochrane SLRs' authors, the remaining search terms have been merged with each of our observational filters, and searches were run. No additional

sources (e.g., grey literature, reports from Health Technology Assessment Agencies) were searched for the purpose of this analysis.

During the results' analysis, the main interest was in the percentage of relevant references covered by each filter and in the difference between the number of titles and abstracts to be screened when using the basic and comprehensive filter. Estimations were done based on studies covered by the Cochrane group, the 'gold standard'. Only studies with relevant study design, that were available in the OVID platform were used for the coverage calculations. Additionally, studies that were missed due to incorrect indexing of terms not related to study design, that otherwise would have been captured using our filter, were not counted as 'missing'.

For studies which were included by Cochrane SLRs' authors and missed in our searches, possible reasons for not capturing them were checked.

Table 9 Comprehensive filter on observational study design performed in Embase and MEDLINE (Ovid interface)

Line number	Search terms	Results
1	Case control study/	532958
2	Clinical Studies as Topic/	163678
3	Clinical study/	168353
4	Cohort analysis/	1356754
5	cohort studies/	1218553
6	Cross-sectional studies/	896111
7	cross-sectional study/	1024817
8	Epidemiologic Methods/	251901
9	Epidemiologic studies/	250869
10	exp case control study/	1635868
11	exp Epidemiologic Studies/	7558704
12	Family study/	25778
13	follow up/	2040742
14	Longitudinal study/	357495
15	Observational Studies as Topic/	334777
16	observational study/	467318
17	Prospective study/	1530446
18	quasi experimental study/	12281
19	Retrospective study/	2573360
20	single-case studies as topic/	516
21	(Cohort adj (study or studies)).mp.	1001611
22	(Case control adj (study or studies)).tw.	295220
23	(follow up adj (study or studies)).tw.	129568
24	(observational adj (study or studies)).tw.	410847
25	(epidemiologic$ adj (study or studies)).tw.	217013
26	(cross sectional adj (study or studies)).tw.	587870
27	Case control.tw.	357711
28	Cohort analy$.tw.	30947
29	Longitudinal.tw.	756712

Line number	Search terms	Results
30	Retrospective.tw.	1979342
31	Cross sectional.tw.	1174026
32	(Observational Study or Validation Studies or Clinical Study).pt.	146669
33	(observational adj3 (study or studies or design or analysis or analyses)).ti,ab,kf.	563955
34	cohort*.ti,ab,kf.	2313162
35	(prospective adj7 (study or studies or design or analysis or analyses)).ti,ab,kf.	1340964
36	((follow up or followup) adj7 (study or studies or design or analysis or analyses)).ti,ab,kf.	438316
37	((longitudinal or longterm or (long adj term)) adj7 (study or studies or design or analysis or analyses or data)).ti,ab,kf.	836461
38	(retrospective adj7 (study or studies or design or analysis or analyses or data or review)).ti,ab,kf.	1816301
39	((case adj control) or (case adj comparison) or (case adj controlled)).ti,ab,kf.	372120
40	(case-referent adj3 (study or studies or design or analysis or analyses)).ti,ab,kf.	1342
41	(population adj3 (study or studies or analysis or analyses)).ti,ab,kf.	583588
42	(descriptive adj3 (study or studies or design or analysis or analyses)).ti,ab,kf.	272877
43	((multidimensional or (multi adj dimensional)) adj3 (study or studies or design or analysis or analyses)).ti,ab,kf.	10677
44	(cross adj sectional adj7 (study or studies or design or research or analysis or analyses or survey or findings)).ti,ab,kf.	1001556
45	((natural adj experiment) or (natural adj experiments)).ti,ab,kf.	6832
46	(quasi adj (experiment or experiments or experimental)).ti,ab,kf.	45258
47	((non experiment or nonexperiment or non experimental or nonexperimental) adj3 (study or studies or design or analysis or analyses)).ti,ab,kf.	4054
48	(prevalence adj3 (study or studies or analysis or analyses)).ti,ab,kf.	120917
49	or/1-48	13517971

Table 10 Basic filter on observational study design performed in Embase and MEDLINE (Ovid interface)

Line number	Search terms	Results
1	Epidemiologic studies/	250869
2	exp case control studies/	1635868
3	exp cohort studies/	3505022
4	Cross-sectional studies/	896111
5	Clinical study/	168353
6	Case control study/	532958
7	Family study/	25778
8	Longitudinal study/	357495
9	Retrospective study/	2573360
10	Prospective study/	1530446
11	Randomized controlled trials/	420794
12	10 not 11	1512202
13	Cohort analysis/	1356754

Line number	Search terms	Results
14	Case control.tw.	357711
15	(Case control adj (study or studies)).tw.	295220
16	(cohort adj (study or studies)).tw.	775643
17	Cohort analy$.tw.	30947
18	(Follow up adj (study or studies)).tw.	129568
19	(observational adj (study or studies)).tw.	410847
20	Longitudinal.tw.	756712
21	Retrospective.tw.	1979342
22	Cross sectional.tw.	1174026
23	(Cohort adj (study or studies)).mp.	1001611
24	(epidemiologic$ adj (study or studies)).tw.	217013
25	(cross sectional adj (study or studies)).tw.	587870
26	or/1-9,12-25	8702065

6.4. Results

The implementation of search filters resulted in a significant reduction in terms of number of references to be screened: compared to the search strategies without any restriction to study design, search strategies restricted by basic or comprehensive filters generated 72 and 53% fewer hits, respectively (Table 11).

The usage of comprehensive filter resulted in 84 to 100% of studies coverage[118-125]; moreover, in case of half of the analyzed reviews, it was higher than 94% (Table 11)[118-121]. At the same time, the used filter consists of search terms which are justified and logical.

To get full coverage the following terms related to study design should have been used additionally: *comparative study/, health survey/, major clinical study/, controlled study/.* The addition of health survey/ (used in indexing of 24 studies) could be reasonable when the full coverage of relevant studies is the major objective of the review, as can be associated with the way of observational studies conducting. Comparative study/ was used in indexing of 19 studies, however, this term is very wide and not specific to observational studies so adding it to the study design filter could generate a great number of irrelevant records. The same conclusions can be made for major clinical study/ (used in indexing of 23 studies) and controlled study/ (10 studies) as they are also broad terms, related more to RCTs.

The application of the basic filter resulted in data coverage between 84 and 96% for all but 2 studies: Iheozor-Ejiofor 2015 and Lansbury 2019 (coverage of 48% and 68%, respectively) (Table 11). The objective of the filter development was to have one that can be used in RRs, therefore some of the search terms that could be reasonable for observational study design have been excluded, and the focus was only on the most important ones, as the full coverage was not the objective of the highest priority. Still, one of the conclusions was that some of the subject headings being used in Cochrane's searches, seemed to be not of relevance. The reduction in the number of abstracts to be

screened when using this filter, in comparison to the comprehensive one, was from 18 to 63%, and it resulted in workload being reduced by around 2 to 7 analyst's workdays.

Most of the studies not covered in searches run with either basic or comprehensive filters were not captured because of poor indexing and not using relevant keywords in abstracts. Authors often do not include specific information regarding study design in titles and abstracts of their publications, which is an issue. It lead to missing a large number of studies even with use of comprehensive filters. The issues with indexing and describing study design properly make creating the filter of high relevance a difficult task.

Table 11 Comparison of workload when using the basic and comprehensive filter, and articles coverage in each search

	% of reduction in number of references - basic	% of reduction in number of references - comprehensive	% of reduction in references, basic vs comprehensive	Reduction of workload, basic vs comprehensive*	% coverage, comprehensive filter	% coverage, basic filter
Chi 2015[118]	83	66	63	6	100	93
Iheozor-Ejiofor 2015[122]	89	78	52	2	84	48
Kyrgiou 2017[123]	68	49	37	2	87	84
Lansbury 2019[124]	75	44	55	7	90	60
Law 2020[119]	60	52	18	5	100	88
Smith 2022[125]	45	20	31	4	94	84
Taylor-Rowan 2021[120]	63	42	36	2	100	96
Vinceti 2018[121]	86	76	44	5	98	94

* presented as analyst's workdays

In Chi 2015[118] authors searched for both RCTs and observational studies. The objective of the SLR was to assess the safety of topical corticosteroid use during pregnancy. It is rare to conduct RCTs of drugs on pregnant women due to ethical concerns, so it was unlikely to capture any studies with this design. However, the authors decided to include RCTs in the search process, since they are considered the gold standard for testing the interventions[118]. Finally, 14 observational studies have been identified. The search strategy included the following terms related to the relevant study design: *exp cohort studies/, cohort$.tw., controlled clinical trial.pt., epidemiologic methods/, exp case-control studies/, and (case$ and control$).tw*. Usage of the comprehensive search strategy covered all the terms, and therefore no abstract was missed. When using the basic filter, the following terms were missed: *cohort$.tw., controlled clinical trial.pt., epidemiologic methods/* and *(case$ and control$).tw.,* and it resulted in 1 study being not captured[126] (Table 13).

In Iheozor-Ejiofor 2015[122] review authors included 155 observational studies. They used very wide search strategy without restrictions to study design, included prospective studies with a concurrent control and screened 4677 records[122]. As shown in Table 12, 21 studies were missed due to the implementation of the comprehensive filter, resulting in the overall coverage of 86%. Few of the missing studies would have been captured if *major clinical study/* and *comparative study/* would have been added to the filters, however, these subject headings are very general and not specific to observational study design. A subject heading that was used in missed abstracts multiple times, and could be considered as useful in reviews requiring high level of coverage, is *health survey/*. Six of the missed studies were not indexed based on study design; additionally, 3 of them did not include any relevant keyword in abstract, and for 3 other abstracts were not provided at all. Again, the conclusion is that the reason for missing this amount of studies is poor indexing and improper defining study design within abstracts. The coverage of basic filter was 57%, with 67 studies not captured (Table 13). Forty-two publications would be found if *exp epidemiologic studies/* from the comprehensive filter would be used.

Sixty-nine observational studies were included in Kyrgiou 2017[123] review. The search strategy developed by authors was not restricted to study design[123]. By using the comprehensive and basic filters, 9 and 11 studies were missed, respectively. In that case, use of *major clinical study/* and *controlled study/* would make it possible to capture additional studies, however, both are broad terms, not related specifically to observational studies.

In Lansbury 2019[124] authors were interested in RCTs, quasi-experimental designs and observational cohort studies. They used filters for study design, but only in Embase database. Other databases were searched without such filter. The following terms referring to observational studies were included: *('cross sectional' NEXT/1 (study OR studies)):ab,ti, (epidemiologic* NEXT/1 (study OR studies)):ab,ti, (observational NEXT/1 (study OR studies)):ab,ti, ('follow up' NEXT/1 (study OR studies)):ab,ti, ('case control' NEXT/1 (study OR studies)):ab,ti, (cohort NEXT/1 (study OR studies)):ab,ti, 'cohort analysis'/de, 'prospective study'/de, 'retrospective study'/de, 'longitudinal study'/de, 'family study'/de, 'case control study'/de*. All of the terms were covered in both of the created search filters. When analysing missing abstracts (1 and 4 out of 11, for the comprehensive and basic filters, respectively), one of the conclusions was that adding *medical record/* and *medical record review/* to the filters could be useful in very detailed searches. Other indexes used for missing studies would not be of use as too broad and, therefore, would generate lots of irrelevant records.

An SLR run by Law et al.[119] included in the scope RCTs, quasi-randomized controlled trials, non-randomized controlled trials, prospective cohort studies, retrospective cohort studies, case-control studies and cross-sectional studies. Nine thousand seven hundred and nine studies were screened by authors and 68 observational studies were finally included. Authors used the following terms related to observational study design in their search strategies: *exp cohort studies/, exp case-control studies/, exp retrospective studies/, exp epidemiologic studies/, case-control studies/, longitudinal studies/, follow-up studies/, prospective studies/, cross-sectional studies/, (cohort$ or prospective$ or retrospective$).mp*.[119]. Out of all those search queries, *exp cohort studies/* and exp

retrospective studies/ were not used in our comprehensive search strategy, however, exp cohort studies/ was implemented in the basic filter. As the comprehensive filter includes also other terms with which some of the observational studies were indexed, using it resulted in 100% coverage. The implementation of the basic filter resulted in missing 8 studies, due to the lack of following entry terms: *exp epidemiologic studies/, follow up/,* and *cohort.tw.* Similarly to previous analyses, *major clinical study/* and *controlled study/* were used when indexing most of the missed studies.

Smith et al.[125] searched for observational studies (cross-sectional and prospective cohort studies) and RCTs. The authors did not restrict the search strategy to any study design, and the initial pool of references was 11,480 records. Finally, 50 observational studies were included[125], out of which 3 were missed when using the comprehensive filter, and 8 when using the basic one. One of the terms that was missing in both filters, and could be considered in reviews requiring high level of coverage, was *questionnaire/.* Indexes often used for missed studies, however, too broad and not specifically related to observational studies, were: *major clinical study/, clinical trial/,* and *controlled study/.*

Taylor-Rowan et al.[120] run an SLR focusing on prospective and retrospective longitudinal cohort and case-control studies, and identified 25 observational studies meeting their inclusion criteria. Authors used broad search strategy without restricting the study design, which generated 11,264 records in total. None of the included studies were missed due to the usage of the comprehensive filter, whereas with the basic filter, 1 study was missed (96% of coverage). As presented in Table 13, the study[127] would have been found with the subject heading *follow-up/* used in the comprehensive search strategy. Additionally, it was indexed as *controlled study/* in the database.

In a review conducted by Vinceti et al.[121] RCTs and observational studies of longitudinal design (cohort studies and nested case-control studies) were included. Eight hundred and fifty nine studies were screened, and 105 observational studies were selected for inclusion[121]. Terms used in search strategy which refer to observational study design included: *exp clinical study/, (observ* or cohort* or prospectiv* or (case* and control*)).mp, exp case-control studies/,* and *exp cohort studies/.* In comparison, the search filters developed for this analysis did not include a broad, truncated keywords in the multipurpose field, as well as the unspecific subject heading for clinical study. Two studies were not captured using the comprehensive search due to lack of indexation [128], or usage of broad subject heading only (*clinical article/*[129]) (Table 12). When applying the basic search filter, 6 out of 105 studies were missed which resulted in 94% of coverage. Among search terms used in the SLR the ones related to study design were general, often consisting of very broad terms searched in all fields, whereas in the tested filter selected terms were often connected with other relevant words or phrases in titles, abstracts and keyword heading word field, which significantly reduced number of abstracts to be screened.

Table 12 Results of the search with the comprehensive filter

Study name	**Studies missed because of the filter**	**MeSHes/Emtree/free terms present in the OVID entry of the study**
Chi 2015[118]	NA	NA
	Al-Alousi 1975[130]	*clinical study Major clinical study/
	Alarcon-Herrera 2001[131]	Major clinical study/
	De Crousas 1982[132]	Comparative study/
	Forrest 1965[133]	None†
	Franzolin 2010[134]	None†
	Goward 1982[135]	Comparative study/
	Haavikko 1974[136]	Prevalence adj4 study.tw.
	Künzel 1976[137]	None‡
	Leverett 1986[138]	Comparative study/
	Louw 2002[139]	None†
Iheozor-Ejiofor 2015[122]	Mandinic 2010[140]	Controlled study/ Major clinical study/
	Marya 2012[141]	Prevalence adj5 study.tw.
	McGrady 2012[142]	None‡
	Ray 1982[143]	Comparative study/
	Russel 1951[144]	None‡
	Saravanan 2008[145]	None‡
	Scheinin 1964[146]	None‡
	Segreto 1984[147]	Comparative study/
	Venkateswarlu 1952[148]	None‡
	Wenzel 1982[149]	None†
	Zimmermann 1954[150]	None‡
	Anderson 1948[151]	None‡
	Buller 1982[152]	Follow-up adj8 study Short survey/
	Guo 2013[153]	Controlled study/ Major clinical study/
Kyrgiou 2017[123]	Kristesen 1985[154]	Clinical article/
	Larsson 1982[155]	Major clinical study/
	Moinian 1982[156]	None†
	Poon 2012[157]	Secondary analysis/
	Sagot 1995[158]	None†
Lansbury 2019[124]	Jain 2009[159]	Controlled study/ Major clinical study/ Medical record/ Medical charts.tw
Law 2020[119]	NA	NA
	Haberer 2011[160]	Prospectively.tw Controlled study/
Smith 2022[125]	Jiamsakul 2014[161]	Major clinical study/ Monitoring study.tw
	Okonji 2012[162]	Major clinical study/
Vinceti 2018[121]	Menkes 1986[129]	Clinical article/
	Peleg 1985[128]	None‡

adj_n – defined adjacent operator, retrieves records that contain search terms within a specified number (n-1) of words from each other in any order; NA – not applicable; tw. – term searched in title and abstract; / - subject heading (MeSH or Emtree); * - focussed term; † if no keywords connected to study design in abstract and no indexing connected to study design were used in Embase and MEDLINE; ‡ if no abstract was available and no indexing connected to study design was available in Embase and MEDLINE

Table 13 Results of the search with the basic filter

Study name	Studies missed because of the filter	Indexing of missed studies: MeSHes/Emtree/free terms included in the comprehensive filter	Indexing of missed studies: Additional MeSHes/Emtree/ free terms used to index the studies
Chi 2015[118]	Mahé 2007[126]	Follow-up/	None
	Acharya 2005[163]	Exp epidemiologic studies/	None
	Adair 1999[164]	Exp epidemiologic studies/	Prevalence/ Health survey/ Comparative study/
	AlDosari 2010[165]	Exp epidemiologic studies/	*health survey
	Arif 2013[166]	Exp epidemiologic studies/	None
	Awadia 2000[167]	Exp epidemiologic studies/	None
	Awadia 2002[168]	Exp epidemiologic studies/	*health survey *analysis
	Bao 2007[169]	Exp epidemiologic studies/	None
	Berndt 2010[170]	Exp epidemiologic studies/	Health survey/
	Birkeland 2005[171]	Exp epidemiologic studies/	Major clinical study/ Controlled study/
	Butler 1985[172]	Exp epidemiologic studies/	Short survey/
	Brothwell 1999[173]	Exp epidemiologic studies/	Pilot study/
	Clarkson 1992[174]	Exp epidemiologic studies/	Questionnaire/
	Correia Sampaio 1999[175]	Exp epidemiologic studies/	Comparative study/
	Cypriano 2003[176]	Exp epidemiologic studies/	Health survey/
	Driscoll 1986[177]	Exp epidemiologic studies/	Health survey/
	Ekanayake 2002[178]	Exp epidemiologic studies/	Health survey/
	Eklund 1987[179]	Exp epidemiologic studies/	Health survey/
	Ellwood 1995[180]	Exp epidemiologic studies/	None
	Ellwood 1996[181]	Exp epidemiologic studies/	Health survey/
	Ermiş 2003[182]	Exp epidemiologic studies/	Health survey/
Iheozor-Ejiofor 2015[122]	Firempong 2013[183]	Exp epidemiologic studies/	Comparative study/ Incidence/
	García-Pérez 2013[184]	Exp epidemiologic studies/	Health survey/
	Grobleri 2001[185]	Exp epidemiologic studies/	Comparative study/ Health survey/
	Hardwick 1982[186]	Exp epidemiologic studies/	Health survey/ Comparative study/
	Heintze 1998[187]	Exp epidemiologic studies/	Comparative study/
	Heller 1997[188]	Exp epidemiologic studies/	Health survey/
	Hong 1990[189]	Exp epidemiologic studies/	None
	Ibrahim 1995[190]	Exp epidemiologic studies/	None
	Ismail 1990[191]	Exp epidemiologic studies/	Health survey/
	Machiulskiene 2009[192]	Exp epidemiologic studies/ (prevalence adj3 (study or studies or analysis or analyses)).ti,ab,kf.	Health survey/ Comparative study/ Clinical trial/ Multicenter study/
	Masztalerz 1990[193]	Exp epidemiologic studies/	Health survey/ Comparative study/
	Mella 1992[194]	Exp epidemiologic studies/	Pilot study/
	Riordan 1991[195]	Exp epidemiologic studies/	None
	Riordan 2002[196]	Exp epidemiologic studies/	Health survey/
	Ruan 2005	Epidemiologic methods/	Comparative study/ Controlled study/ Major clinical study/ Questionnaire/

Study name	Studies missed because of the filter	Indexing of missed studies MeSHes/Emtree/free terms included in the comprehensive filter	Additional MeSHes/Emtree/ free terms used to index the studies
	Rugg-Gunn 1997[197]	Exp epidemiologic studies/	None
	Rwenyonyi 1999[198]	Exp epidemiologic studies/	Comparative study/
	Selwitz 1998[199]	Exp epidemiologic studies/	Comparative study/ Health survey/
	Stephen 2002[200]	Exp epidemiologic studies/ (prevalence adj3 (study or studies or analysis or analyses)).ti,ab,kf.	Health survey/ Single blind procedure/ Single blind method/
	Szpunar 1988[201]	Exp epidemiologic methods/	None
	Tabari 2000[202]	Exp epidemiologic studies/ (prevalence adj3 (study or studies or analysis or analyses)).ti,ab,kf.	Comparative study/ Health survey/ Single blind procedure/ Single blind method/
	Tsutsui 2000[203]	Exp epidemiologic studies/	Health survey/
	Vignarajah 1993[204]	Exp epidemiologic studies/	Comparative study/ Health survey/
	Villa 1998[205]	Exp epidemiologic studies/	Comparative study/ Comparative study.tw. Health survey/
	Warnakulasuriya 1992[206]	Exp epidemiologic studies/	Health survey/
	Zheng 1986[207]	Exp epidemiologic studies/	Health survey/
Kyrgiou 2017[123]	Braet 1994[208]	case control*	Controlled study/ Major clinical study/
	Haffenden 1993[209]	Follow up/	Major clinical study/
	Delgado-Rodríguez 2013[210]	Exp epidemiologic studies/	Medical record review/ Clinical charts.tw
Lansbury 2019[124]	Kinikar 2012[211]	Exp epidemiologic studies/	Major clinical study/ Hospital records.tw
	Mady 2012[212]	Exp epidemiologic studies/	Retrospective*.tw
	Ayoub 2018[213]	Exp epidemiologic studies/	Controlled study/ Major clinical study/
	Ferrante 2017[214]	Cohort*.ti,ab,kf.	Major clinical study/ Medical record review/
	Guasch 2016[215]	Follow up/	Clinical trial/ Controlled study/ Major clinical study/
	Nguyen 2014[216]	Cohort*.ti,ab,kf.	Prospective cohort.tw.
Law 2020[119]	Schluender 2007[217]	Exp epidemiologic studies/	Clinical assessment/ Controlled study/ Major clinical study/ Medical record review/
	Shaib 2017[218]	Exp epidemiologic studies/	Controlled study/ Major clinical study/
	Uchino 2013[219]	Exp epidemiologic studies/	Controlled study/ Major clinical study/
	Wilson 2014[220]	Cohort*.ti,ab,kf.	Major clinical study/
	Avong 2015[221]	Cross-sectional study/, observational study/	None
	Duarte 2015[222]	Cohort*	Questionnaire/
Smith 2022[125]	Ekstrand 2010[223]	Follow-up/	Major clinical study/
	Oette 2006[224]	Clinical study/	Clinical trial/ Controlled study/ Major clinical study/

Study name	Indexing of missed studies		
	Studies missed because of the filter	**MeSHes/Emtree/free terms included in the comprehensive filter**	**Additional MeSHes/Emtree/ free terms used to index the studies**
Taylor-Rowan 2021[120]	Parienti 2010[225]	Cohort analysis/	Major clinical study/
	Shah 2013[127]	Follow up/	Controlled study/
	Fex 1987[226]	(population adj3 (study or studies or analysis or analyses)).ti,ab,kf.	None
Vinceti 2018[121]	Heinen 2007[227]	(prospective adj7 (study or studies or design or analysis or analyses)).ti,ab,kf.	None
	Li 2005[228]	Follow up/	None
	Schober[229]	Exp epidemiological studies/	None

adj_n – defined adjacent operator, retrieves records that contain search terms within a specified number (n-1) of words from each other in any order; Exp – retrieves exploded search results; NA – not applicable; tw. – term searched in title and abstract; / - subject heading (MeSH or Emtree); * after term – unlimited truncation; * before term - focussed term

6.5. Discussion

In general, the overall performance of the developed filters was satisfactory – for the comprehensive filter, which was designed to be more sensitive, the average coverage of available data was over 94%, while for the basic filter, which could be implemented in RRs aiming to synthesize key trends in a time-efficient manner, it was between 84 and 96% for all but 2 studies.

A few general trends were observed. Firstly, a significant number of missed observational studies was added to databases with very broad, generic indexations, instead of specific observational study designs – such as subject headings for major clinical study, comparative study or controlled study, which generate more than 5 million, 2.9 and 10 million hits in MEDLINE and Embase databases, respectively (as of June 2023). Using such unspecific terms in the search filter would decrease the precision of the filter significantly, retrieving numerous studies with irrelevant study design, such as RCTs.

Secondly, some studies do not indicate the specific study design (i.e., cross-sectional, cohort, case control) in neither title nor abstract, instead providing the general description of methodology. Therefore, it would be impossible to develop a filter that would be able to catch each of the references based on keywords used in titles and abstracts, while remining focused enough to be useful.

Lastly, the majority of studies missed when using the comprehensive filter (63%) were published prior to the year 2000. Older references tend to be very scarcely indexed in the database; they often lack subject headings related to anything other than disease, and abstract is not available. Therefore, the methodological filters might not be justified when looking for older records, for instance, when conducting an SLR on a drug that have been used for many decades. However, when such records are expected, the methodological filter could still be applied to the records published from 2000 onwards.

This analysis has several strengths and limitations. Firstly, the filters were created based on several sources recommended by the ISSG Search Filters Resource. In one interview, the information specialists and project managers working for NICE, the NICE collaborating centres and the NICE Evidence Review Groups (ERGs) admitted that they

sometimes used search filters when carrying out searches for a particular study type, and the ISSG Search Filters Resource was among the most frequently reported sources[230]. Moreover, to test the filters, Cochrane SLRs conducted among different disease areas have been selected and treated as the 'gold standard'. The searches used for tests were created by combining the filters with exactly the same search terms as used previously by the Cochrane Group for population, intervention and comparators (when applicable), and outcomes (when applicable). This methodology allowed for precise testing and making relevant conclusions. When analyzing the results, different points influencing the filters usefulness were considered, such as indexing, date of articles publishing, KWs used in titles and abstracts, relevance of some subjects headings.

However, some limitations arose despite the project' strengths. The main one is the fact that some sources suggest the use of methodological search filters should be avoided when conducting an SLR to not miss any study that should be included into the analysis[231]. However, nowadays, an increasing amount of data being generated every day, which makes identification of the data relevant in decision-making that much difficult. It is also important to factor in the human error and the false exclusion rates, which can also lead to data omission, especially when dealing with overwhelming amount of data[232]. Across all reviews analysed in this work, the inclusion of methodological filters into the search strategy allowed for significant workload reduction; in comparison to broad, unrestricted searches, up to 89% less references would need to be screened when using basic filter, and 78% for comprehensive.

In general, to account for possible data gaps when conducting an SLR, it is advised to include additional data sources, such as scanning reference lists of identified papers or other reviews in search of relevant studies.

When conducting an RR, which can streamline or omit a variety of methods used in an SLR to produce evidence in a resource-efficient manner, usage of the study design filter, even a basic one, which is less sensitive than a comprehensive one, is justified – the filter is designed to capture the most relevant studies, while allowing for significant reduction in workload. Therefore, the created filters can be used by researchers interested in observational studies. The filters should be tested from time to time, to check if no changes have been made when it comes to subject headings in MEDLINE and Embase, and can be translated to be used in other databases when needed.

The relevant filters are of great value, and further work on them is needed so they can support reviews for all types of study designs.

7. Conclusions

SLR is a highly complex, rigorous, and transparent process of synthesizing evidence. SLRs are used to inform decisions, usually in healthcare. However, health policymakers may have limited time frames, and a limited budget or short staff might confine the process of SLR. RR might be a solution to consider. This type of knowledge synthesizing is becoming more commonly discussed every year, especially since the global COVID-19 pandemic when time was of the essence. A timely and affordable approach could provide valuable and relevant evidence to enhance healthcare and systems, but SLRs are still the gold standard in healthcare policymaking. It is critical to adapt RRs to the various decision makers' needs, to improve the methodology of RRs further, and analyze all the potential biases that could be introduced with new solutions, as it might allow widespread and accurate use of this form of evidence synthesis.

This thesis aimed to support placing SLRs and RRs in the area of healthcare policy and assess their perspectives in the era of new technologies. The objectives were achieved by:

- analyzing the process of SLRs development and publishing over the years, including the COVID-19 pandemic
- systematically reviewing and assessing HTAs' requirements regarding LRs supporting the funding process
- summarizing and assessing guidelines and studies on RRs' methodology
- creating and testing three different methodologies of RRs and comparing results of searches, studies selection and meta-analyses with the ones achieved by the Cochrane group in SLRs with the identical scopes
- developing and testing two different Ovid filters enabling identification of observational studies, and comparing their effectiveness to the ones used by the Cochrane group among various SLRs
- analyzing new AI tools and software supporting the development of LRs.

First part of work was conducted to assess requirements regarding methodologies of LRs supporting the refund process, to analyze how many SLRs are being produced within time, what are the reasons this amount is increasing every year, and, finally, what approaches are suggested when RR can be a potential solution.
It has been demonstrated that HTAs' recommendations regarding LRs give a broad possibilities of interpretation, however, at the same time, each agency expects from researchers to capture all the relevant studies, and provide deliverables of the highest quality. As amount of scientific publications is exploding, there is a need to do more SLRs, with a massive amount of data to be analyzed. It results in high human resources demands, but, on the other hand, timelines for different activities, such as simple manuscripts submissions, but also HTA submissions, do not change. The final effect is a low quality of reviews, and, as this PhD thesis's results show, even reviews done by the

Cochrane group are not always the ones of the highest quality. In that case, RRs can be a solution. However, despite the fact there are many guidelines and methodological recommendations on how to run and assess the quality of LRs, there has yet to be a unified guideline on distinguishing between SLR and RR, and which approach should be used under what conditions. Reviews authors often use the terms SLR, RR, targeted literature review, and desk research interchangeably, which can result in significant consequences, especially in the healthcare area. There is a need for a unified and clear vocabulary to avoid confusion, an aligned methodology for performing SLR and RR, and recommendations on when SLRs or RRs are appropriate. It remains important to define how to interpret SLR/RR and how to adopt pragmatic attitude in using SLR/RR to set policies based on LRs' outcome. This last point is rarely addressed, as if this was obvious, not critical, or left to decision makers judgement.

The second part of this work was a practical one. The objective was to test 3 chosen RRs' methodologies and to what extend they can replace SLRs, to analyze and check which AI tools and software can support LRs process and how far they can replace human analysts, and finally, to present how to create study design filters in a well-founded way; effectiveness of filters have been tested against the ones used by Cochrane group in 8 different SLRs.
It has been showed that using 3 different approaches to RRs resulted in 84, 89 and 100% of efficacy (defined as the number of studies captured during the RR divided by the number of studies identified in SLR), and efficacy of 100% corresponded to the methodology requiring the least human resources, related to work reduction by 18 analyst workdays, and based on searching of only one database. The most complex approach to RR, including searches of several sources and the AI tool support, resulted in 89% of efficiency and 11 workdays saved. In total, 6 out of 56 trials have been lost as a result of modified methodology in all 3 RRs (3 in RR_1 and 3 in RR_2). However, the three points that should be highlighted are: 1) among those six studies, five have been assessed as having high RoB, and for 1 the RoB was not clear, 2) four studies were not available in MEDLINE and Embase and two were indexed inappropriately, 3) missed studies were included by Cochrane in 28 meta-analyses in total, however, their lack in analyses run based on RRs' results impacted in a significant way results of only 4 of them. Those observations highlight the necessity of relevant indexing of records, as well as using filters comprehensive enough to capture all important studies, and, at the same time, specific enough to make running LR possible within reasonable timelines. Filters to be used are especially crucial when searching for specific study design. Part of this PhD thesis dedicated to creating filters for observational studies showed that this type of publications is often being indexed with terms not relevant for study design (e.g. major clinical trials) or by using broad terms for names of outcomes of some RWE studies in general, instead of focusing on terms related to the analysed study's design. It results in creating filters that are not methodologically justified, and one of the reasons can be just to make sure the specific studies will be captured in the review. By using the developed comprehensive filter consisted of only justified keywords, 84 to 100% of studies identified by Cochrane group were captured also in RR.

The last step of this work was dedicated to analysis of AI tools and software supporting LRs. The project consisted of RR identifying what is available on the market, on summary analysis based on the review's results and information from producers' websites, and on running tests of several selected programs. The general conclusions demonstrate that the most commonly supported stage of LRs is screening, where the tools assist the reviewer by suggesting, classifying, or selecting records. The conducted research show the usage of the AI-supported solutions during titles and abstracts reviewing can be a time-saving solution. None of the tested software showed good outcomes when it comes to automatic or semi-automatic data extractions. Especially for this step, further collaboration and information sharing between researchers seems to be one of the crucial points needed for future development of the area. This observation is in line with other authors' opinions, caused by the fact that data extraction is the process that needs a lot of aspects to be considered, and it is highly probable that extraction of objective and some basic characteristics of a study, such as a sample size, will be supported in the future, but there is a low probability for automatic extraction of complex data that can be interpreted in multiple ways, such as main conclusions[93]. Two additional points that should be considered by researchers making decision on the potential usage of the software are: to draw special attention when deciding for using a free programs (especially when working on commercial projects), as they can be open-source aiming data collection from different reviewers and building database to be used then for further tests, and therefore, may not be secured enough; and to ask for free trial period when considering paid software, as prices of some of tools are relatively high, and functionalities are comparable with the ones available for free or at a lower price.

The most important goal of this work is to facilitate an understanding of how crucial it should be for reviewers working on SLRs and RRs to make sure the chosen methodology is relevant, as well as for policymakers to critically evaluate LRs' quality and consider potential biases when analyzing the results and making recommendations. Ultimately the use of SLR or RR seems to be able to lead to similar conclusions. Improving indexation and experience in filtering appears to be a way forward to enhance RRs effectiveness and efficiency over SLRs. However, an appropriate indexation is the responsibility of the publications' reviewers and journals. Introducing specific procedures to validate indexation is critical to enhancing the effectiveness of RRs. Filtering remains a know-how. As any know-how is acquired by experience, serendipity and mirroring experienced expert practice. The challenge will be to conceptualize know-how so it may be shared at a broader level and benefit widely. This will require maturity or research and abstraction abilities. It may be achieved in the future by operating workshops between experts with the aim to structure and standardize such know-how.

8. Appendices

8.1. LRs and HTA agencies' requirements

Table 14 Detailed HTA's requirements regarding clinical LRs

Agency	SLR needed	PICOS	Databases&sources	Search strategy	Selection process	Critical appraisal
NICE	+++ Required	+++ Definition prior to SLR running needed	+++ Electronic databases + published/unpublished research + scanning references + many other The selection of electronic databases to search should depend upon the review topic. Search should not be restricted to use of electronic databases; searching databases and registers that include unpublished studies, such as records of ongoing research, conference proceedings and theses, can reduce the impact of publication bias. For reviews of health care interventions, MEDLINE and Embase are the databases most commonly used to identify studies. CENTRAL includes details of published articles taken from bibliographic databases and other published and unpublished sources Published and unpublished research may also be obtained by	+++ Should be described. The methods used should be justified with reference to the decision problem. Sufficient detail should be provided so that the results may be reproduced. This includes a full list of all information sources and the full electronic search strategies for all databases, including any limits applied. limiting searches to English language papers can introduce language bias. (CRD)	+++ Description of the inclusion and exclusion selection criteria, language restriction and the study selection process in a table. -A log of ineligible studies with the rationale for why studies were included or excluded. -Having more than 1 reviewer. The procedure for resolving disagreements between reviewers should be reported. -A flow diagram of the numbers of studies included and excluded at each stage should be provided using a validated statement for reporting systematic reviews and meta-analyses. -Complete reference list of included and excluded studies are required.[1]	+++ All evidence should be critically appraised, and potential biases must be identified. Describe the methods used for assessing risk of bias and generalisability of individual trials (including whether this was done at the study or outcome level), and how this information is to be used in any data synthesis. Key aspects of quality to be considered can be found in Systematic reviews: CRD's guidance for undertaking reviews in health care (section 2.2.5).[3] A suggested table format for the quality assessment results for parallel group RCTs is provided by NICE.

Agency	SLR needed	PICOS	Databases&sources	Search strategy	Selection process	Critical appraisal
			using one or more of the following methods: scanning reference lists of relevant studies, handsearching key journals, searching trials registers, contacting experts and manufacturers, searching relevant Internet resources, citation searching, etc. (Systematic Reviews: CRD's guidance for undertaking reviews in health care)			
HAS	+++	- NR, However, they do specify that they perform searches in accordance with PRISMA-P and Cochrane and both specify the necessity of the usage of PICOS	++ All clinical data sources for the intervention being evaluated and its comparators should be identified through a systematic, reproducible methodology, and then presented in accordance with international standards and HAS guidelines	++ Data should be found through a systematic literature review conducted in accordance with the PRISMA-P guidelines or Cochrane Handbook, and reported according to the PRISMA guidelines	- Where a study is not selected, its conclusions should be briefly presented, and the reason for its exclusion should be explained. The selected studies should be presented in a detailed manner in a table (methodology, results, main elements required for their interpretation, level of evidence). When known, the sources of variability of the expected effect should be documented and described so they can be considered in the	+ Available data sources to document the effectiveness and tolerability of each intervention should undergo a stringent critical appraisal.

Agency	SLR needed	PICOS	Databases&sources	Search strategy	Selection process	Critical appraisal
					analysis (interaction variables such as age, gender, disease severity, etc.)	
IQWIG	+++ Required	+++ Required	+++ In information retrieval, superordinate aspects such as reporting bias (including publication bias) should be taken into account as they have a fundamental effect on the selection of the different information sources. Published studies usually appear in scientific journals and thus can largely be searched for via bibliographical databases such as MEDLINE or Embase. Bibliographical databases (besides trial registries) represent a main source for identifying study results, especially if no clinical study reports are available. The Cochrane Handbook lists MEDLINE, Embase and CENTRAL as the 3 most important bibliographic databases. Depending on the research question of the	+++ The search is usually not limited to specific years. However, a restriction to English and German publications can be made, since evaluations have shown that the exclusion of non-English publications does not change the conclusion of most systematic reviews. A combination of subject headings (including publication type) and free-text terms is necessary for the development of search strategies. The PRESS checklist is therefore used to support the process of quality assurance. Searches in trial registries should show high sensitivity, be simple, and, if possible, only consider one component (usually the therapeutic indication or intervention)	++ Due to the primarily sensitive approach, the literature search in bibliographic databases results in a large number of citations that are not relevant to the assessment. The selection of relevant publications is made in several steps: - Exclusion of definitely irrelevant publications (i.e. publications not fulfilling the inclusion or exclusion criteria of the report plan or project outline) through perusal of the titles, and, if available, the abstracts. - The full texts of the remaining potentially relevant publications are obtained. The decision on the inclusion of the study in the assessment concerned is then made on the	+++ The risk of bias is substantially affected by the study quality; however, its assessment is not equivalent to the quality assessment of a study. For example, individual outcomes may also be considerably biased in a high-quality study. Other studies, however, may provide high certainty of results for specific outcomes in individual cases, despite being of low quality. As a rule, the Institute will therefore estimate the extent of the risk of bias in a problem-orientated manner for all relevant results (both for the study and the specific outcomes).

Agency	SLR needed	PICOS	Databases&sources	Search strategy	Selection process	Critical appraisal
			benefit assessment, regional or topic-specific databases can also be considered. It is also important to perform searches in clinical trials databases. A systematic search always considers several trial registries, as no trial registry includes all studies Other sources are considered either as a standard or optionally and do not need to be searched systhematically, eg regulatory authorities, conferences, etc.		basis of these documents. All selection steps are performed by 2 persons independently of each other. Discrepancies are resolved by discussion. In the first selection step, if doubts exist as to the relevance of a study, the corresponding full text is obtained and assessed. The documentation of study selection is performed as transparently as possible and contains the decisions on the inclusion or exclusion of each citation (only on the full-text level). Study selection is performed in IQWiG's internal web-based trial selection database (webTSDB). The references evaluated are further documented in the course of report production using the reference management software Endnote.	

Agency	SLR needed	PICOS	Databases&sources	Search strategy	Selection process	Critical appraisal
ICER	+++ Required	+++ Definition prior to SLR running needed	+++ MEDLINE, Embase, and CENTRAL for studies relevant to the appraisal. When appropriate, discipline-specific databases should also be searched. To supplement the database searches, investigators perform a manual check of the bibliographies of recent systematic reviews and invite key stakeholders to share references germane to the scope of an appraisal. Investigators also search for data contained in grey literature sources, such as regulatory documents, conference proceedings, data shared by life science companies, and open-input submissions from stakeholders that may be relevant to the topic.	++ The search strategies include a combination of indexing terms (MeSH terms in MEDLINE and Emtree terms in Embase), as well as free-text terms. Searches are generally limited to English-language studies of human subjects and exclude articles indexed as guidelines, letters, editorials, narrative reviews, case reports, or news items	+++ Study selection occurs through two levels of screening: at the title/abstract and full-text levels. Initially, each title and abstract identified through electronic searches is independently screened by two investigators. A third reviewer works with the initial two reviewers to resolve any issues of disagreement through consensus. Reasons for exclusion are categorized according to the PICOTS elements during both title/abstract and full-text review. A flow diagram illustrating the selection of studies is presented in the appendix of each report.	+++ For all HTAs, the identified studies are summarized in the text and in evidence tables of the Evidence Report. This summary is key to understanding the existing evidence base pertaining to the HTA topic. Relevant data include those specified in the data extraction section. Any key differences between the studies in terms of study design, patient characteristics, interventions (including dosing and frequency), outcomes (including definitions and methods of assessments), patient subgroups, and study quality are described. The methodological quality of included studies is evaluated using appropriate tools for the evidence base under review. For each review, ICER will select or adapt a quality rating tool that is relevant to the

Agency	SLR needed	PICOS	Databases&sources	Search strategy	Selection process	Critical appraisal
						topic and document the choice in the protocol. The assessment of risk of bias requires judgment about how particular aspects of a study may lead to biased results. The rationale for the assessments should be explicitly determined a priori in the protocol for the evidence review, and the judgment itself is provided in the appendix of each report
CADTH	+++ Required	+++ Definition prior to SLR running needed. At the initiation of the process, CADTH develops a protocol to ensure that the review will reflect the most relevant clinical information.	++ When drafting the review protocol, CADTH considers a variety of information, such as clinical practice guidelines, availability of comparator drugs, clinical trial protocols, and stakeholder input (i.e., information from patient groups, clinical experts, drug plans, and expert review committee members)	+++ CADTH designs and conducts one or more independent systematic literature searches according to the protocol and to supplement the submission material provided by the sponsor. The search strategy used and the relevant literature that is identified are included n the clinical review.	+ A list of studies that will be included in the systematic review portion of the clinical review is sent to the sponsor for information purposes. Additional relevant evidence from studies that are not included in CADTH's systematic review may be included in other portions of the clinical report. CADTH summarizes and critically appraises the relevant studies in the clinical report.	+++ CADTH summarizes and critically appraises the relevant studies in the clinical report. Strengths and limitations with respect to both internal validity are documented. A tailored review consists of the review team conducting an appraisal of the clinical evidence and pharmacoeconomic evaluation filed by the sponsor using a

Agency	SLR needed	PICOS	Databases&sources	Search strategy	Selection process	Critical appraisal
					Strengths and limitations with respect to both internal validity (i.e., how well the study was designed, conducted, and reported) and external validity (i.e., how well the results of the study could be applied to the target population in Canada) are documented	CADTH-provided review template. CADTH validates and critically appraises the information provided by the sponsor in the template. Strengths and limitations with respect to both internal validity (i.e., how well the study was designed, conducted, and reported) and external validity (i.e., how well the results of the study could be applied to the target population in Canada) are documented. CADTH includes its assessment of the submitted information and comments directly into the appropriate sections of the tailored review template. A single report combining both clinical and pharmacoeconomic information is prepared by CADTH for tailored reviews (i.e., CADTH Clinical and Economic Review Report

Agency	SLR needed	PICOS	Databases&sources	Search strategy	Selection process	Critical appraisal
PBAC	+++ Required	+++ Definition prior to SLR running needed	+++ MEDLINE, Embase, Cochrane Library, clinicaltrials registries (clinicaltrials.gov, International Clinical Trials Registry Platforms, Australian Clinical Trials Registry), internal registries, other (report on the details of supplementary searches, including manual checking of the references in retrieved papers, searches of the Therapeutic Goods Administration dossier and searches of grey literature).	+++ Search filters should initially be set to include only randomised trials. For population, MeSH terms, text words and synonyms for the target population/disease/condition should be included. It should be ensured that the search terms for population are broad and should only be applied if the proposed medicine is used for multiple indications. For intervention and/or comparator, known proprietary and nonproprietary names, MeSH terms and developmental/provisional medicine names should be included. For study design – Cochrane Highly Sensitive Search Strategies for identifying randomised trials in MEDLINE, or MeSH and text word terms for nonrandomised study designs should be used. The search terms for the systematic literature search in MEDLINE should be presented. For other sources (such as Embase), any key differences from the search strategy should be provided. It should be ensured that the	++ Identified trials (e.g., in a spreadsheet) should be listed, with indication of which trials were excluded and the basis for their exclusion. PRISMA flowchart should be constructed. If neither direct randomized trials nor other randomized trials suitable for an indirect comparison are retrieved, investigators should broaden the original search for the proposed medicine to identify all nonrandomized studies of the proposed medicine, preferably compared with the main comparator, that recruited participants whose characteristics overlap with the target population. Relevant study types include cohort studies, case-control studies and quasi-experimental studies. Double screening not mentioned	+++ The preferred approach for randomized trials is described in Chapter 8 of the Cochrane handbook for systematic reviews of interventions (version 5.1.0). Assessment of the risk of bias for included individual trials within an identified systematic review or meta-analysis. Where individual trials are not able to be retrieved and the submission relies on a pooled treatment effect from the published systematic review and meta-analysis, researchers should report the risk of bias assessment undertaken by the authors of the systematic review; also assess the quality of the systematic review using a validated tool (e.g., AMSTAR4 or ROBIS). Assessment of risk of bias appropriate for the study design

Agency	SLR needed	PICOS	Databases&sources	Search strategy	Selection process	Critical appraisal
				search terms are consistent with the search criteria. Search strategies for all data sources should be summarized.		and conduct of the included nonrandomized studies (The Risk Of Bias In Non-randomized Studies – of Interventions (ROBINS-I) tool).
C2H	+++ Required	++ When an SLR is conducted, clinical questions (CQs) should be clearly presented. For example, a definition of structured CQs according to PICOS.	- NR	++ Detailed information about databases used, search algorithm and research selection process is required	+++ PRISMA flowchart is required in order to clarify a flow from searching to the selection of clinical studies for a final review. A description of the inclusion/exclusion criteria, databases used, search algorithm, and research selection process (inclusion flow of information) is required in accordance with the PRISMA. Target studies (articles) for the final review should be listed and explained. A time point between determining the framework of analysis and manufacturer's submission can be used as a cut-off date for the literature search in the SLR.	++ It is desirable to evaluate the quality of studies that have been identified in the systematic review and ultimately used to evaluate additional benefit

Agency	SLR needed	PICOS	Databases&sources	Search strategy	Selection process	Critical appraisal
					When no studies or only insufficient studies are available based on the result of SLR, additional benefit is evaluated by SLR of comparative non-RCT (e.g. observational) studies, if agreed upon in consultation. In that case, sufficient explanation on research quality is needed (e.g., study design, differences in background factors between groups, methods of statistical analysis, sample size, and number of institutions).	
EUnetHTA						

Table 15 Detailed HTA's requirements regarding economic LRs

Agency	Preferred analytical technique/scope/ objective	Approach to search strategy	Studies selection	Critical appraisal	Description of model structure	Model/perspective/time horizon/discounting/utility estimates/incremental analysis/costs/PSA/DSA
NICE	**Preferred analytical technique**: CUA **Scope/objective**: Clear description of the scope required Objective includes	Models: Economic evaluations may be based on new analyses. However, a review of published, relevant economic evaluations of	Manufacturer should the rationale for any inclusion and exclusion criteria used.	Each study's results should be interpreted with reference to a critical appraisal of its methodology.	Model structure description and diagram of the model submitted should be provided (including the	**Perspective**: on outcomes – all direct health effects, whether for patients or, when relevant, carers; on costs – NHS and PSS **Time horizon**: The time horizon for estimating clinical and cost effectiveness should be sufficiently

Agency	Preferred analytical technique/scope/ objective	Approach to search strategy	Studies selection	Critical appraisal	Description of model structure	Model/perspective/time horizon/discounting/utility estimates/incremental analysis/costs/PSA/DSA
	CEA studies relevant to decision making in the UK.	interventions should also be done. Search for economic evaluations using transparent and reproducible approaches until sufficient appropriate and relevant evidence has been identified. Reviews may not be exhaustive if additional studies identified would merely provide further support that is consistent with the already-identified evidence (rather than necessarily identifying all relevant studies). Once identified, critically assess economic evaluations using a suitable tool and assess external validity related to the decision problem. Clearly state and rationalise if no relevant economic	When studies have been identified and not included, justification for this should be provided.	Complete quality assessment for each relevant cost-effectiveness study identified should be provided. Use an appropriate and validated instrument.	following: type of analysis; justification of the chosen structure; how the model structure and its health states capture the disease or condition for patients; where appropriate, state the cycle length and whether a half-cycle correction has been applied). Features of the analysis should be presented in table. If there have been NICE technology appraisals in the same disease area, main inputs to the economic models accepted by appraisal committees should be summarized. If the model in this appraisal uses different inputs, rationale is needed. Manufacturer should compare	long to reflect all important differences in costs or outcomes between the technologies being compared **Discounting**: The same annual rate for both costs and health effects (currently 3.5%) **Utility estimates:** Health effects should be expressed in QALYs. The EQ-5D is the preferred measure of HRQoL in adults. For the reference case, the measurement of changes in HRQoL should be reported directly from patients and the utility of these changes should be based on public preferences using a choice-based method. **Incremental analysis**: is required. ICERs compared with baseline (usually standard care) and then incremental analysis, ranking technologies in terms of dominance and extended dominance. If the company has formally agreed a patient access scheme with the Department of Health, present the results of the base-case incremental cost-effectiveness analysis with the patient access scheme. **Costs**: All parameters used to estimate cost effectiveness should be presented clearly in a table with details of data sources. Relevant cost and healthcare resource use data for England should be presented. **PSA/DSA**: PSA is preferred for translating the imprecision in all input

Agency	Preferred analytical technique/scope/ objective	Approach to search strategy	Studies selection	Critical appraisal	Description of model structure	Model/perspective/time horizon/discounting/utility estimates/incremental analysis/costs/PSA/DSA
		evaluations are found. Utilities: Describe how systematic searches for relevant health-related quality-of-life data were done. Consider published and unpublished studies, including any original research commissioned for the technology. Provide the rationale for terms used in the search strategy and any inclusion and exclusion criteria used. The search strategy used should be provided in the appendix Costs: Describe how relevant cost and healthcare resource use data for England were identified. Include the search strategy and inclusion criteria, and consider published and unpublished studies to			and justify your chosen values with the methods specified by NICE in the reference case **Model:** Decision tree, Markov model, discrete event simulation model	variables into a measure of decision uncertainty in the cost effectiveness of the options being compared. In non-linear decision models, probabilistic methods provide the best estimates of mean costs and outcome. It should be identified which variables were subject to deterministic sensitivity analysis, how they were varied, and the rationale behind this. Only analyses when there is genuine uncertainty about a parameter should be reported. Results of deterministic sensitivity analysis should be presented, focusing on the key drivers of the model. Use of tornado diagrams should be considered.

Agency	Preferred analytical technique/scope/ objective	Approach to search strategy	Studies selection	Critical appraisal	Description of model structure	Model/perspective/time horizon/discounting/utility estimates/incremental analysis/costs/PSA/DSA
		demonstrate how relevant cost and healthcare resource use data for England were identified. The search strategy used should also be provided in the appendix. If the systematic search yields limited data for England, the search strategy may be extended to capture data from other countries.				
HAS	**Preferred analytical technique**: CEA (if health-related quality of life is not a major consequence, health outcome evaluated is the length of life), CUA (when health-related quality of life is a major consequence). The cost-benefit analysis is not recommended for the reference case. **Scope/objective :** Within the framework of HAS mandate, the prime	Required databases are not clearly specified. The values associated with health outcome parameters (effectiveness, tolerability, quality of life) and cost parameters should stem from a systematic and exhaustive research process that can cover numerous sources of data: systematic literature review, meta-analyses, use of medical or	The reference case analysis should identify all clinically relevant interventions in the population analyzed. The arguments supporting the inclusion or exclusion of an intervention in the reference case analysis should be	Clear, justified discussion should make it possible to estimate the robustness of the conclusion of the evaluation and define the conditions under which the conclusion would be altered. This discussion should rest on a critical appraisal of the methods and	The type of model and its structure should be defined to represent the patients' clinical progression and care, without introducing any needless complexity. The technical characteristics of the model should be suited to the specificities of the evaluation (mode of progression over time, degree of heterogeneity and interaction	**Perspective**: The reference case analysis uses a collective perspective, covering all individuals or institutions affected by the production of an intervention – whether in terms of health effects or cost – within the scope of the overall care provided. Otherwise, the choice of a healthcare system perspective in the reference case analysis should be duly justified. **Time horizon**: The choice of the time horizon should be based on a trade-off between the information produced by considering the expected or observed consequences of the intervention over the long term and the uncertainty generated by extrapolation over time. The time horizon used should be the same for all interventions. The two recommended options are: a lifetime horizon, i.e.,

Agency	Preferred analytical technique/scope/ objective	Approach to search strategy	Studies selection	Critical appraisal	Description of model structure	Model/perspective/time horizon/discounting/utility estimates/incremental analysis/costs/PSA/DSA
	objective of an economic evaluation is to guide public decision-making regarding the allocation of collective resources, by documenting the cost-effectiveness criteria.	administrative databases, clinical trials, observational epidemiological studies (cohort studies, case-control studies, cross-sectional studies), surveys, registries, etc	clearly set out. The inclusion of non-recommended interventions or the exclusion of recommended interventions should be justified in detail, on a case-by-case basis (e.g., context specificity, data availability, etc.). The interest of envisaging options such as non-medicinal therapeutic options or therapeutic abstention is reiterated.	data used, and on the sensitivity analyses conducted. Available data sources to document the effectiveness and tolerability of each intervention should undergo a stringent critical appraisal. The sources with the highest level of evidence should be selected Sample grid to evaluate the quality of an cost-effectiveness study Source: (adapted from Drummond et al.) is provided	among individuals, and degree of randomness) and comply with applicable recommendations. The structure of the model (statuses, events, links) should be defined so as to capture the costs and health outcomes associated with the progression of the disease and comparative care	until death; or a specific time horizon (e.g., to a defined age or over a defined period) **Discounting**: Future costs and health outcomes should be discounted to present values whenever the time horizon exceeds 12 months. The reference case analysis should use the public discount rate applicable at the time of the evaluation. On the date of publication of this guidance, that rate was 2.5% for time horizons of less than 30 years. Beyond that, the rate gradually decreases to a floor level set at 1.5%. In the reference case analysis, costs and health outcomes should be discounted using the same rate. **Utility estimates**: The utility scores used to weight life-years should be estimated using a multi-attribute approach, comprising the collection of health state data from patients via a generic questionnaire and the valuation of health states according to the preferences of the general population. Among available classification systems, EQ-5D-5L is recommended (French EQ-5D-5L value set and EQ-5D-5L questionnaire). Other approaches are not recommended for the base-case analysis of the reference case. **Incremental analysis**: two metrics may be used: ICER or net benefit (NB) expressed in monetary units (net monetary benefit – NMB) or

Agency	Preferred analytical technique/scope/ objective	Approach to search strategy	Studies selection	Critical appraisal	Description of model structure	Model/perspective/time horizon/discounting/utility estimates/incremental analysis/costs/PSA/DSA
						in health units (net health benefit – NHB). **Costs**: Only direct costs should be considered in the reference case analysis. A direct cost analysis may be presented as a supplemental analysis. The evaluation of the total cost of an intervention is based on the intervention's production costs. This requires the identification, measurement and valuation of the resources consumed. The scope of the cost items evaluated depends on the perspective adopted. Under a collective perspective, all resources consumed in the production of the overall patient care are taken into consideration. They cover the domestic sphere (e.g., informal care), the healthcare sphere (e.g., stays, procedures, and health products) and the medico-social sphere (e.g., stays, personal care services). Under the healthcare system perspective, the resources considered are those involved in the production of care (stays, procedures, and healthcare products). **PSA/DSA**: PSA based on a second-order Monte Carlo simulation is systematically required when the parameters' theoretical or empirical distributions are known or can be estimated. Depending on the number of comparators, an acceptability curve (two options compared) or multi-option acceptability curve (three or

Agency	Preferred analytical technique/scope/ objective	Approach to search strategy	Studies selection	Critical appraisal	Description of model structure	Model/perspective/time horizon/discounting/utility estimates/incremental analysis/costs/PSA/DSA
						more options) should be presented. Only parameters whose value is a random variable should be integrated in a PSA, Where the modelling methodology greatly hinders the use of a probabilistic analysis based on a second-order Monte Carlo simulation, this point should be clearly explained, and an appropriate methodology should be put forward. Univariate DSA (or multivariate ones if necessary) should systematically be conducted to identify the parameters that can have the greatest impact on the results of the evaluation. The choice of the parameters involved in a deterministic sensitivity analysis* should be justified. They may consist of parameters with a wide variation amplitude, parameters stemming from low-evidence studies, parameters relating to behaviors for which actions can be taken, etc. When a multivariate deterministic analysis is conducted, the author should explain the reasons behind the selection of the parameters. For the cases requiring a CEESP opinion on healthcare products, a deterministic analysis should be conducted on the price of the intervention being evaluated by testing at least three prices below the price used in the base-case analysis
IQWiG	Preferred analytical technique: Not specified; however, rationale	Data considered in the HEE to illustrate the provision of health	The choice of data basis, the size and relevant	NR	Not specified, but the choice needs to be justified.	**Perspective**: Depending on the commission, the following perspectives can be considered: the (pure) perspective of SHI, the

Agency	Preferred analytical technique/scope/ objective	Approach to search strategy	Studies selection	Critical appraisal	Description of model structure	Model/perspective/time horizon/discounting/utility estimates/incremental analysis/costs/PSA/DSA
	for the modelling technique adopted has to be provided	care, epidemiology, and costs, can be collected in various ways and originate from various sources. Analyses of secondary data should follow guidelines and recommendations on the good practice of secondary data analysis. If guidelines are used, they should be evidence-based and originate from the German health care system or from an Organization for Economic Co-operation and Development (OECD) country with health care structures comparable to Germany. If no evidence-based guidelines are available in the therapeutic area to be assessed, it is to be considered and presented transparently whether other	characteristics of the sample and the study population (incl. inclusion and exclusion criteria), the statistical methods, and the control of confounding factors, should be described transparently and justified. The generalizability and representativeness of results should be explained. The individual analysis steps must be comprehensible; plausibility checks should be ensured.			perspective of the community of SHI insurants (short: "SHI insurant perspective"), the social insurance perspective or the perspective of individual social insurance branches, as well as the societal perspective. In contrast to the pure SHI perspective, in the SHI insurant perspective the costs borne by the insurants (e.g. from co-payments) are also considered. Depending on the commission, it can be required for an HEE to consider the perspective of individual social insurance branches in addition to the SHI insurant perspective. The decision on whether further perspectives are included in an HEE depends solely on the question as to whether this is relevant for the decision maker. **Time horizon:** The time horizon must at least represent the average study duration and thus consider the differences in costs and benefits between the interventions of an HEE that are relevant for the decision. A longer time horizon should preferably be chosen in particular for chronic diseases. Costs and benefits should always be modelled over the same time horizon. The appropriate time horizon is often longer than the period covered by the available primary data from prospective studies. **Discounting**: If costs and benefits are incurred in periods lasting longer than a year, in the base case they are discounted after the first year with an

Agency	Preferred analytical technique/scope/ objective	Approach to search strategy	Studies selection	Critical appraisal	Description of model structure	Model/perspective/time horizon/discounting/utility estimates/incremental analysis/costs/PSA/DSA
		German guidelines can be used or whether expert surveys should be drawn upon. Expert surveys follow the generally recognized methods and procedures of quantitative social research. This means that in expert surveys, explicit information should be provided on the recruitment, number and expertise of experts, the research question, individual answers (not only mean values), the manner of achieving a consensus, as well as on the presentation and handling of results. Price catalogues or lists must be current and represent the prices relevant for the SHI. There are very different health economic questions for which focused information				identical constant rate of 3% to the current period. Likewise, identical constant rates of 0 and 5% should be used in sensitivity analyses; any deviations must be justified. Additionally, if cost data originate from different time periods, they must be adjusted for inflation. **Utility estimates**: If generic index instruments are used, a scale validated in Germany must be used for the determination of the utility value. The use of QALYs, as well as their determination and conversion into a German scale, must in each case be presented in a comprehensible manner and justified. Apart from that, all usual standards for the respective procedures and instruments apply: i.e. evidence of objectivity, reliability, validity, and responsiveness must be available. Parallel to the use of a generic instrument, disease-specific instruments to determine quality of life in clinical studies should be applied. The mapping of disease-specific to generic instruments is therefore discouraged. **Costs**: In general, when determining costs it should be evaluated in each perspective whether these costs and, if applicable, cost savings, are relevant for the interventions, therapeutic areas, and patient groups investigated. **PSA/DSA**: Parameter uncertainty as well as types of uncertainty that

Agency	Preferred analytical technique/scope/ objective	Approach to search strategy	Studies selection	Critical appraisal	Description of model structure	Model/perspective/time horizon/discounting/utility estimates/incremental analysis/costs/PSA/DSA
		retrieval is to be conducted. These include the search for health economic evaluations, decision analytic models, the measure of overall benefit, cost determination, as well as, if necessary, epidemiologic data, if no data from Germany are available. A search is conducted in at least the bibliographic database MEDLINE. In addition, manufacturers can be asked to provide health economic evaluations.				cannot be reduced are quantified. The Institute considers both univariate and multivariate deterministic as well as PSA, and in its work follows the recommendations of the conjoint Modeling Good Research Practices Task Force Working Group of ISPOR and the Society for Medical Decision Making (SMDM). All analyses performed for this purpose should be fully documented, with minimum and maximum values for the parameters used and underlying assumptions. The following aspects must be specified for PSAs: probability distributions used and their sources, correlations between input parameters, and any structural variants. Structural sensitivity analyses are performed to investigate the impact of a variation of assumptions in the model structure, for example, the number or type of the model states.
ICER	**Preferred analytical technique**: CUA, CEA; As a measure of cost-effectiveness, ICER's voting panels follow common academic and health technology assessment standards by using the cost per QALY gained as the	ICER aims to use data from published or publicly available sources, including peer-reviewed journals, supplementary appendices, briefing documents used by regulatory authorities, and conference proceedings. In	NR	NR	The type of model is described, including a textual and/or graphic depiction of the model structure, process, and outputs. The links between each intervention and its short-term effects, long-term effects, and final	**Perspective**: The economic model contextualizes the evidence from a US health system perspective, which includes estimates of the costs and outcomes associated with each intervention. **Time horizon** (typically lifetime) used in primary analyses are specified. For potential BIA five-year time horizon is explored. With a five-year timeframe, some of the potential cost-offsets provided by new interventions are captured.

Agency	Preferred analytical technique/scope/ objective	Approach to search strategy	Studies selection	Critical appraisal	Description of model structure	Model/perspective/time horizon/discounting/utility estimates/incremental analysis/costs/PSA/DSA
	primary measure of cost-effectiveness as well as the equal value life years gained (evLYG), with formal voting on "long-term value for money". **Scope/objective :** The decision problem (i.e., the specific question to be addressed from the analysis) is specified in terms of the overall objective, the interventions of interest and their key comparators, the relevant population groups and subgroups being considered, and the outcomes	addition, because life science companies may have relevant information that is currently held in confidence, ICER has structured a process to accept and use such data. Results from the evidence review, including the results from any meta-analysis, are used to inform input parameters when possible.			outcomes are described, including the health states and sequence of possible transitions between them. Typical choices for model structure include decision trees, Markov cohort, or microsimulation models. Key assumptions regarding the model structure are articulated with their corresponding rationale. The economic modeling team identifies the software used to create the model and run analyses, which may include statistical programming software (e.g., R), decision analysis software (e.g., TreeAge), or Microsoft Excel	**Discounting**: Discount rates for costs and outcomes (typically 3% per year) are specified. **Utility estimates**: NR **Incremental analyses:** Alternative interventions are compared in terms of cost per unit of effectiveness, and the resulting comparison is presented as a cost-effectiveness ratio. As with comparative clinical effectiveness, relative certainty in the cost and outcome estimates is also a consideration. As a measure of cost-effectiveness, ICER's voting panels follow common academic and health technology assessment standards by using the cost per QALY gained as the primary measure of cost-effectiveness as well as the equal value life years gained (evLYG), with formal voting on "long-term value for money". **Costs**: Costs are reported in terms of total and incremental costs. **PSA/DSA**: Expected values of costs and outcomes for each intervention are also estimated through "probabilistic sensitivity analysis", which captures some of the uncertainty in the input parameter estimates. This type of analysis takes repeated samples, typically 1,000 or more, from the (joint) distribution of all key model input parameters simultaneously; results are presented tabularly in terms of the percentage of simulations that achieve $50,000,

Agency	Preferred analytical technique/scope/ objective	Approach to search strategy	Studies selection	Critical appraisal	Description of model structure	Model/perspective/time horizon/discounting/utility estimates/incremental analysis/costs/PSA/DSA
						$100,000, $150,000, and $200,000 per QALY thresholds, and graphically using scatter plots and cost-effectiveness acceptability curves, which reflect the percentage of simulations that result in incremental cost-effectiveness ratios that fall at or under different cost-effectiveness thresholds. The economic modeling team also conducts one-way sensitivity analyses and presents the results in "tornado diagrams" that display the findings across a feasible range for each input parameter estimate. A table contains the distributional assumptions about the input parameters varied. Specific scenario analyses (e.g., using a modified societal perspective that incorporates estimates such as productivity losses, caregiver burden, and other indirect costs) and subgroup analyses are conducted when appropriate. In addition, the economic modeling team reports results from long-term cost-effectiveness threshold analyses, which estimate the intervention costs or prices that correspond to commonly-cited thresholds extending from $50,000 per QALY and evLYG to $200,000 per QALY and evLYG.
CADTH	**Preferred analytical technique**: In the reference case, the economic evaluation	NR	The selection of data sources for health state utility	NR	Model conceptualization and development should address the decision problem.	**Perspective:** In the reference case, the perspective should be that of the publicly funded health care payer. The perspective of the economic

Agency	Preferred analytical technique/scope/ objective	Approach to search strategy	Studies selection	Critical appraisal	Description of model structure	Model/perspective/time horizon/discounting/utility estimates/incremental analysis/costs/PSA/DSA
	should be a CUA with outcomes expressed as QALYs. CEA with outcomes expressed in natural units is not an appropriate reference case. If convincing evidence is available to show that important patient outcomes are equivalent on virtually all measures, except for survival or quality of life, then a CUA remains the appropriate approach. This allows for the necessary comparison, using the same benefit metric, across all the technologies being considered. CMA is not an appropriate reference case analysis. A cost-consequence analysis (CCA) should be viewed as a complement to, and not a substitute for, a CUA. A CCA		values should be based on their fitness for purpose, credibility, and consistency. One should describe the trade-offs among these criteria and provide justification for the selected sources. When valuing resources, researchers should select data sources that most closely reflect the opportunity cost, given the perspective of the analysis. Fees and prices listed in schedules and formularies of Canadian ministries of		The model should be consistent with the current understanding of the clinical or care pathway for the health condition and the interventions being compared. The scope, structure, and assumptions should be clearly described and justified. Researchers should consider any existing well-constructed and validated models that appropriately capture the clinical or care pathway for the condition of interest when conceptualizing their model. The choice of modelling technique should be justified. The approach should be no more complex than is necessary to	evaluation should be related to the decision problem **Time horizon**: Time horizon should be based on the condition and the likely impact of the intervention **Discounting**: In the reference case, costs and outcomes that occur beyond one year should be discounted to present values at a rate of 1.5% per year. The impact of uncertainty in the discount rate should be assessed by comparing the results of the reference case to those from non-reference case analyses, using discount rates of 0% and 3% per year. **Utility estimates**: Health preferences should reflect the health states in the model and be conceptualized to address the decision problem. Health preferences should reflect the general Canadian population. In the reference case, researchers should use health preferences obtained from an indirect method of measurement that is based on a generic classification system (e.g., EQ-5D, HUI, SF-6D). Researchers must justify where an indirect method is not used. **Incremental analysis**: The economic evaluation should be assessed based on the ICER. Estimates of net monetary benefit may also be provided. **Costs**: Resource use and costs should be based on Canadian sources and reflect the jurisdiction(s) of

Agency	Preferred analytical technique/scope/ objective	Approach to search strategy	Studies selection	Critical appraisal	Description of model structure	Model/perspective/time horizon/discounting/utility estimates/incremental analysis/costs/PSA/DSA
	aids in the transparency of the reporting of an economic evaluation, as disaggregated results are presented in terms of costs and outcomes (e.g., events predicted, survival, gains in quality of life). Where there are important health outcomes from a technology that cannot be captured in a CUA, then these should be reported as additional components within a CCA. If such outcomes can be valued in monetary terms then, additionally, a CBA can be undertaken as a non-reference case analysis, with full details provided on the derivation of monetary values for all outcomes included in the evaluation, or justification for why		health are recommended as unit-cost measures when considering the perspective of the public payer, as long as they reflect actual payments. In other instances, total average costs (including capital and allocated overhead costs) may be relevant. Where costs are directly calculated or imputed, they should reflect the full economic cost borne by the decision-maker		address the decision problem	interest (as specified in the decision problem). When valuing resources, researchers should select data sources that most closely reflect the opportunity cost, given the perspective of the analysis. Fees and prices listed in schedules and formularies of Canadian ministries of health are recommended as unit-cost measures when considering the perspective of the public payer, as long as they reflect actual payments. In other instances, total average costs (including capital and allocated overhead costs) may be relevant. Where costs are directly calculated or imputed, they should reflect the full economic cost borne by the decision-maker. **PSA/DSA:** In the reference case, uncertainty regarding the value of each parameter should be examined through probabilistic analysis. The final results should be based on expected costs and expected outcomes. These should be estimated through a probabilistic analysis.

Agency	Preferred analytical technique/scope/ objective	Approach to search strategy	Studies selection	Critical appraisal	Description of model structure	Model/perspective/time horizon/discounting/utility estimates/incremental analysis/costs/PSA/DSA
	the outcomes were excluded. **Scope/objective**: The decision problem addressed by the economic evaluation should be clearly stated					
PBAC	**Preferred analytical technique**: CEA/CUA, or CMA: A full CEA is appropriate where the clinical evaluation has concluded that the proposed medicine is: therapeutically superior to the main comparator, but likely to result in additional costs to the health system; or therapeutically inferior to the main comparator, but likely to result in lower costs to the health system. A CUA is preferred over a CEA, particularly where: there is a claim of incremental life-years gained in the economic	Model: Present the results of a literature search for economic evaluations involving the proposed and similar medicines or alternative managements, or similar treatment algorithms, focusing on the structure of the existing models. This may include published reports and models considered by other health technology assessment agencies. Utilities: Where utility weights or QALY changes cannot be directly estimated from data collected in the clinical studies from Section 2, or there	Manufacturer should review the literature for relevant economic references and any additional clinical or epidemiological literature relevant to the model. Any additional literature (eg additional clinical trials, guidelines, natural history studies, burden of disease studies, utility studies, surveys) that informs the	NR	Submitting party should ensure that the model structure captures all relevant health states or clinical events along the disease or condition pathway, and that it is consistent with the treatment and disease or condition algorithms. Model structure needs to be informed using the results of the literature review of economic evaluations, and other clinical and economic literature, including clinical trials, clinical guidelines, natural	Health care system perspective should be used to inform the base-case analysis. The PBAC's preferred health care system perspective includes health and health-related resource use (costs and cost offsets), and health-related outcomes. Costs include those incurred by the patient, and the public or private health care provider; outcomes are those associated with the patient. Costs and outcomes that are not specifically related to 'health and/or provision of health care' should not be included in the base case. To show a broader societal perspective and quantitatively incorporate considerations beyond the patient and the health care system, supplementary analysis should be presented in addition to the base case. A well-justified and well-supported analysis will form a more compelling case. Where interventions do not affect mortality and have temporary health or quality-of-life effects, a relatively short time horizon may be appropriate. Where there is evidence that a treatment affects mortality or

Agency	Preferred analytical technique/scope/ objective	Approach to search strategy	Studies selection	Critical appraisal	Description of model structure	Model/perspective/time horizon/discounting/utility estimates/incremental analysis/costs/PSA/DSA
	evaluation (to assess the impact of quality adjusting that survival gain); or there is an improvement in quality, but not quantity, of life; or relevant direct randomised trials report results using a multiattribute utility instrument. A cost-minimisation approach is appropriate where there is a therapeutic claim of noninferiority, the safety profile is equivalent or superior and use of the proposed medicine is anticipated to result in equivalent or lesser costs to the health system. **Scope/objective :** NR	are significant concerns about the reliability and relevance of trial-based utility, transform the Section 2 health outcomes to estimate QALY gains (eg by applying utility weights to the time spent in different health states that represent the experience of clinical outcomes). Additional studies (either published or done for the submission) may be needed to estimate utility weights for health states in the economic model. These studies should be identified. Where available, use the source of costs recommended by the Manual of resource items and their associated costs	model structure or inputs needs to be presented. Utility estimates may be available from the literature. The validity of the derived utility weights depends on the applied elicitation methods and the relevance of the study populations. Present details of search strategies, and inclusion and exclusion criteria used to identify relevant utility studies. Assess the validity of all identified studies, including:		history studies and burden of disease studies. Patient-relevant events should be disaggregated if there are important differences in mortality, disease or condition progression, associated costs, or QoL effects, and the distribution differs between the intervention and comparator. During the model development, it should be considered whether, for a given patient, an event experienced in the model should influence the risk of experiencing subsequent events – this may inform the choice of computational method.	long-term/ongoing quality of life, then a lifetime time horizon is appropriate. Discounting: 5% per year for all costs and health outcomes that occur or extend beyond one year in the base case, and fixed discount rates of 3.5%, and 0% per year (applied to both costs and outcomes) in sensitivity analysis. If relevant, supplementary analyses using other discounting methodologies (eg a different uniform rate, differential rates, time-varying rates) should be used and justified. If available, trial-based quality-of-life or utility data should be used to estimate QALYs in the model, or the use of alternative indirect methods to estimate QALYs should be justified when direct data are available. No preference is stated for a specific MAUI. If a MAUI has been used in an included study to estimate utility weights, it should be stated where and when the scoring algorithm was derived, and considered how applicable it is to the general Australian population. It is preferred that Australian-based preference weights are used in the scoring algorithm used to calculate utility weights. The cost-outcome analysis may use two metrics: ICER or NB expressed in monetary units (NMB) or in health units (NHB). Health care resource costs including those incurred by the

Agency	Preferred analytical technique/scope/ objective	Approach to search strategy	Studies selection	Critical appraisal	Description of model structure	Model/perspective/time horizon/discounting/utility estimates/incremental analysis/costs/PSA/DSA
			-how representative the health state in each identified study is of the health state in the economic evaluation (including the type and severity of symptoms, and the duration of the health state) -how the health state was captured (eg MAUI, scenario based) -how the preference was elicited (eg standard gamble, time trade-off) -what sample was chosen to respond to the MAUI questionnaire or scenario (eg the general		Assess the model structure(s) to establish face validity. Justify the exclusion of any potentially relevant states or events identified in the literature, and reference data sources and expert input. Discuss the potential impact of any exclusions on the model outputs. Where the model structure differs from existing models, explain the basis for the selection of the submission's approach. If relevant, define multiple plausible model structures and test them as part of a structural sensitivity analysis.	patient, and the public or private health care provider. Univariate deterministic sensitivity analyses should be performed for all uncertain input parameters, or natural groups of input parameters (e.g. cost or utility weights for all target clinical outcomes). The following requests are based on good-practice guidelines for model parameter estimation and uncertainty analysis. PSA may be provided in addition to the DSA.

Agency	Preferred analytical technique/scope/ objective	Approach to search strategy	Studies selection	Critical appraisal	Description of model structure	Model/perspective/time horizon/discounting/utility estimates/incremental analysis/costs/PSA/DSA
			public, patients, carers, health care professionals) -what assessment was made of the nature and direction of bias that might arise, given the sample and methods -how the sensitivity analyses examined variation in the identified utility options. Using different published studies to inform utility weights for alternative health states is discouraged because of the potential for inconsistency			

Agency	Preferred analytical technique/scope/ objective	Approach to search strategy	Studies selection	Critical appraisal	Description of model structure	Model/perspective/time horizon/discounting/utility estimates/incremental analysis/costs/PSA/DSA
			in the methods and populations from which utilities were derived.			
C2H	**Preferred analytical technique**: CEA should be used (defined as an analysis that calculates cost and effectiveness separately without converting effectiveness into monetary units). A cost comparison with the comparator (CMA) should be performed if the analysis based on results of SLR does not demonstrate an additional benefit compared to the comparator, but the outcome of the selected technology appears equivalent to that of the comparator. **Scope/objective:** NR	The selection of effectiveness, safety, and QOL data on the basis of an SLR of Japanese and overseas clinical research is recommended. This review may also include unpublished clinical study/trial data if deemed appropriate. It is recommended that claims databases established in Japan, which reflect actual clinical practice should be used to estimate the costs of each health state, if deemed appropriate. However, this recommendation does not apply to cases in which it is difficult to define health states using only information	Data with a high evidence level should be used preferentially. The use of data deemed appropriate from the viewpoints of research quality, target population, and external validity is recommended Data by re-analysis of existing study and/or registry data can be used if deemed appropriate. In that case, detailed information		All parameters and data sources used for model creation should be shown. The assumption used to create the model should be described clearly. The model used and the calculation processes should be submitted in the form of electronic files. Model analysis should present the validity of the model Decision analytic model, Markov model, and/or other models. Half-cycle correction should be used in the Markov model if the length of the Markov cycle is long and its	**Perspective**: Public healthcare payer's perspective is a standard perspective that pertains to factors such as costs, comparator(s), and target populations within the range of the public healthcare insurance in Japan. Even when an analysis is conducted from a perspective other than the "public healthcare payer's perspective," an analysis from the "public healthcare payer's perspective" should also be submitted. There are some healthcare technologies that are not covered by the public healthcare insurance but are publicly funded, such as some prophylactic procedures (e.g., health check-ups, vaccinations). Analyses including these technologies should be submitted from the "public healthcare payer's perspective. If the effect on public long-term care costs is important with regard to the selected technology, it is acceptable to perform an analysis from the "public healthcare and long-term care payer's perspective." **Time horizon**: should be sufficiently long to evaluate the influence of the technology on cost and

Agency	Preferred analytical technique/scope/ objective	Approach to search strategy	Studies selection	Critical appraisal	Description of model structure	Model/perspective/time horizon/discounting/utility estimates/incremental analysis/costs/PSA/DSA
		from claims data, insufficient data have been accumulated in the database, and so on. .	on patient background, statistical methods, etc. must be provided. Japanese data should be used preferentially if there is evident heterogeneity between Japanese and overseas data. If the data do not differ statistically significantly between the selected technology and the comparator(s), pooled data of both groups should be applied. Regarding evaluation of medical devices, if there are reliable and		influence on the results is not negligible	effectiveness. The same time horizon should be applied for both cost and effectiveness. The reason for setting this time horizon should be specified. **Discounting**: Both cost and effectiveness should be discounted at a rate of 2% per year. The discount rate should be subjected to sensitivity analysis and should be changed at the same rate of 0–4% per year for both costs and effectiveness. Discounting is not needed if the time horizon is one year or less or is otherwise sufficiently short to ignore the influence of discounting **Utility estimates:** QOL data (inclusion of parameters such as transition probability for model analysis) derived from high-quality research, with a high evidence level reflective of practical clinical results in Japan. The EQ-5D is one currently available measure for which a scoring algorithm has been developed in Japan. **Incremental analysis:** If the analysis allows a judgment that reveals additional benefit, ICER should be calculated from the expected cost and effectiveness in each treatment group. **Costs**: Only public healthcare costs should be included in the case of analysis from the public healthcare payers' perspective. Healthcare costs of each health state include only related costs that are directly affected

Agency	Preferred analytical technique/scope/ objective	Approach to search strategy	Studies selection	Critical appraisal	Description of model structure	Model/perspective/time horizon/discounting/utility estimates/incremental analysis/costs/PSA/DSA
			quantitative data, analysis reflecting learning effect (i.e., improvement of treatment effect by the accumulation of clinicians experience) or product improvement effect can be submitted in addition to analysis not considering the effects, upon agreement in consultation			by the selected technology and do not include unrelated costs. Healthcare costs of each health state should reflect the average resource consumption and standard clinical practices in Japan. Unit costs should be derived from the latest medical fee schedule, drug price lists, or similar resources. It is particularly essential to use the latest unit costs for the selected technology or comparator(s). The estimation should include not only the costs of the selected technology and the comparator(s) but also the costs of factors such as adverse events and related future events. An analysis of the public healthcare costs should include not only the portion of costs paid by the insurer but also those paid by the government and patients as copayment. Public long-term care costs and productivity losses arising from an inability to perform work should not be included in the base-case analysis. It is acceptable to include public long-term care costs and productivity losses in additional analyses only if they can be estimated by Japanese data. However, judgments regarding the appropriateness of including productivity losses should consider the possibility of working in the context of the illness characteristics **PSA/DSA**: PSA is desirable. In such cases, the distribution used for analysis, scatter plots of the cost-

Agency	Preferred analytical technique/scope/ objective	Approach to search strategy	Studies selection	Critical appraisal	Description of model structure	Model/perspective/time horizon/discounting/utility estimates/incremental analysis/costs/PSA/DSA
						effectiveness plane, and cost-effectiveness acceptability curves (CEAC) must be presented. If the patterns of clinical practice or other factors are not uniform and this discrepancy could affect the results, analyses based on multiple scenarios should be conducted. For situations in which the uncertainty is large because of a long time horizon, a shorter-term analysis is necessary, such as an analysis of the period for which clinical study data are available. If no available studies involve a comparison with the comparator, particularly when a comparison has been made concerning results between single-arm studies, a sensitivity analysis with a sufficiently wide range is required because of the large uncertainty. Sensitivity analyses are needed for parameters with large variances, those based on assumptions rather than actual data, those with possible heterogeneity between overseas and domestic data, and others. When the variance of the estimator should be considered (parametric uncertainty), the range moving parameter in the sensitivity analysis can refer to the 95% confidence interval of the estimator.

8.2. 3RRs methodologies vs Cochrane's SLRs

8.2.1. Search strategies for 3 RRs

Table 16 RR_1 - search strategy (Searched on 19th October 2022; Database: MEDLINE)

#	Searches	Results
1	Obesity/	215069
2	Adiposity/	15383
3	Body Weight Changes/	7
4	Overweight/	33180
5	Body Weight.ti,ab,kw.	238897
6	Weight Loss.ti,ab,kw.	107212
7	(overweight* or over weight*).ti,ab,kw.	88447
8	adipos*.ti,ab,kw.	133035
9	obes*.ti,ab,kw.	382448
10	(weight adj3 (cyc* or reduc* or los* or maint* or decreas* or control* or gain* or chang* or improv* or modif* or exessiv*)).ti,ab,kw.	271003
11	or/1-10	851804
12	Fluoxetine/	9917
13	fluoxetin*.ti,ab,kw.	13727
14	(prozac or sarafem).ti,ab,kw.	469
15	or/12-14	15664
16	Randomized controlled trials as Topic/	162372
17	Randomized controlled trial/	594048
18	randomized.ti,ab.	658257
19	or/16-18	1012913
20	11 and 15 and 19	409

Table 17 RR_2 - search strategy (Searched on 20th October 2022; Databases: Embase, MEDLINE)

#	Searches	Results
1	Melatonin/ or melatonin.mp.	75599
2	exp anxiety/	369122
3	(anxiet* or Angst or Hypervigilance or Nervousness or Anxiousness or worry or apprehens* or consternation or uneas* or fearfulness or fear).mp.	950718
4	preoperative period/ or preoperative care/ or exp anesthesia recovery period/ or premedication/ or postoperative period/ or postoperative care/ or (preoperat* or pre operat* or pre procedur* or preprocedur* or before surg* or pre surg* or presurg* or postoperat* or post operat* or post surg* or postsurg* or post procedur* or postprocedur*).mp.	3037563
5	or/2-4	3953049
6	(Randomized Controlled Trial or Controlled Clinical Trial or Pragmatic Clinical Trial or Equivalence Trial or Clinical Trial, Phase III).pt.	674096
7	Randomized Controlled Trial/	1312262
8	exp Randomized Controlled Trials as Topic/	399077
9	randomi?ed controlled trial*.tw.	528987
10	RCT.tw.	78057
11	randomi?ed trial.tw.	145973
12	((randomized or placebo).ab. or drug therapy.fs. or randomly.ab. or trial.ab. or groups.ab.) not (animals not (humans and animals)).sh.	12822399
13	(crossover procedure or double blind procedure or single blind procedure or crossover* or cross over* or placebo* or (doubl* adj blind*) or (controlled adj3 (study or design or trial))).tw.	1281968
14	or/6-13	13258408
15	1 and 5 and 14	2211
16	remove duplicates from 15	1863

#	Searches	Results
17	limit 16 to yr="2021 -Current"	322
18	16 not 17	1615

Table 18 RR_3 - search strategy (Searched on 19th October 2022; Database: MEDLINE)

#	Searches	Results
1	Respiratory Distress Syndrome/	23715
2	Acute Respiratory Distress Syndrome.mp.	21167
3	((Shock adj2 Lung) or ARDS or Respiratory Distress Syndrome*).mp.	56592
4	((severe or hypoxic) adj4 (respiratory and failure)).mp.	9183
5	(acute adj (lung injur* or distress syndrome)).mp.	17937
6	or/1-5	75195
7	Enteral Nutrition/ and Immune System/	61
8	((dietary or nutrient*) adj3 (modulation or immune system)).mp.	1076
9	dietary nutrient*.tw.	1322
10	immunonutrition.tw.	690
11	antioxidants/ or antioxidant*.tw.	286620
12	Vitamin E/ or Vitamin E.tw.	40442
13	Carotenoids/ or carotenoid*.tw.	33789
14	Ascorbic Acid/ or Ascorbic Acid.tw.	60393
15	Selenium/ or selenium.tw.	37377
16	Zinc/ or zinc.tw.	154256
17	exp Amino Acids, Essential/ or Glutamine/ or Arginine/	241157
18	exp Fatty Acids, Essential/	71419
19	Fatty Acids, Omega-3/	15520
20	(omega-3 or omega3).tw.	18820
21	(beta?caroten* or glutamine or arginine).tw.	139065
22	or/7-21	954810
23	(Randomized Controlled Trial or Controlled Clinical Trial or Pragmatic Clinical Trial or Equivalence Trial or Clinical Trial, Phase III).pt.	674063
24	Randomized Controlled Trial/	579156
25	exp Randomized Controlled Trials as Topic/	162170
26	randomi?ed controlled trial*.tw.	230931
27	RCT.tw.	28935
28	randomi?ed trial.tw.	61002
29	((randomized or placebo).ab. or drug therapy.fs. or randomly.ab. or trial.ab. or groups.ab.) not (animals not (humans and animals)).sh.	4684685
30	or/23-29	4852504
31	6 and 22 and 30	599
32	limit 31 to yr="2019 -Current"	190
33	31 not 32	409

8.2.2. RR_1: outcomes of meta-analyses

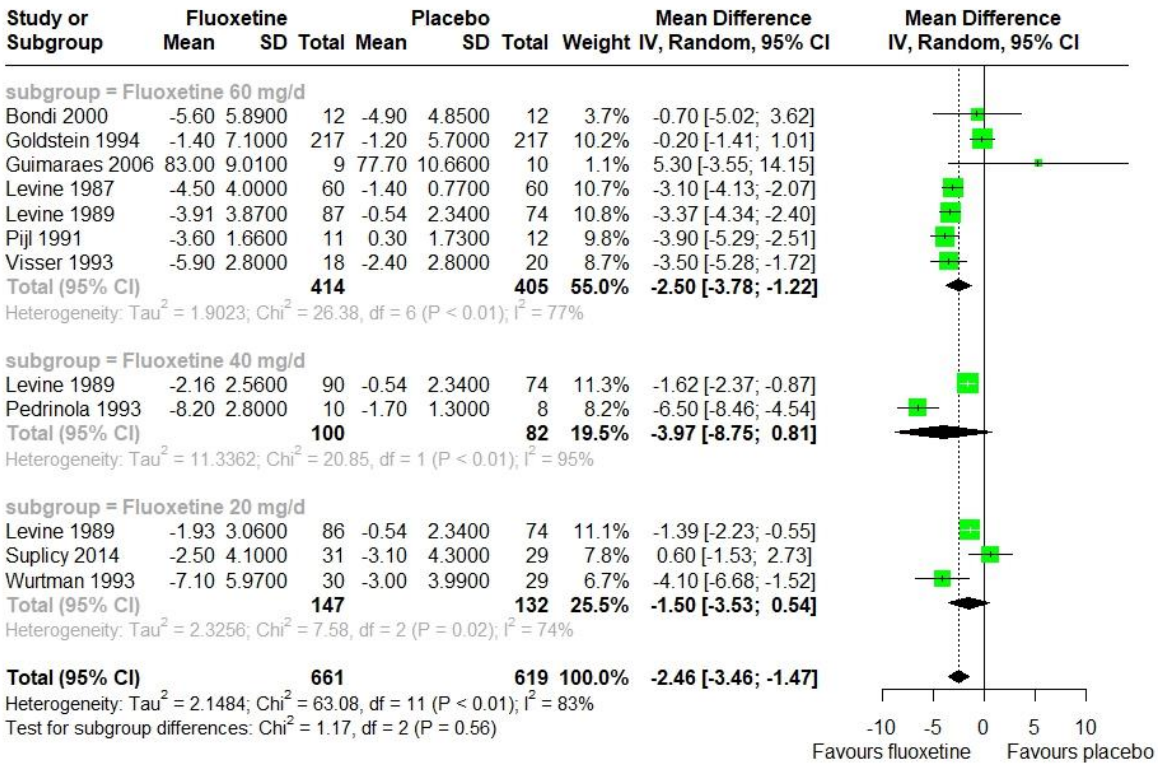


Figure 16 Analysis 1.1. Comparison 1: Fluoxetine versus placebo, Outcome 1: Weight loss (end of trial weight)(Serralde-Zúñiga, Gonzalez Garay et al. 2019) - without Levine 1987, Pedrinola 1993 and Wurtman 1993

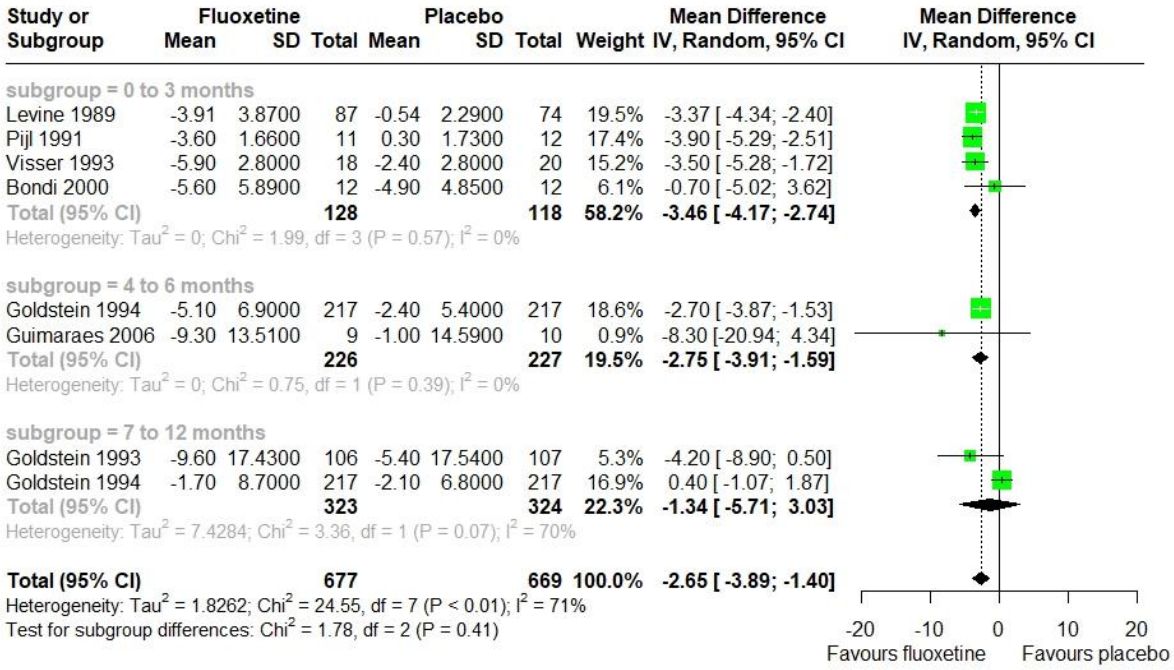


Figure 17 Analysis 1.2. Comparison 1: Fluoxetine versus placebo, Outcome 2: Weight loss (duration of intervention) - without Levine 1987

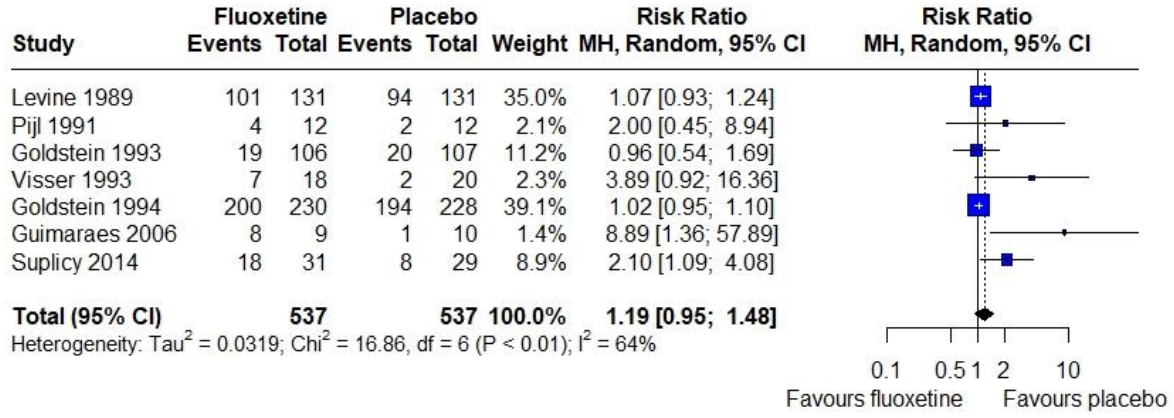


Figure 18 Analysis 1.4. Comparison 1: Fluoxetine versus placebo, Outcome 4: Any adverse event - without Levine 1987 and Wurtman 1993

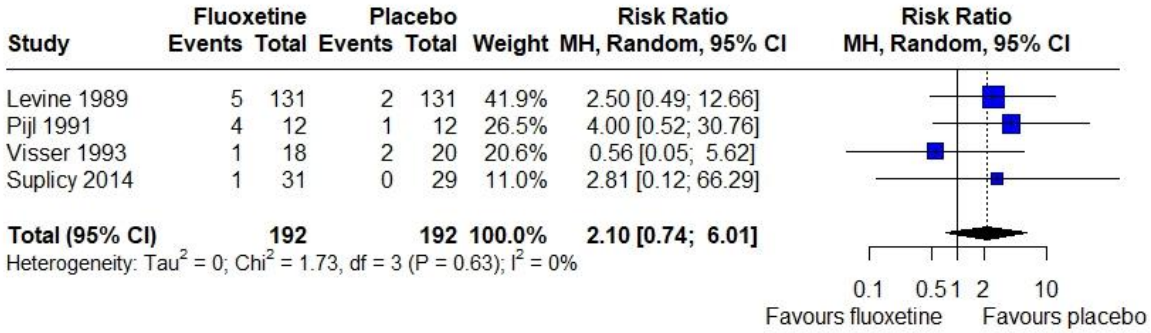


Figure 19 Analysis 1.5.1 Comparison 1: Fluoxetine versus placebo, Outcome 5: Specific adverse events, abdominal pain - without Levine 1987

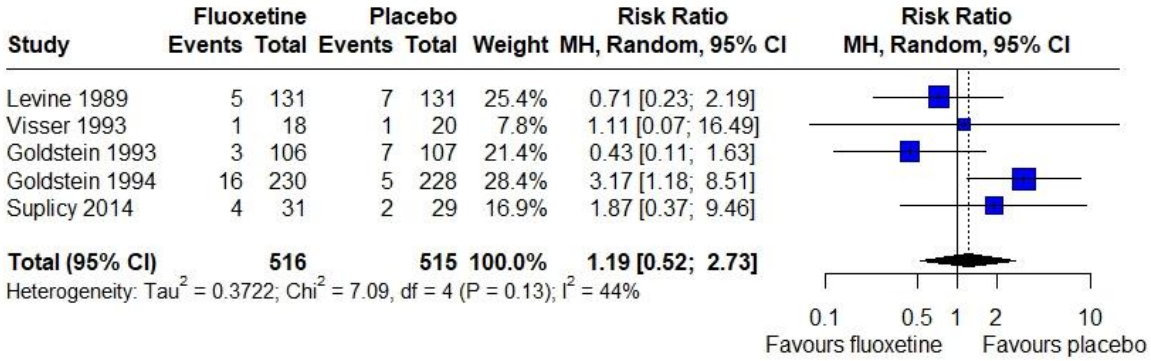


Figure 20 Analysis 1.5.5 Comparison 1: Fluoxetine versus placebo, Outcome 5: Specific adverse events, anxiety - without Levine 1987 and Wurtman 1993

Figure 21 Analysis 1.5.6 Comparison 1: Fluoxetine versus placebo, Outcome 5: Specific adverse events, constipation - without Wurtman 1993

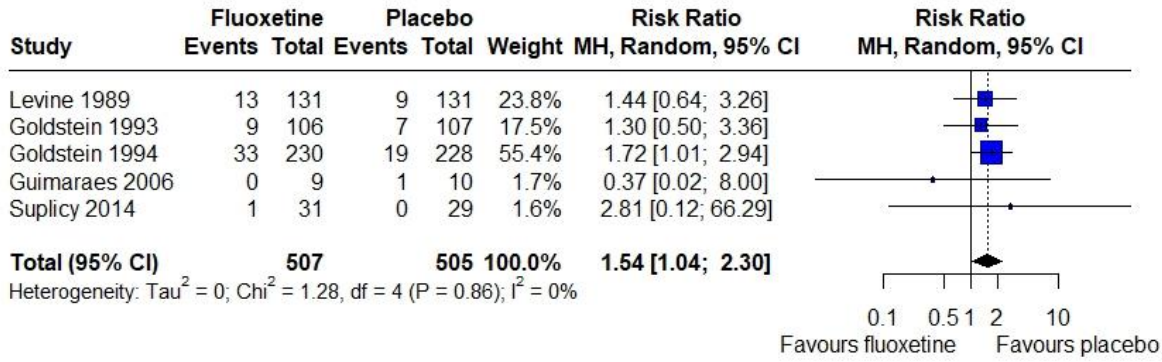


Figure 22 Analysis 1.5.7 Comparison 1: Fluoxetine versus placebo, Outcome 5: Specific adverse events, diarrhoea – without Levine 1987 and Wurtman 1993

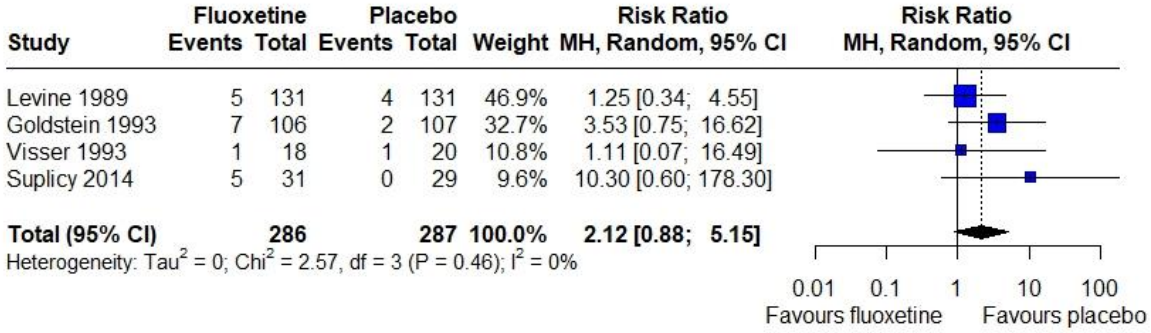


Figure 23 Analysis 1.5.8 Comparison 1: Fluoxetine versus placebo, Outcome 5: Specific adverse events, dizziness – without Levine 1987

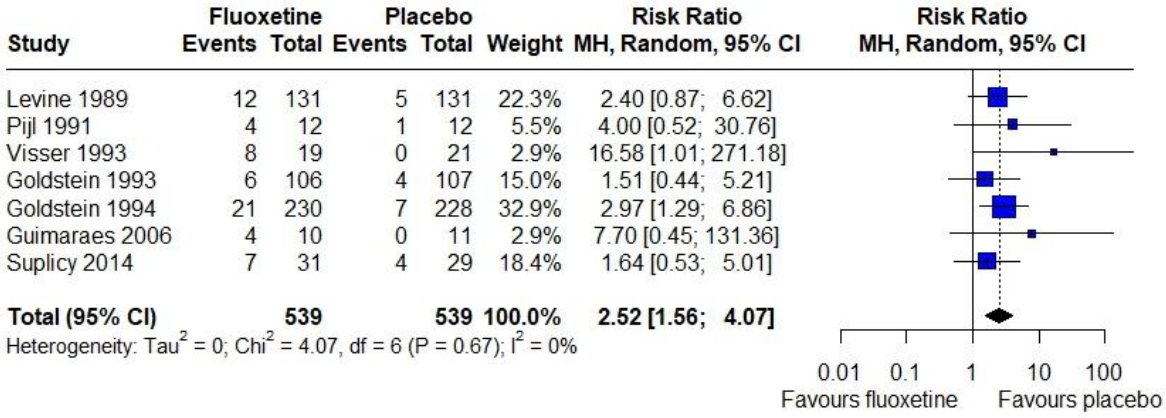


Figure 24 Analysis 1.5.9 Comparison 1: Fluoxetine versus placebo, Outcome 5: Specific adverse events, drowsiness – without Levine 1987 and Wurtman 1993

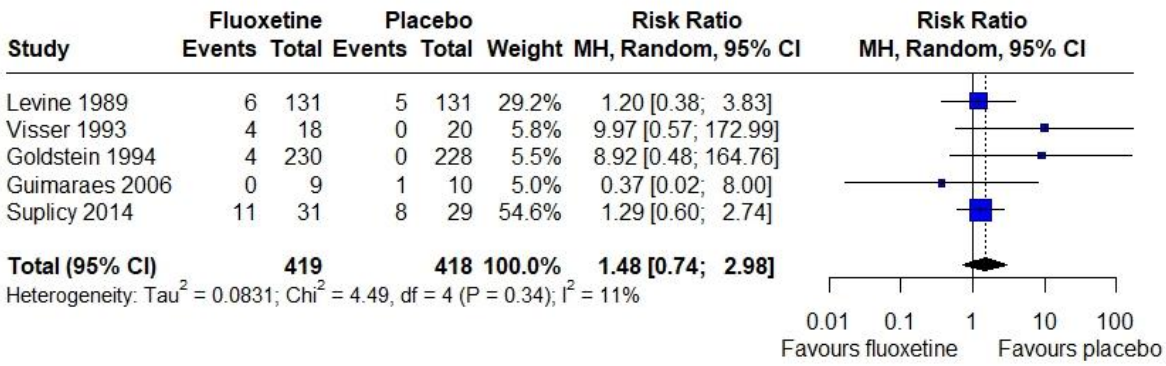


Figure 25 Analysis 1.5.10 Comparison 1: Fluoxetine versus placebo, Outcome 5: Specific adverse events, dry mouth – without Wurtman 1993

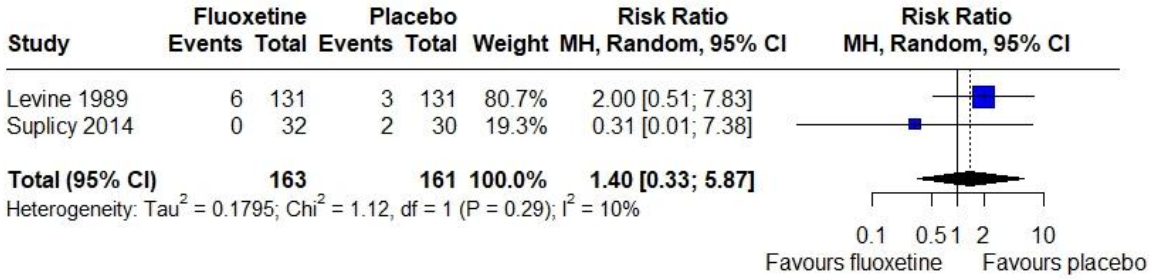


Figure 26 Analysis 1.5.11 Comparison 1: Fluoxetine versus placebo, Outcome 5: Specific adverse events, dyspepsia – without Levine 1987 and Wurtman 1993

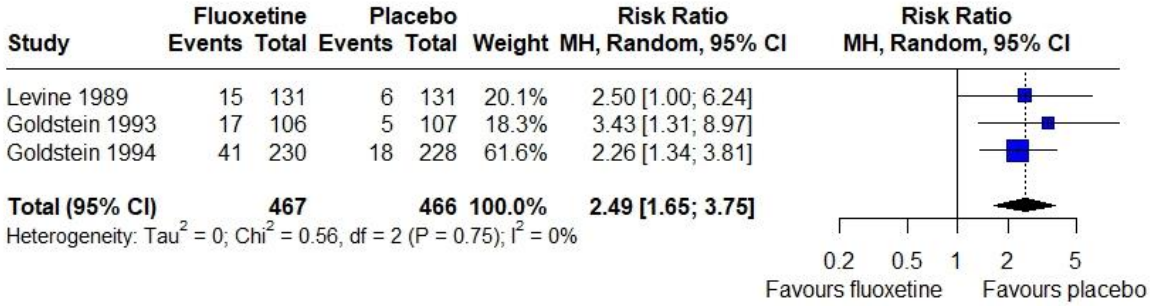


Figure 27 Analysis 1.5.12 Comparison 1: Fluoxetine versus placebo, Outcome 5: Specific adverse events, fatigue – without Levine 1987 and Wurtman 1993

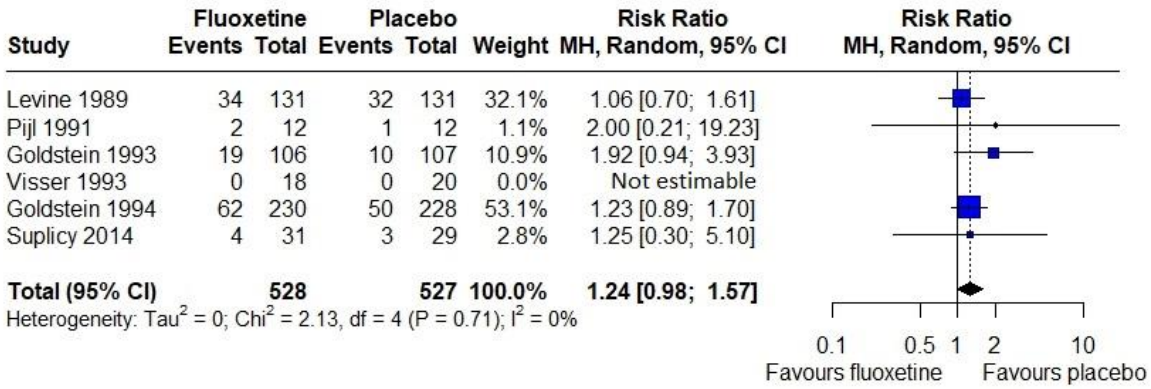


Figure 28 Analysis 1.5.13 Comparison 1: Fluoxetine versus placebo, Outcome 5: Specific adverse events, headache – without Levine 1987 and Wurtman 1993

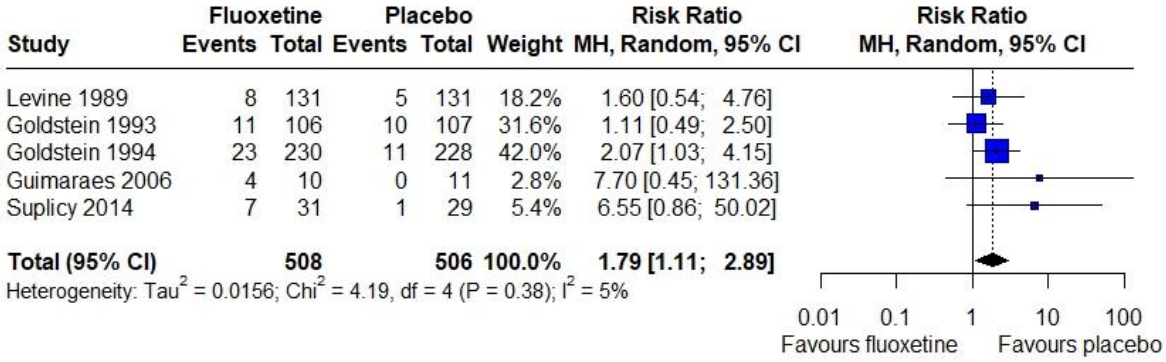


Figure 29 Analysis 1.5.14 Comparison 1: Fluoxetine versus placebo, Outcome 5: Specific adverse events, insomnia – without Levine 1987 and Wurtman 1993

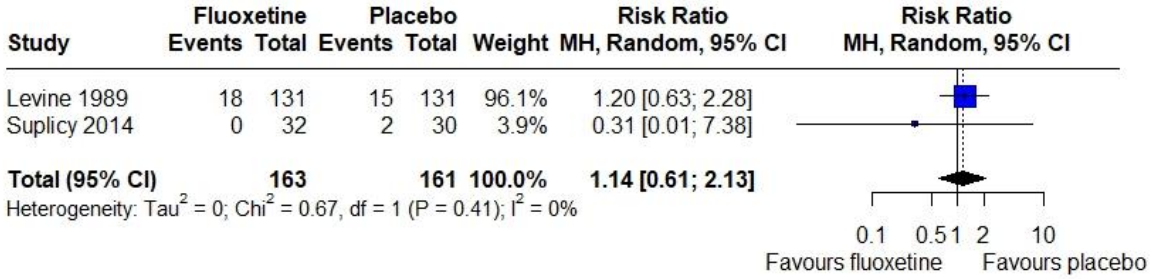


Figure 30 Analysis 1.5.15 Comparison 1: Fluoxetine versus placebo, Outcome 5: Specific adverse events, irritability – without Levine

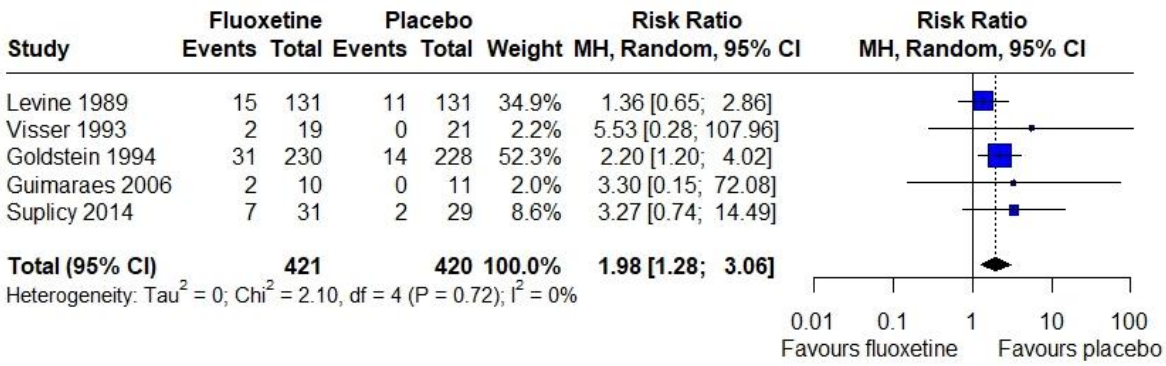


Figure 31 Analysis 1.5.17 Comparison 1: Fluoxetine versus placebo, Outcome 5: Specific adverse events, nausea – without Levine and Wurtman 1993

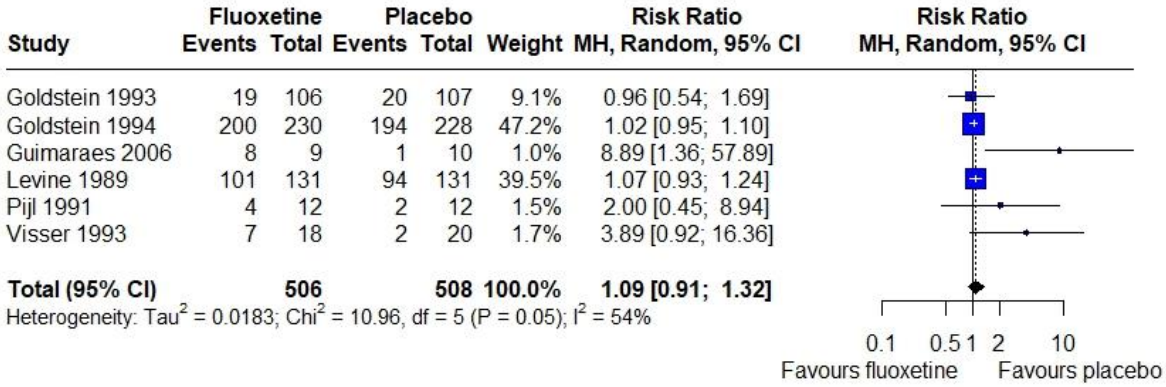


Figure 32 Analysis 1.6.1 Comparison 1: Fluoxetine versus placebo, Outcome 6: Participants with any adverse event (per dose), Fluoxetine 60 mg/d – without Levine 1987

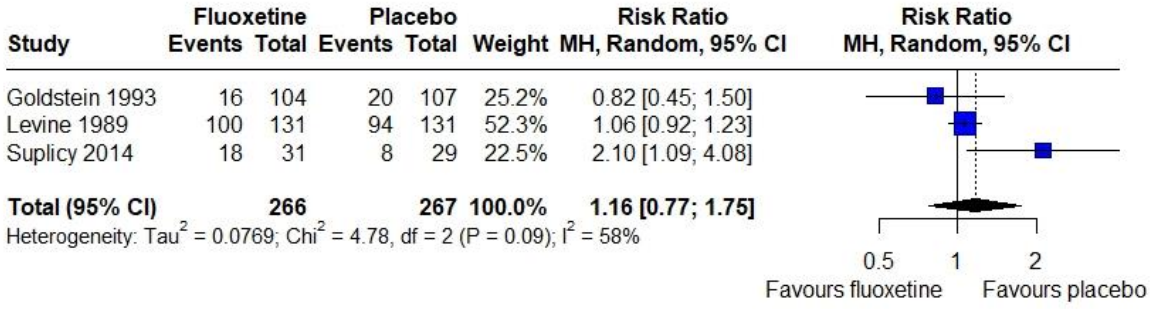


Figure 33 Analysis 1.6.3 Comparison 1: Fluoxetine versus placebo, Outcome 6: Participants with any adverse event (per dose), Fluoxetine 20 mg/d – without Wurtman 1993

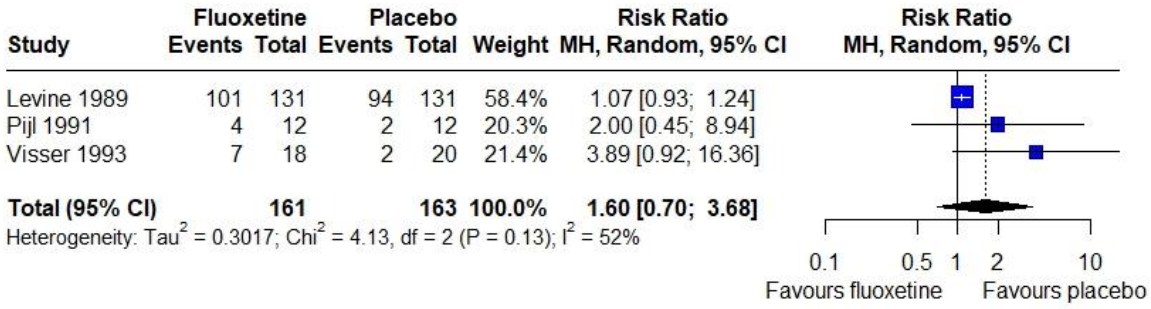


Figure 34 Analysis 1.7.1 Comparison 1: Fluoxetine versus placebo, Outcome 7: Participants with any adverse event (duration of intervention), 0 to 3 months – without Levine 1987

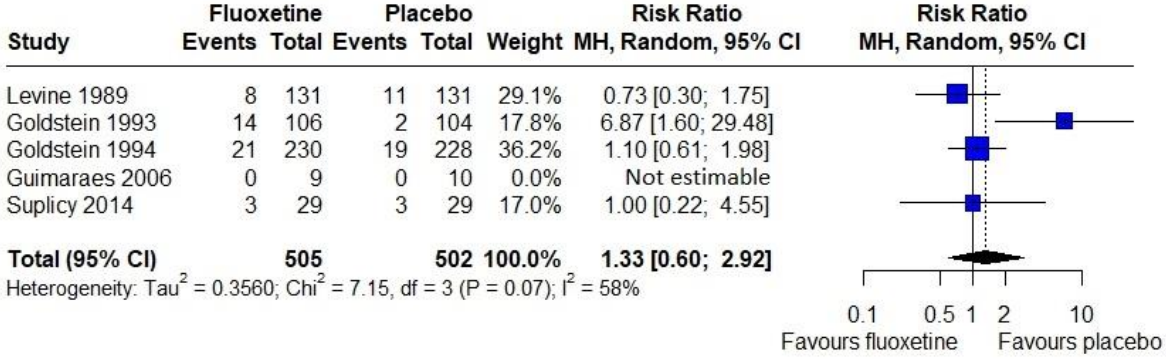


Figure 35 Analysis 1.8 Comparison 1: Fluoxetine versus placebo, Outcome 8: Participants discontinuing due to adverse events – without Levine 1987, Pedrinola 1993 and Wurtman 1993

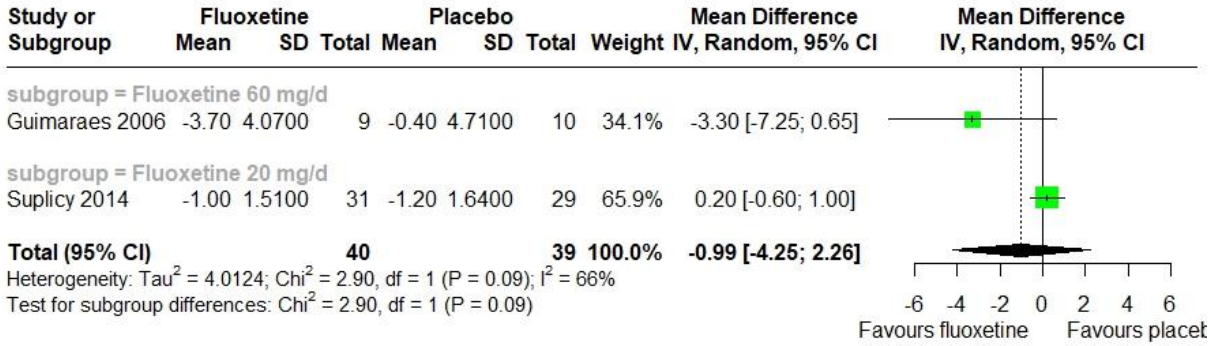


Figure 36 Analysis 1.9 Comparison 1: Fluoxetine versus placebo, Outcome 9: Anthropometric measurements other than weight loss (BMI) - without Pedrinola 1993

Figure 37 Analysis 1.10 Comparison 1: Fluoxetine versus placebo, Outcome 10: Morbidity (depression) - without Levine 1987

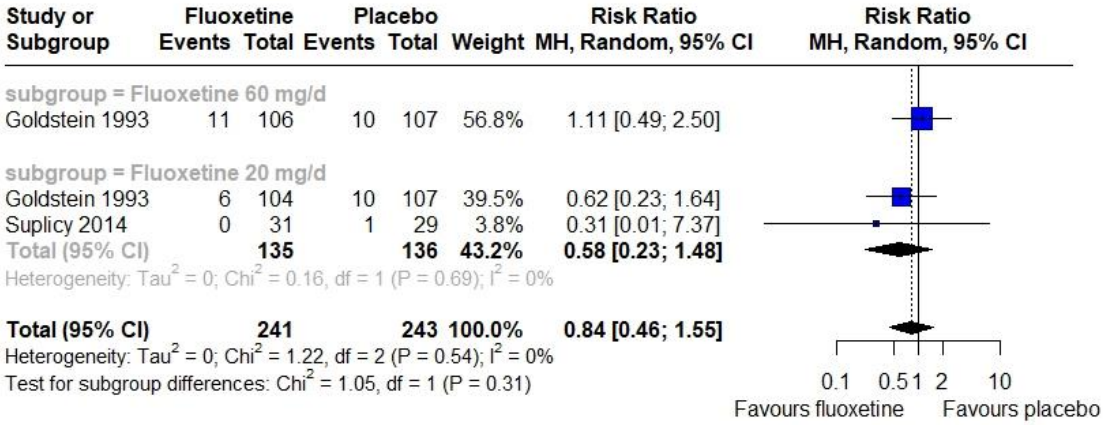


Figure 38 Analysis 1.11 Comparison 1: Fluoxetine versus placebo, Outcome 10: Morbidity (depression per dose) - without Levine 1987

8.2.3. RR_2: outcomes of meta-analyses

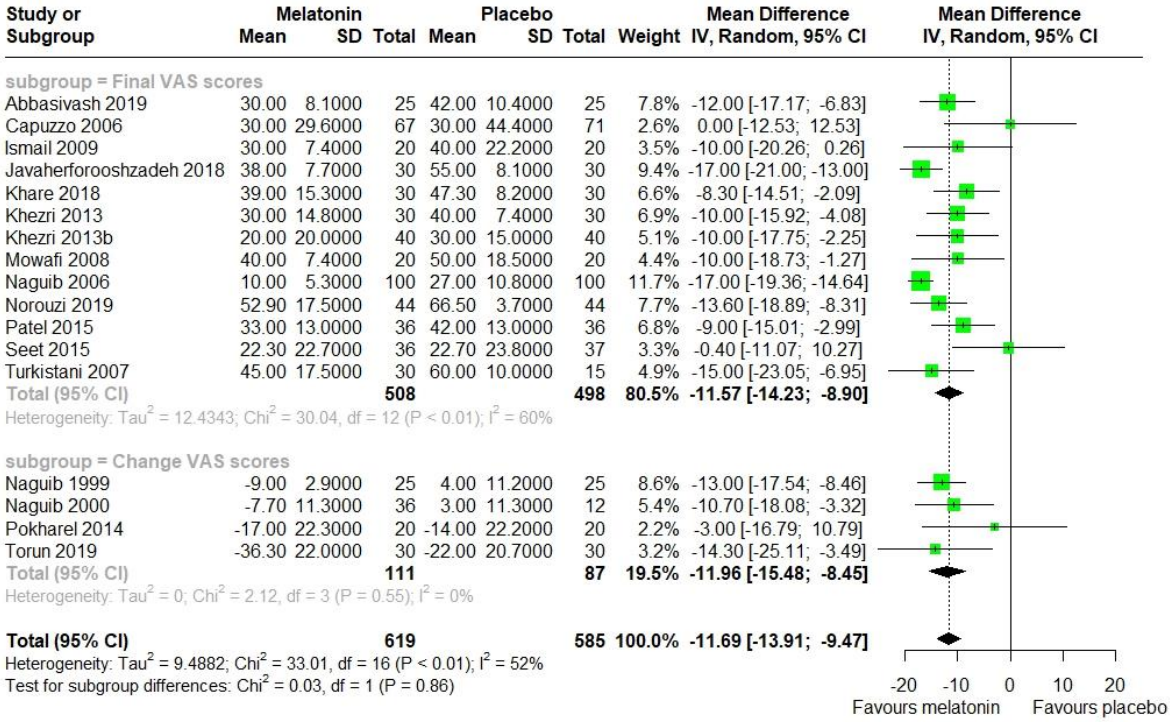


Figure 39 Analysis 1.1 Comparison 1: Melatonin versus placebo, Outcome 1.1 Preoperative anxiety (VAS) (mm) with subgroup 1.1.1 Final VAS scores and subgroup 1.1.2 Change VAS score - without Jain 2019

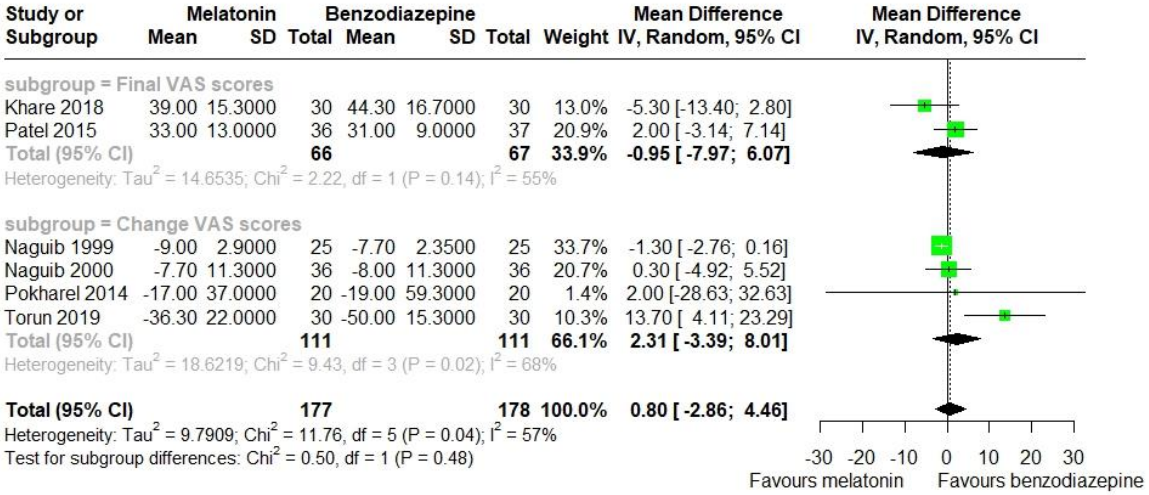


Figure 40 Analysis 2.1 Comparison 2: Melatonin versus benzodiazepine – preoperative anxiety, Outcome 2.1 Preoperative anxiety (VAS) (mm) - without Marzban 2016

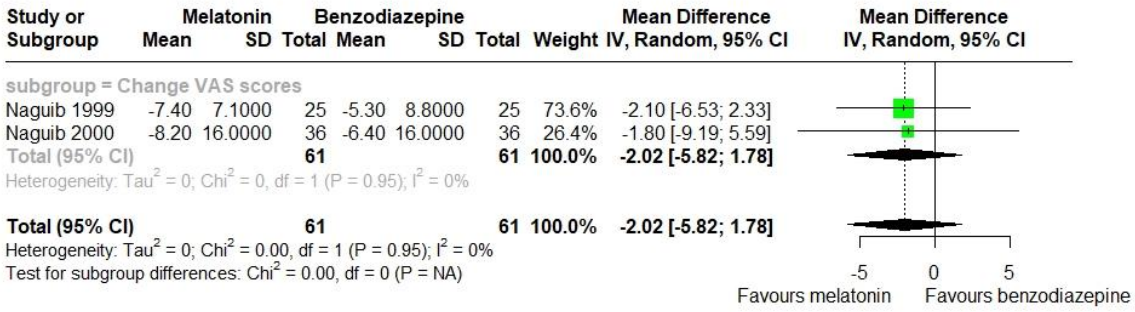


Figure 41 Analysis 2.2 Comparison 2: Melatonin versus placebo, Outcome 2 Postoperative anxiety (VAS) (mm) - without Marzban 2016

8.3. LRs in the era of new technologies

Table 19 Comparison of tools supporting all or several stages of LRs

Tool name	KWs highlighting	PDFs searching	AI support for screening	AI support for extraction	Comments	Price
CADIMA	The functionality is being developed	Software not integrated with publication databases	Text mining capabilities in development	-Text mining capabilities in development -The data extraction sheet will automatically be generated by CADIMA and the reviewer can mark those data that are needed -On- or offline extraction	-The software leads user through the process with instructions (excluding searching and quantitative synthesis) -Manually entering a search string in RIS format -Deduplication possible -Article screening for study selection possible - Title/abstract/full text co-displayed -RoB assessment possible Does not support qualitative/ quantitative syntheses of results -Gives output of text, figures or tables to assist with report writing -The extracted data are compiled in a form of a sheet -Automated calculation of a kappa-statistic to test inter-reviewer agreement when applying the defined criteria Where records are independently assessed by more than one reviewer and inconsistencies between reviewers occur, they will be automatically identified by CADIMA and the respective reviewers asked to solve those conflicts -Allows review steps to be modified and/or updated	Free

Tool name	KWs highlighting	PDFs searching	AI support for screening	AI support for extraction	Comments	Price
Colandr†	NR	Manual finding and importing of PDFs, once uploaded it can not be deleted	-Smart sorting and text mining (ML and NLP approaches) -After every 10 citations screened in the same review (including by multiple collaborators) the citation screener learns and the relevance ranking is updated	Data extraction from full texts powered by NLP (needs to be trained on 50 hand-coded metadata extractions in the same review (including by multiple collaborators) before it starts offering some suggestions for metadata extraction)	-Available stages: screening, data extraction, search -Deduplication available -PDF uploading available -Export of screening decisions and extracted data in csv format -Uses NLP and SL	Free
Covidence	KWs highlighting and a lightning quick interface available	Automatically uploads open access studies in FTR and transfer PDFs stored in reference manager	Integrated with Cochrane Randomized Controlled Trial classifier, which allows for filtering out the studies that are not RCTs	NR	-A tool for: screening, data extraction and critical appraisal -Works with Endnote or any tool that support RIS or PubMed formats -Keeps full records of who made decision, and supports single and dual screening -Adding notes and highlighting keywords of phrases in FTRs possible -Creating custom templates for data extraction possible -Exporting a single, machine-readable file that integrates into all the common statistics packages	One review with unlimited collaborators: $240/year Up to 3 reviews with unlimited collaborators: $635/year
EPPI Reviewer 4†	Presented in the in-depth analysis's summary, section 5.4.2					
Import.io	NA	NR	NA	NR	Facilitates search process and extracting data from websites	$199-$1099/month
IRIS.IA†	NR	-Open Access content from Core.ac.uk, PubMed, the US Patent Office,	NR	-Automatically extracts and systematizes any key data points from text and tables into a table layout of chosen	-Identifies the most meaning bearing words in text, enrich it with contextual synonyms and hypernyms -Summary tool does abstractive summarization -NLP and SL technologies involved	Payment required

Tool name	KWs highlighting	PDFs searching	AI support for screening	AI support for extraction	Comments	Price
		and CORDIS, all EU-funded research projects available -Each user can also simply choose to upload a list of documents; direct upload from computer, through personal cloud storage, or from reference manager possible		design -Allows filtering based on information extracted from documents like specific entities, data points or data ranges		
SUMARI	Presented in the in-depth analysis's summary, section 5.4.2					
MetaGear	NR	PDF downloader to automate the retrieval of journal articles from online databases	NR	-Automated data extraction from scatter-plots, box-plots -Simple imputation tools to fill gaps in incomplete or missing study parameters	-A package for the R statistical programming language -Supports: screening, data extraction and meta-analysis -Can assign screening across multiple collaborators and assess inter-reviewer reliability by kappa statistics -Can generate PRISMA flow chart	Free
Open-meta.app	NR	Uploading a citation file from a bibliographic database or doing a live search of PubMed possible	NR	NR	It's in beta phase and there might be occasional bugs -Search for duplicates available -PICO can be set up by reviewers - RoB assessment included -Synthesize section with: PRISMA diagram, sequential analysis, forest plot, bias blots and meta-regression	Free
Parsifal	NR	NR	NR	NR	-Help with the objectives, PICOS, research questions, search string, keywords and synonyms, selecting the	NR, signing up necessary

Tool name	KWs highlighting	PDFs searching	AI support for screening	AI support for extraction	Comments	Price
					sources, the inclusion and exclusion criteria -Provide mechanisms to build a quality assessment checklist and data extraction forms -Importing BibTex files and selecting the studies, finding duplicates among all the different sources, executing the quality assessment and extracting data from the papers possible	
Pvtopic†	NR	NA	Topic detection model to accelerate the performance of the active learning, text classification model used for citation screening	Method that encodes documents into a vector representation that is subsequently used to train the text classification model	-Neural embeddings based on a PubMed corpus -Identifies latent topics in a collection of documents and selects a set of meaningful and comprehensive words to describe the content of each topic -Supports search and data extraction process -Uses SL	Free
RAx†	NR	Uploading the literature only in the following formats .pdf, .fdf, .xfdf, .xfae	Smart recommendations: RAx scans the records uploaded and matches it with research papers indexed in their databases and recommends the most relevant and useful papers (recommended resources available in form of URLs	Automatically extracts articles' meta-data and references- user can click on the title of a reference and it will take her/him to the source website/Google Scholar	-Available stages: screening, data extraction, critical appraisal, reporting, search -Importing from a reference manager under development -Customizable review templates with question-answer format -Collaboration feature constantly scan the internet to bring the most current content (discover page) -Whenever entering a search query in Discover page, RAx enriches the query by identifying relevant concepts and resources that may not have the exact keywords in it but is related to what is searching for. This significantly increases the chances of discovering useful content that otherwise would be hidden. -Uses NLP and SL	From $10/ month to $25/ month depending on version
ReLiS	NR	NR	NR	NR	-Support for multi-user and multi-reviewer -Statistics on classified papers available -No more information identified	Free
SESRA	NA	NA	NA	NA	Software Engineering SLR Application, no further	NR

Tool name	KWs highlighting	PDFs searching	AI support for screening	AI support for extraction	Comments	Price
					information	
SRDB.PRO	NR	NR	Automatic removal of duplicates	NR	-Designed specifically for the pharmaceutical industry and healthcare consultancies -Supports all aspects of the SLR process -Assigns tasks automatically, confirms completion, and continually prompts project managers to instigate steps, or check on activities -Assigning users to the project or unassigning them possible -Set milestones which can be used to estimate the end-date of the project -Anonymous mode, which hides analyst names on the conflict reconcilement steps to prevent possible bias available -Provides an advanced feature for exporting extracted data into text and spreadsheet documents. All data can be exported from the SRDB.PRO database using filters, grouping, sorting and aggregation functions to provide a fully customisable output	£45,000/year/5 users
StArt	NR	NA	Automatic scoring calculated according to the number of times the keywords defined in the protocol appear	NR	-StArt provides support to the SLR process activities, except to the automated search of primary studies in electronic databases (BibTex file) -Duplicates identification available -Extraction form available, defined in the protocol	Free
SyRF†	NR	PDFs have to be uploaded with Endnote/csv files	Uses NLP	NR	-Supported stages: screening, data extraction, reporting -Dual screening with automatic reconciliation, manual reconciliation and single screening available -Not supporting deduplication -Creating annotation questions possible -Data export: bibliographic data, screening data and annotation data	Free

Tool name	KWs highlighting	PDFs searching	AI support for screening	AI support for extraction	Comments	Price
SysRev†	Presented in the in-depth analysis's summary, section 5.4.2					
Thoth	NR	NA	NA	NR	-Automated search for Scopus, overall search process not supported -Supported functions: study selection (handles disagreements between reviewers), quality assessment, data extraction, data synthesis, reporting, plot custom graphs, multiple users, role management, multiple projects	Free

AI – artificial intelligence, FDA – Food and Drug Administration, FTR – full-text review, KWs – keywords, NA-not applicable, NLP – natural language processing, NR – not reported, PICOS – population, intervention, comparators, outcomes, study design ; † ML tool

Table 20 Software and ML tools assisting search process (elaboration of search strategy, running the search and/or obtaining records retrieved by the search) with short description of functionalities, computational methods involved and price

Tool name	Description of functionalities	Computational methods involved	Price
2dsearch†	Instead of entering Boolean strings into one-dimensional search boxes, queries are formulated by manipulating objects on a two-dimensional canvas. This eliminates syntax errors, makes the query semantics more transparent, and offers new ways to collaborate, share, and optimize search strategies and best practices	NLP	Free
A2A	Enables the evaluation of query formulation techniques within a predefined framework	NA	Free
Aggregator†	Clusters together RCT articles predicted to arise from the same underlying clinical trial	SL	Free
Ananse	Partially automates search term selection and writing search strategies for SLRs	NA	Free
APSE	Recommends PubMed articles relevant to a given article. It automatically extracts keywords from the title and abstract indexed in PubMed to generate recommendations	NA	Free
BADERI	Provides an administration subsection, from which the roles of users are managed, a references subsection, where information associated to identified controlled clinical trials can be entered, a reports subsection, from which reports can be generated to track and analyse the results of handsearching activities, and a built-in free text search engine	NA	Free
BIBOT†	Interacts with MEDLINE database through the NCBI API, in order to retrieve abstracts, articles and meta-data from these articles (year of publication, author list, journal of publication, language, conflict of interest statement, etc.)	NLP	Free
BEST	No website found	No website found	No website found
Bioreader	Helps to distil reading list by ranking articles by relevance	NA	Free

Tool name	Description of functionalities	Computational methods involved	Price
Carrot2†	Automatically discovers groups of related documents and labels them with short key terms or phrases	NLP	Free
Chilibot†	Searches the PubMed literature database based on specific relationships between proteins, genes, or keywords; the results are returned as a graph	NLP	Free
citationchaser	Contains functions to automate process of citations chasing. An input article list can be used to return a list of all referenced records, and/or all citing records in the Lens.org database (consisting of PubMed, PubMed Central, CrossRef, Microsoft Academic Graph and CORE)	NA	Free
CitNetExplorer	Visualizes and analyses citation networks of scientific publications. The tool allows citation networks to be imported directly from the Web of Science database. Citation networks can be explored interactively, for instance by drilling down into a network and by identifying clusters of closely related publications	NA	Free
CloudSERA	No website found	No website found	No website found
Cochrane Register of Studies†	Web-based data repository and management system, developed for Cochrane. The specialised trials registers maintained by Cochrane information specialist are stored and managed within the Cochrane Register of Studies. It is a 'meta-register' of all the trials identified by Cochrane	NLP	Free with Cochrane Account
Coremine Medical†	A free Internet service for searching, updating, and sharing medical information – both search and social network	NLP	Free
Costumer†	Provided the data, the functions, scripts for the analyses and the documentation (report) within the relative templates	SL	Free
CTD	No website found	No website found	No website found
DOC search	No website found	No website found	No website found
Doctor Evidence†	Real-time medical search engine that generates actionable insights and answers critical business and research questions based on the published medical information	NLP	Payment required
Epistemonikos†	Combines all of the evidence relevant for health decision-making, including world's largest systematic review database, curated and annotated by the network of collaborators	SL	Free
FACTA+†	Helps discover associations between biomedical concepts from MEDLINE articles	NLP	Free
Findpapers	Helps researchers who are looking for references for their work by performing searches in several databases (currently ACM, arXiv, bioRxiv, IEEE, medRxiv, PubMed, and Scopus) from a user-defined search query	NA	Free
Gscraper	Harvests URLs at a speed of about 5 URLs per second per thread	NA	$68/lifetime
Heoro†	Pre-screens thousands of abstracts and indexes them by disease, type of intervention, study methodology and geographical setting	NLP	Free

Tool name	Description of functionalities	Computational methods involved	Price
Inciteful.xyz	Builds a network of papers from citations, uses network analysis algorithms to analyse the network. Allows entering two papers and it will gives an interactive visualization showing how the papers are connected by the literature	NA	Free
JANE	Compares a paper to documents in PubMed to find the best matching journals	NA	Free
JSTOR Text Analyzer	Analyses the text within the document to find key topics and terms used, and then uses the ones it deems most important — the "prioritized terms" — to find similar content in JSTOR (Journal Storage, a digital library)	NA	Free
LDS Shiny	No website found	No website found	No website found
Leaf	Searches for and retrieves records from a (possibly cloud-hosted) encrypted database while ensuring the confidentiality of the queries. Specifically, at the core of LEAF are three novel methods, referred to as Localization, Extraction, and Reconstruction	NA	NR
Leximancer†	Analyses any text to identify the high-level concepts, delivering the key ideas and actionable insights needed with powerful models, interactive visualisations and data exports	NLP USL	NR
Lingo3G	Organizes collections of text documents into clearly-labeled hierarchical folders	NA	NR
Linguamatics†	Transforms the free (unstructured) text in documents and databases into normalized, structured data suitable for analysis	NLP	Payment required
Litmaps	Discovers and traverses the citations and references of articles to identify highly relevant literature	NA	Free
Litesearchr†	Facilitates quick, objective, reproducible search strategy development using text-mining and keyword co-occurrence networks to identify important terms to include in a search strategy. It can automatically write Boolean search strings in up to 53 different languages, with stemming support for English. It can also check the results of a search against a set of known, relevant articles to get performance metrics	NLP	Free
Mapping MEDLINE	Mapping MEDLINE searches results against geographical MeSHes	NA	Free
Medline (PubMed) Trend	Displays the number of entries (articles) in PubMed (MEDLINE) published every year (having the 'DP', date of publication, field set to that year), that conform to search strategy	NA	Free
Medline Transpose	Translates a search query between Ovid MEDLINE and PubMed syntax	NA	Free
MeSHSIM	Measures semantic similarity over MeSH headings and Medline documents, supports querying hierarchy information of a MeSH heading and retrieving MeSH headings of a query document, and can be easily integrated into pipelines for any biomedical text analysis tasks	NA	Free
NAILS	No website found	No website found	No website found
OmixLitMiner	Helps researchers reduce time spent on literature research post analysis and streamline	NA	Free

Tool name	Description of functionalities	Computational methods involved	Price
	the decision about which proteins or genes are the most interesting and most promising for follow-up experiments		
paperscraper	Facilitates scraping publication meta-data as well as full PDF files from PubMed or from preprint servers such as arXiv, medRxiv, bioRxiv and chemRxiv. It provides a streamlined interface to scrape meta-data and comes with simple postprocessing functions and plotting routines for meta-analysis	NA	Free
PDQ-Evidence	Provides rapid access to systematic reviews of health systems evidence. It links together systematic reviews, broad syntheses of reviews and primary studies, thus providing a highly efficient method for searching. It includes translations of the titles and abstracts of included records to facilitate searching in different languages and it is continually updated by searching multiple sources of systematic reviews and broad syntheses of reviews	NA	Free
Pex	Identifies copyright use in custom databases, or within uploads to digital platforms via the Pex Registry	NA	NR
Publish or Perish	Retrieves and analyses academic citations. It uses a variety of data sources to obtain the raw citations, then analyses these and presents a range of citation metrics, including the number of papers, total citations and the h-index	NA	Free
PubReminer	Allows to directly enter, for example, a set of relevant PMIDs and determine high frequency words and subject headings assigned to those specific records, and can also be used to determine the most prolific authors on a topic, for example, authors who have the highest frequency of publications	NA	Free
PubVenn	Enables exploring PubMed using venn diagrams. By entering any multi-term search, one can see the relative size of the citation set for each term as well as how those sets interact	NA	Free
RCT Tagger†	Identifies human randomized controlled clinical trial articles (up to 25,000 most recent articles) retrieved by a PubMed query	NLP	Free
Researchr	Platform for finding, collecting, sharing, and reviewing scientific publications	NA	NR
RoBotAnalyst†	Designed for searching and screening reference collections obtained from literature database queries. It combines search engine functionality with ML and text mining technology, including topic modelling and relevancy feedback-based text classification models, to minimise the human workload involved in identifying relevant references	SL	Contacting authors required
Sample Size Search (SSS) Tool for PubMed†	Automatically extract information on sample size from clinical trial reports by processing their abstracts	NLP	Free

Tool name	Description of functionalities	Computational methods involved	Price
Sci2 Tool	A modular toolset that supports the temporal, geospatial, topical, and network analysis and visualization of scholarly datasets at the micro (individual), meso (local), and macro (global) levels	NA	Free
searchrefiner	An interactive interface for visualising and understanding queries used to retrieve medical literature for systematic reviews	NA	Free
SensPrecOptimizer	Software for developing Highly Sensitive Search Strategies	NA	Free
SLRqub	No website found	No website found	No website found
SLR_SearchStrings†	NR	NLP	Free
SRA†	Supports: search, deduplication and screening (includes a functionality of identifying disagreements and resolve them). Allows exporting databases from Endnote or Zotero and Mendeley. Word Frequency Analyser - outputs the most common words/phrases that appear within library	NLP	NR
Swift-Review†	Supports: searching, categorizing, and prioritizing of literature in an interactive manner. Utilizes newly developed statistical text mining and ML methods that allow users to uncover over-represented topics within the literature corpus and to rank order documents for manual screening	SL	Free
Textpresso	An information extracting and processing (text mining) package for biological literature. The two key elements are the collection of the full text of scientific articles split into individual sentences, and the implementation of semantic categories, for which a database of articles and individual sentences can be searched	NA	NR
TextRank†	The tool assists: phrase extraction (gets the top-ranked phrases from a text document), summarization of a text document, helps infer concepts from unstructured text into more structured representation	NLP USL	Free
Thalia†	Recognizes eight different types of concepts (chemicals, diseases, drugs, genes, metabolites, proteins, species, anatomical entities) occurring in biomedical abstracts. Available via a web-based interface or a RESTful API. It is updated from PubMed on a daily basis	NLP	Free
Trial2rev†	Requires programming knowledge. Currently known as ES3. Web-based interface to a database that collates and augments information from multiple sources including bibliographic databases, the ClinicalTrials.gov registry, and the actions of registered users	SSL	Free
TrialStreamer†	Finds and summarises new trial publications, registrations, and preprints in COVID-19	NLP SL	Free

Tool name	Description of functionalities	Computational methods involved	Price
VOSviewer	A software tool for constructing and visualizing bibliometric networks. These networks may for instance include journals, researchers, or individual publications, and they can be constructed based on citation, bibliographic coupling, co-citation, or co-authorship relations. VOSviewer also offers text mining functionality that can be used to construct and visualize co-occurrence networks of important terms extracted from a body of scientific literature	NA	Free
Voyant Tools†	Allows analysing text or collection of text to educate others and present results in a way that is more easily observable to an online audience	NLP	Free
WordStat 9.0†	Text analysis software to quickly extract and analyse information from large amounts of documents. WordStat's seamless integration with SimStat – statistical data analysis tool – QDA Miner – qualitative data analysis software – and Stata – the comprehensive statistical software from StataCorp, gives flexibility for analysing text and relating its content to structured information, including numerical and categorical data	NLP	NR
Yale MeSH Analyzer	Can help identify the problems in search strategy by presenting the ways articles are indexed in the MEDLINE database in an easy-to-scan tabular format	NA	Free

MeSH – medical subject heading, NA – not applicable, NLP – natural language processing, NR – not reported, SL – supervised learning, SSL – semi-supervised language, USL – unsupervised learning; † ML tool

Table 21 Software and ML tools assisting screening process with short description of functionalities, computational methods involved and price

Tool name	Short description of functionalities	Computational methods involved	Price
AbstrackR†	Allows to semi-automate title and abstract screening (recognizes patterns in relevant and irrelevant records, as labelled by the user and presents records in order of relevance based on a predictive model)	SL	Free
Active_learning_document_screening†	Contains the complementary material used for automatic document screening of medical literature using word and text embeddings in an active learning setting	NLP SL	Free
Active-learning-for-systematic-review†	Active learning for workload reduction of automated screening in SLRs	NLP SL	Free
ArticleNet	No website found	No website found	No website found

Tool name	Short description of functionalities	Computational methods involved	Price
ASReview†	Can be used for: screening new text datasets, teaching using existing data sets, science to investigate the performance of different models and datasets. Compatible with EndNote, Rayyan, Distiller, PubMed	NLP SL	Free
BioReader	No website found	No website found	No website found
Cochrane RCT Classifier†	Helps to distinguish between reports of RCTs and non-RCTs	SL	Free
Concept Encoder†	Expresses similarities and relationships between documents and words by vectorization, and in order to capture the characteristics of documents in a multifaceted manner	NLP SL	NR
DAE-FF†	A neural network-based feature extraction method to facilitate citation screening for systematic reviews	NR	Free
DoCTER†	Classifies a set of documents as relevant or non-relevant to a topic. Assesses the probability of a document being relevant to the topic of interest Training dataset needed (a set of documents annotated as being relevant or not to the topic of interest) to train the ML algorithm. Clusters a set of documents into a user-specified number of bins. For each bin, identifies the defining topics/keywords.	NLP SSL	NR
Doctor†	No website found	No website found	No website found
FASTREAD†	Supports primary study selection in SLR	SL	NR
GAPscreener†	Assists in screening PubMed abstracts for human genetic association studies	NLP SL	Free
Inclusion criteria†	A code for machine classification of inclusion criteria from Cochrane systematic reviews	NLP	Free
Machine Learning Functions†	Excludes abstracts according to the specific reasons defined according to the PICOS framework	SL	Free
Papers-review	No website found	No website found	No website found
PICO Portal†	Offers fast and accurate deduplication technology, to review the most relevant articles first. Facilitates collaboration, assigns articles automatically (single, dual or custom review)	NLP SSL	Price not public
PubmedClassifier†	Reviews and classifies the medical literature on cancer susceptibility genes	NLP	Free
Pvtopic†	Identifies latent topics in a collection of documents and selects a set of meaningful and comprehensive words to describe the content of each topic	SL	Free
RapidMiner†	Enables open communication, easy sharing, and widespread re-use of work without breaking policies and regulations	NLP SL SSL	Available at request

Tool name	Short description of functionalities	Computational methods involved	Price
		USL	
Rayyan†	No detailed information	NLP SL	$4-8.25 per month
Reasearch Screener†	Actively learns from a reviewer's decisions and present the most relevant articles for the study during screening	NLP SL	Not available to purchase yet
RevTools†	Facilitates screening and deduplication, including dividing screening tasks among a team and manual article screening (screening of titles and abstracts)	NLP	Free
RoBotAnalyst†	Supports screening references obtained from database queries. Facitlitates uploading references in RIS format or from PubMed, browsing automatically generated clusters labelled by keywords, finding the most similar references, adding notes to references and filter the results by the notes, prioritising references by relevancy predictions and automatically finalise screening by predictive model suggestions without manual screening	SL	NR
RoBotSearch†	Performs full-text indexed searches on the contents	NLP SL	$19.99-199.99/1 license
Rules_cochranereviews †	Contains an updated version of the implementation of rules for automatically extracting relevant information from Cochrane reviews	NLP	NR
Screen4Me†	Assists through the screening of search results much more by identifying already known RCTs, removing very obvious non-RCT records from the set via ML, and sending the remaining records to be assessed by a crowd via Cochrane Crowd	NLP	Available for Cochrane intervention reviews only
Screen-IT	No website found	No website found	No website found
SLR_SearchStrings†	NR	NLP	Free
SRA (Systematic Review Accelerator)†	Allows exporting databases from Endnote or Zotero and Mendeley. Outputs the most common words/phrases that appear within library. Identifies disagreements and resolves them	NLP	NR
Subscreen	Facilitates screening and identifying subgrouping factors	NR	Free
SWIFT-Active Screener†	Automatically prioritizes articles as they are reviewed, using user feedback to push the most relevant articles to the top of the list. A separate statistical model estimates the number of relevant articles remaining in the unscreened document list. Supports real-time collaboration and conflict resolution	SL	NR, payment required

Tool name	Short description of functionalities	Computational methods involved	Price
SWIFT-Review†	Allows users to uncover over-represented topics within the literature corpus and to rank order documents for manual screening	SL	Free
Sys_review_ml†	No specific information provided	SL	Free

NA – not applicable, NLP – natural language processing, NR – not reported, RCT – randomized controlled trial, SL – supervised learning, SSL – semi-supervised learning, USL – unsupervised learning
† ML tool

Table 22 Software and ML tools assisting data extraction process with description of functionalities, computational methods involved and price

Tool name	Description of functionalities	Computational methods involved	Price
A Keyword-Based Literature Review Data Generating Alhorithm†	Automatically extracts keywords from the meta-information of each article and generate the basic data for review articles	NLP	Free
Cochrane_scraper	Searches through a range of potential Cochrane IDs and download data for all found meta-analyses	NA	Free, only with institutional access to Cochrane Library
CERC	Aims at developing new tools and methodologies that will allow enormous volumes of data from multiple sources to be processed and analysed in real-time — in order to obtain usable knowledge and to automate decision-making	NA	NR
ContentMine	Facilitates extracting "facts" with a number of "facet" tools: word search, regexes, bespoke text tools, chemical NLP (OSCAR), and certain diagram types (phylogenetic trees)	NA	Free
EvidenceSET	Facilitates extracting data from the primary studies, such as bibliographical data, aims, context, and results; coding data by identifying and coding such as interesting concepts, categories, findings, and results. Allows translating codes into themes, subthemes, and higher order themes	NA	Free
ExaCT†	Extracts accurate efficacy and safety information from CT.gov in TSV spreadsheet format via series of steps (based on NCT ID)	SL	NR, payment required
Graph2Data	Converts 2D cartesian graph to csv file	NA	Free
import.io	Facilitates extracting data from websites	NA	$199-1099/month
MetaDigitise	Extracts data from figures in primary research papers. Automatically calculates summary statistics for users from box plots, bar plots (e.g., mean and errors), scatter plots and histograms	NA	Free

Tool name	Description of functionalities	Computational methods involved	Price
Numbat	Facilitates managing the extraction of large volumes of data from primary sources among multiple users, and then reconciling the differences between them. Provides different levels of extraction (e.g. title-and-abstract vs full extraction). Can generate reference networks among the publications in the database	NA	Free
PICOtron	Extracts as much data as possible directly from Cochrane Reviews	NA	NR
PlotDigitizer	Allows users to extract data from images in numerical format	NA	$49/1 lifetime license
PubMed2XL	Downloads PubMed data into an Excel spreadsheet based on PMID	NA	Free
Pvtopic†	Identifies latent topics in a collection of documents and selects a set of meaningful and comprehensive words to describe the content of each topic	SL	Free
Raptor†	Helps visualize algorithms and avoid syntactic baggage	NLP	Free
Refchaser	No detailed information on extractions support	NA	Free
Rnatlp†	Set of scripts to extract meaningful information when performing an SLR	NLP	Free
RoBotReviewer†	Allows users to upload RCT articles and see automatically determined information concerning the trial conduct (the 'PICO', study design, and whether there is a RoB)	NLP SL	Free
rules_cochranreviews†	The repository contains an updated version of the implementation of rules for automatically extracting relevant information from Cochrane reviews	NLP	Payment required
Scholarcy	Reads research articles, reports and book chapters and breaks them down into bite-sized sections. Creates a summary flashcard of any article, report or document in Word or PDF format. It creates links to open-access versions of cited sources and can be configured to extract figures, tables and images	NA	Personal library for $7.99 Academic institution license: $8,000+ (unlimited access, annual subscription)
TableBuilder	Builds tables based on underlying microdata; displays counts, percentages and relative standard errors, calculates means and medians for continuous variables such as income	NA	Free
Topictagger	Helps organize content, to improve both, the search and browsing experience; applies tags	NA	Free
WebPlotDigitizer	Works with a wide variety of charts (XY, bar, polar, ternary, maps etc.), automatically extracting a large number of data points	NA	Free
WordStat	Facilitates fast extraction of themes and trends, or careful and precise measurement with state-of-the-art quantitative content analysis tools	NA	One-year commercial license: $2,900

Tool name	Description of functionalities	Computational methods involved	Price
			One-year academic research license: $375

CSV – comma-separated values, NA – not applicable, NLP – natural language processing, NR – not reported, RoB – risk of bias, SL – supervised learning, TSV – tab-separated values; † ML tool

Table 23 Software and ML tools assisting critical appraisal with short description of functionalities, computational methods involved and price

Tool name	Description of functionalities	Computational methods involved	Price
AMSTAR Checklist	The website is not functioning	The website is not functioning	The website is not functioning
CrowdCARE	Crowdsourcing critical appraisal for health-care practice	NA	NR
GRADEpro	Guides through the process of guideline development while seamlessly making sure it adheres to the GRADE methodology. Facilitates project management tasks and collaborate; summarizing and grading evidence with several available table formats and make judgments using the "Evidence to Decision" framework	NA	Free for groups up to 3 researchers $2200/active project/year (15 team members)
ORBIT Matrix Generator	Provides a transparent way for systematic reviewers to show the outcomes that are reported for each study in the review, and those that are missing or partially reported; allows reviewers to assess the risk of outcome reporting bias according to the ORBIT classification system	NA	Free
RoBotReviewer†	Allows users to upload RCT articles and see automatically determined information concerning the RoB	NLP SL	Free
RoBvis	Facilitates visualising RoB assessments performed as part of a systematic review. Creates: "traffic light" plots of the domain-level judgements for each individual result and weighted bar plots of the distribution of RoB judgements within each bias domain	NA	Free
Systematic Review Assistant	The sentences are highlighted by relevance with respect to each RoB property. A stronger highlighting colour (higher opacity) indicates higher relevance	NA	Free

NA – not applicable, NLP – natural language processing, NR – not reported, RoB – risk of bias, SL – supervised learning; † ML tool

Table 24 Software tools assisting meta-analysis process with short description of functionalities and price

Tool name	**Description of functionalities**	**Price**
Comprehensive Meta-Analysis	Facilitates entering data directly or import data from other program (any spreadsheet-based program), computing the treatment effect (or effect size) automatically for binary and continuous data, mixing and matching different data formats (sample size, odds ratio etc.). Allows running the core meta-analysis and creates a display that serves as a roadmap for all that follows.	$95-895/1 copy/year
Dmetar	Facilitsts performing meta-analyses in R from scratch with focus on biomedical and psychological research synthesis. Covers topics such as power analysis, effect size calculation, small-study effects, publication bias, meta-regression, subgroup analysis, RoB assessment, and network meta-analyses	Free
MAVIS	Provides support for both random effects and fixed effects models and additional support for calculating effect sizes plus single case designs, graphical output, and detecting publication bias	Free
MetaDTA	Facilitates meta-analysis of diagnostic test accuracy studies using a bivariate model, and displays the individual study and meta-analysis results using a summary receiver operating characteristic curve	Free
Meta-CART	Integrates classification and regression trees into meta-analysis. Provides a flexible approach to identify interaction effects between moderators in meta-analysis	Free
Meta-Essentials	Facilitates the integration and synthesis of effect sizes from different studies. Consists of a set of workbooks designed for Microsoft Excel that, based on an input, automatically produces all the required statistics, tables, figures, etc.	Free
metafor	Facilitates calculating various effect sizes or outcome measures and then allows the user to fit equal-, fixed-, and random-effects models to these data. For meta-analyses of 2×2 tables, proportions, incidence rates, and incidence rate ratios, the package also provides functions that implement specialized methods, including the Mantel-Haenszel method, Peto's method, and a variety of suitable generalized linear mixed-effects models (i.e., mixed-effects logistic and Poisson regression models). For non-independent effects/outcomes (e.g., due to correlated sampling errors, correlated true effects or outcomes, or other forms of clustering), the package also provides a function for fitting multilevel and multivariate models to meta-analytic data	Free
MetaGenyo	Facilitates a meta-analysis of genetic association studies, including Hardy-Weinberg table, association values, forest plots, publication bias, subgroup analysis and sensitivity analysis	Free
MetaInsight	Provides a structured representation of knowledge extracted from multi-dimensional data aiming to facilitate exploratory data analysis automatically and effectively. A novel formulation of basic data patterns allows to capture essential characteristics of raw data distribution to achieve knowledge extraction	Free
MetaLight	Supports the teaching and learning of meta-analysis	Free

Tool name	Description of functionalities	Price
metamisc	Facilitate frequentist and Bayesian meta-analysis of diagnosis and prognosis research studies. Includes functions to summarize multiple estimates of prediction model discrimination and calibration performance, to evaluate funnel plot asymmetry and for developing multivariable prediction models from datasets with clustering	Free
MetaXL	Facilitates the quality effects model, the inverse variance heterogeneity model, the Doi plot and LFK index for the detection of publication bias, network meta-analysis, and cumulative meta-analysis	Free
MIX 2.0	Facilitates professional meta-analyses from inside Excel	$185-265 (general, 2 computers)
NetMetaXL	Provides an Excel-based interface for conducting a Bayesian network meta-analysis using WinBUGS from within Microsoft Excel; allows the user to easily prepare and enter data, set model assumptions, and run the network meta-analysis, with results being automatically displayed in an Excel spreadsheet. Generates evidence network diagrams, forest plots, league tables of pairwise comparisons, probability plots (rankograms), and inconsistency plots	Free
OpenMEE	Provides the following functionalities for ecological and evolutionary meta-analyses: standard, cumulative, leave-one-out, subgroup, phylogenetic meta-analyses, meta-regression, bootstrapped meta-analysis, multiple imputations for missing data, data exploration, publication bias	Free
Open Meta-Analyst	Facilitates meta-analyses of binary, continuous, or diagnostic data, using a variety of fixed and random-effects methods, including Bayesian and maximum likelihood analysis, cumulative, leave-one-out, subgroup, and meta-regression analyses	Free
Pymeta/PythonMeta	Facilitates evidence-based medicine tasks, such as: combining effect measures (odds ratio, risk ratio, risk difference for count data and weighted mean difference, standard mean difference for continuous data); heterogeneity test(Q/Chi-square test); subgroup analysis; cumulative meta-analysis; sensitivity analysis (one or two factors); plots drawing: forest plot, funnel plot, and bar-line, cross-block plot	Free
RevMan	Performs meta-analysis of the data entered, and present results graphically	Pricing for organizations: on request For individuals: £60-99 + VAT (EU&UK)
X-META	Facilitates the following functions: fixed effect and random effect meta-analysis, plots (forest, funnel, radial, etc.), statistical tests to evaluate bias, meta-regression	NR

NR – not reported

Table 25 Software tools assisting reporting with short description of functionalities and price

Tool name	**Description of functionalities**	**Price**
EviAtlas	Provides with overviews of the quantity and quality of evidence in a form of systematic maps. Allows categorizing documents according to the type, location, and publication information available for each work within a particular topic	Free
EvidenceSET	Facilitates extracting data from the primary studies (e.g., bibliographical data, aims, context, and results), and coding data by identifying interesting concepts, categories, findings, and results. Then, translates codes into themes, subthemes, and higher-order themes	Free
Gephi	Visualization and exploration software for all kinds of graphs and networks. Can be used in exploratory data analysis, link analysis, social network analysis, biological network analysis and poster creation	Free
GRADEpro	Offers team management, questions and outcomes generation, presentation formats, multiple comparisons, references management, evidence tables, evidence to decision framework	Up to $2200/active project/year
RevMan	Facilitates preparation of protocols and full reviews, including text, characteristics of studies, comparison tables, and study data. It can perform a meta-analysis of the data entered, and present results graphically	Pricing for organizations: on request For individuals: £60-99 + VAT (EU&UK)

Table 26 Software tools assisting reference management with short description of functionalities and price

Tool name	**Description of functionalities**	**Prices**
BibDesk	Facilitates managing bibliographies and references when writing essays and articles	Free
Citavi	Assists in organizing, tracking, and formatting references. Includes automatic citation formatting for over 10,000 styles	€255-425, depending on version
colwiz	Incorporates reference management, collaboration and networking tools, as well as productivity features	Free
docear	Assists organizing, creating, and discovering academic literature	Free
JabRef	Assists collecting citations, retrieving full texts, and links full texts. Facilitates collecting complete bibliographic info, editing citations, organizing and citing	Free

Tool name	Description of functionalities	Prices
KBibTeX	A reference management software primarily for BibTeX which is typically used in conjunction with TeX/LaTeX. Beyond normal editing capabilities, KBibTeX offers features such as searching and importing new references from Google Scholar or BibSonomy	Free
Mendeley	Facilitates adding papers directly from browser or importing any documents from your desktop. Generates references, citations and bibliographies in a whole range of journal styles	Free
Paperpile	Facilitates managing research library from browser	$10/month
Papersapp	Allows searching across search engines and 1-click downloads to add references and full-text PDFs to the library. Full-screen enhanced PDF reader available (highlighting, underlining, striking through, making inline notes, drawing and adding sticky notes possible). Users can create private shared collections to collaborate with up to 30 other papers users at a given time	$10/month
Qiqqa	Combines PDF reference management tool, a citation manager, and a mind map brainstorming tool.	Free
refbase	Multi-user interface for managing scientific literature and citations	Free
Wikindx	Provides bibliographic and quotations/notes management and article authoring system (Virtual Research Environment) designed either for single use or multi-user collaborative use	Free
Zotero	Facilitates collecting, organizing, annotating, citing, and sharing research	Free

Table 27 Software tools assisting data visualization with short description of functionalities and price

Tool name	Description of functionalities	Price
ALBATROSS	Facilitates presenting results of diversely reported studies in a systematic review. The plot allows an approximate examination of underlying effect sizes and the potential to identify sources of heterogeneity across studies	Free
Gephi	Visualization and exploration software for all kinds of graphs and networks. Can be used in exploratory data analysis, link analysis, social network analysis, biological network analysis and poster creation	Free
PRISMA 2020	Facilitates making PRISMA2020 flow diagrams	Free
PrismadiagramR	Creates PRISMA diagram from a minimal dataset of included and excluded studies and allows for more custom diagrams	Free
PRISMAstatement	Connected to PRISMA 2020	Free
ROSES flowchart	Provides with templates for recording the flow of articles through searching, screening, coding/meta-data extraction, data extraction, critical appraisal and synthesis for systematic reviews and systematic maps	Free
yEd Graph Editor	Generate high-quality diagrams (manually or by importing external data for analysis). Arranges even large data sets	Free

Table 28 An overview of SLR databases, which register, collect and disseminate SLRs

Database name	**Description**
BEME	An international group of individuals, universities and professional organisations committed to the development of evidence informed education in the medical and health professions through the dissemination of information which allows teachers and stakeholders in the medical and health professions to make decisions on the basis of the best evidence available; the production of reviews which present the best available evidence and meet the needs of the user; and the creation of a culture of best evidence education amongst individuals, institutions and national bodies
Campbell Collaboration	An international social science research network that produces high quality, open and policy-relevant evidence syntheses, plain language summaries and policy briefs
Cochrane Library	A collection of databases (including Cochrane Database of Systematic Reviews, CENTRAL and Cochrane Clinical Answers) that contain high-quality, independent evidence to inform healthcare decision-making
Epistemonikos	Combines the best of evidence-based health care, information technologies and a network of experts to provide a unique tool for people making decisions concerning clinical or health-policy questions
JBI EBP	Promotes and supports evidence-based decisions that improve health and health service delivery, and the sustainability of improved healthcare practice and health outcomes globally by developing and delivering a range of unique evidence-based resources
MAGICapp	A web-based tool that helps users and organizations to author, publish and update digitally structured clinical practice guidelines based on best current evidence, enabling clinicians and patients to make well-informed healthcare decision
PEDro	A web-based database of randomized controlled trials and SLRs in physiotherapy. The database contains bibliographic details and abstracts of most English-language randomized trials and systematic reviews in physiotherapy, and of many trials and reviews in other languages. Trials on the database are rated on the basis of their methodological quality so that users of the database can quickly identify trials of high quality. Trials and SLRs are extensively indexed to facilitate searching. The database provides an important information resource to support evidence-based clinical practice

9. List of abbreviations

9.1. Abbreviations used in the text and tables

Abbreviation	Definition
AI	Artificial intelligence
AMSTAR	Tool for the assessment of multiple systematic reviews
API	Application programming interfaces
ARDS	Acute respiratory distress syndrome
BMI	Body Mass Index
BMJ	British Medical Journal
CADTH	Canada's Drug and Health Technology Agency
CENTRAL	Cochrane Central Register of Controlled Trials
ChatGPT	Chat Generative Pre-Trained Transformer
CI	Confidence interval
CINAHL	Cumulative Index to Nursing and Allied Health Literature
CONSORT	Consolidated Standards of Reporting Trials
CRD	Centre for Reviews and Dissemination
CT.gov	Clinicaltrials.gov
C2H	Core2 Health
Embase	Excerpta Medica Database
Emtree	Embase subject heading
ERGs	Evidence Review Groups
EUnetHTA	European Network for Health Technology Assessment
FDA	U.S. Food and Drug Administration
FTR	Full texts review
GPT-3	Generative Pre-trained Transformer 3, a large language model
GPT-4	Generative Pre-trained Transformer 4, a multimodal large language model
GRADE	The Grades of Recommendation, Assessment, Development and Evaluation for Systematic Reviews
HAS	Haute Autorité de Santé
HTA	Health Technology Assessment
ICD-10	International Classification of Diseases,Tenth Revision
ICER	Institute for Clinical and Economic Review
ICTRP	International Clinical Trials Registry Platform
IHME	Institute for Health Metrics and Evaluation
IQWIG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen in Germany
ISSG	InterTASC Information Specialists' Sub-Group
JBI	Joanna Briggs Institute
KWs	Keywords
LILACS	Latin American and Caribbean Health Science Information database
LLM	Large language model
LR	Literature review
MEDLINE	Medical Literature Analysis and Retrieval System Online
MeSH	Medical Subject Headings; it is a controlled and hierarchically-organized vocabulary produced by the National Library of Medicine. It is used for indexing, cataloging, and searching of biomedical and health-related information
ML	Machine learning
MS	Microsoft
NICE	The National Institute for Health and Care Excellence in the UK
NLP	Natural language processing
NA	Not applicable
NR	Not reported
PBAC	Pharmaceutical Benefits Advisory Committee in Australia
PICOS	Population, intervention, comparators, and study design
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-P	Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols
QSR	Quick scoping reviews
QUORUM	Quality of Reporting of Meta-Analyses checklist
PRESS	Peer Review of Electronic Search Strategies

PROSPERO	International prospective register of systematic reviews
RCT	Randomized Controlled Trial
REA	Rapid evidence assessments
RoB	Risk of bias
ROBIS	Tool to assess risk of bias in systematic reviews
ROBINS-I	The Risk Of Bias In Non-randomized Studies of Interventions tool
RR	Rapid review
RWD	Real-world data
RWE	Real-world evidence
SIGN	Scottish Intercollegiate Guidelines Network
SL	Supervised learning
SLR	Systematic literature review
SSL	Semi-supervised learning
UK	The United Kingdom
US	The United States
USL	Unsupervised learning
VAS	Visual analogue scale
WHO	World Health Organization

9.2. Abbreviations used only in detailed HTA recommendations

Abbreviation	Definition
BIA	Budget impact analysis
CBA	Cost-benefit analysis
CCA	Cost-consequence analysis
CEA	Cost-effectiveness analysis
CEAC	Cost-effectiveness acceptability curves
CEESP	The Economic and Public Health Committee
CMA	Cost minimization analysis
CUA	Cost-utility analysis
DSA	Deterministic sensitivity analysis
EQ-5D	A standardised measure of health-related quality of life developed by the EuroQol Group to provide a simple, generic questionnaire for use in clinical and economic appraisal or population health status surveys
HEE	Health economic evaluation
HRQoL	Health-related quality of life
HUI	Health Utilities Index
ICER	Incremental Cost-Effectiveness Ratio or Institute for Clinical and Economic Review, depending on the context
MAUI	Multiattribute utility instrument
NB	Net benefit
NHB	Net health benefit
NHS	The National Health Service
NMB	Net monetary benefit
OECD	Organization for Economic Co-operation and Development
PSA	Probabilistic sensitivity analysis
PSS	Personal Social Services
QALY	Quality-adjusted life year
SF-6D	Short-Form Six-Dimension health index
SHI	Statutory health insurance
SMDM	Society for Medical Decision Making

10. List of tables

11. List of figures

12. References

1. Mulrow, C.D., *The medical review article: state of the science.* Ann Intern Med, 1987. 106(3): p. 485-8.
2. Oxman, A.D. and G.H. Guyatt, *The science of reviewing research.* Ann N Y Acad Sci, 1993. 703: p. 125-33; discussion 133-4.
3. Cochrane, A.L., *Effectiveness and efficiency: random reflections on health services*. Taylor & Francis, 1999. 1-103.
4. Sackett, D.L., *Rules of evidence and clinical recommendations on the use of antithrombotic agents.* Chest, 1989. 95(2 Suppl): p. 2s-4s.
5. Greenhalgh, T., *How to read a paper. Getting your bearings (deciding what the paper is about).* Bmj, 1997. 315(7102): p. 243-6.
6. Chalmers, I., *The Cochrane collaboration: preparing, maintaining, and disseminating systematic reviews of the effects of health care.* Ann N Y Acad Sci, 1993. 703: p. 156-63; discussion 163-5.
7. Higgins, J., Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors), *Cochrane Handbook for Systematic Reviews of Interventions.* 2nd Edition ed. 2019: Chichester (UK): John Wiley & Sons.
8. Mikulic, M. *Global pharmaceutical industry - statistics & facts.* . 2023 [cited 2023 Jun]; Available from: https://www.statista.com/topics/1764/global-pharmaceutical-industry/.
9. EUnetHTA. *HTA definition. Internet Online edition*. 2007; Available from: http://www.eunethta.net/HTA/.
10. Kristensen, F.B., et al., *Identifying the Need for Good Practices in Health Technology Assessment: Summary of the ISPOR HTA Council Working Group Report on Good Practices in HTA.* Value in Health, 2019. 22(1): p. 13-20.
11. NICE. *Guide to the methods of technology appraisal 2013. Process and methods [PMG9]*. 2013 [cited 2023 Jan]; Available from: nice.org.uk/process/pmg9.
12. NICE. *Single technology appraisal and highly specialised technologies evaluation: User guide for company evidence submission template*. 2015 (updated in 2022) [cited 2023 Jan]; Available from: www.nice.org.uk/process/pmg24.
13. IQWIG. *Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. General Methods. Version 6.1 of 24 January 2022*. [cited 2023 Mar]; Available from: https://www.iqwig.de/en/about-us/methods/methods-paper/.
14. ICER, *A Guide to ICER's Methods for Health Technology Assessment. Institute for clinical and economic review. 2020: ICER. p: 1-43. Available from: https://icer.org/wp-content/uploads/2021/01/ICER_HTA_Guide_102720.pdf*.
15. PBAC. *Pharmaceutical Benefits Advisory Committee. Guidelines for preparing a submission to the Pharmaceutical Benefits Advisory Committee. Version 5.0, 2016*. [cited 2023 Mar]; Available from: https://pbac.pbs.gov.au/.
16. HAS. *Choices in methods for economic evaluation. Methodology Guide. Haute Autorité de Santé*. 2020 [cited 2023 Mar]; Available from: https://www.has-sante.fr/jcms/r_1499251/en/choices-in-methods-for-economic-evaluation.
17. Khangura, S., et al., *Evidence summaries: the evolution of a rapid review approach.* Syst Rev, 2012. 1:10: p. 1-9.

18. van der Knaap, L.M., et al., *Combining Campbell Standards and the Realist Evaluation Approach:The Best of Two Worlds?* American Journal of Evaluation, 2008. 29(1): p. 48-57.
19. Gordon, M., et al., *Twelve tips for undertaking a focused systematic review in medical education.* Med Teach, 2019. 41(11): p. 1232-1238.
20. Swingler, G.H., J. Volmink, and J.P. Ioannidis, *Number of published systematic reviews and global burden of disease: database analysis.* Bmj, 2003. 327(7423): p. 1083-4.
21. Tricco, A.C., et al. *Rapid reviews to strengthen health policy and systems: a practical guide. World Health Organization*. 2017 [cited 2023 Mar]; Available from: https://apps.who.int/iris/handle/10665/258698.
22. Collins, A., et al. *The Production of Quick Scoping Reviews and Rapid Evidence Assessments: A How To Guide*. 2015 [cited 2023 May]; Available from: https://www.gov.uk/government/publications/the-production-of-quick-scoping-reviews-and-rapidevidence-assessments.
23. Dobbins, M. *Rapid Review Guidebook: Steps for conducting a rapid review. National Collaborating Centre for Methods and Tools (NCCMT)*. 2017 [cited 2022 Sep]; Available from: https://www.nccmt.ca/capacity-development/rapid-review-guidebook.
24. Garritty, C., et al., *Cochrane Rapid Reviews Methods Group offers evidence-informed guidance to conduct rapid reviews.* Journal of Clinical Epidemiology, 2021. 130: p. 13-22.
25. Garritty, C., et al., *Cochrane Rapid Reviews Methods Group to play a leading role in guiding the production of informed high-quality, timely research evidence syntheses.* Syst Rev, 2016. 5:184: p. 1-5.
26. Plüddemann, A., et al., *Redefining rapid reviews: a flexible framework for restricted systematic reviews.* BMJ Evid Based Med, 2018. 23(6): p. 201-203.
27. Wilson, M.G., et al., *Paper 3: Selecting rapid review methods for complex questions related to health policy and system issues.* Systematic Reviews, 2021. 10:286: p. 1-8.
28. Haby, M.M., et al., *What are the best methodologies for rapid reviews of the research evidence for evidence-informed decision making in health policy and practice: a rapid review.* Health Res Policy Syst, 2016. 14:83: p. 1-12.
29. WHO. *Global Health Estimates: Life expectancy and leading causes of death and disability*. 2020 [cited 2023 Jun]; Available from: https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates.
30. IHME. *GBD Results. University of Washington.* 2020 [cited 2023 May]; Available from: https://vizhub.healthdata.org/gbd-results/.
31. Glass, G.V., *Primary, Secondary, and Meta-Analysis of Research.* Educational Researcher, 1976. 5(10): p. 3-8.
32. Begg, C., M. Cho, S. Eastwood, R. Horton, D. Moher, I. Olkin, R. Pitkin, D. Rennie, K. F. Schulz, D. Simel and D. F. Stroup, *Improving the quality of reporting of randomized controlled trials. The CONSORT statement.* Jama, 1996. 276(8): p. 637-639.
33. Moher, D., A. Liberati, J. Tetzlaff and D. G. Altman, *Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement.* PLoS Med, 2009. 6(7): e1000097: p. 1-6.
34. J. Thomas, J.B., S. Graziosi, *EPPI-Reviewer 4.0 software for research synthesis. EPPI Centre Software. 2023 London: Social Science Research Unit, Institute of Education, University of London. [Software] Available from: https://eppi.ioe.ac.uk/eppireviewer-web/home*. 2010.

35. The Endnote Team, *Endnote. [Software] Available from: https://endnote.com/*. 2013, Clarivate: Philadelphia, PA.
36. ICD-10-CM. *ICD-10-CM Chapters*. [cited 2023 May]; Available from: https://icd.codes/icd10cm.
37. Crowley, P., I. Chalmers, and M.J. Keirse, *The effects of corticosteroid administration before preterm delivery: an overview of the evidence from controlled trials.* Br J Obstet Gynaecol, 1990. 97(1): p. 11-25.
38. WHO. *WHO reveals leading causes of death and disability worldwide: 2000-2019*. 2020 [cited 2023 May]; Available from: https://www.who.int/news/item/09-12-2020-who-reveals-leading-causes-of-death-and-disability-worldwide-2000-2019.
39. Evans, K., *Real World Evidence: Can We Really Expect It to Have Much Influence?* Drugs Real World Outcomes., 2019. 6(2):43-45. .
40. *World Health Organization website. Health Technology Assessment.* [cited 2023 Jan]; Available from: https://www.who.int/health-topics/health-technology-assessment.
41. Kristensen, F.B., et al., *Practical tools and methods for health technology assessment in Europe: Structures, methodologies, and tools developed by the European network for Health Technology Assessment, EUnetHTA.* International Journal of Technology Assessment in Health Care, 2009. 25(S2): p. 1-8.
42. Kristensen, F.B., et al., *European network for Health Technology Assessment, EUnetHTA: Planning, development, and implementation of a sustainable European network for Health Technology Assessment.* International Journal of Technology Assessment in Health Care, 2009. 25(S2): p. 107-116.
43. European Parliament, C.o.t.E.U., *Regulation (EU) 2021/2282 of the European Parliament and of the Council of 15 December 2021 on health technology assessment and amending Directive 2011/24/EU (Text with EEA relevance). Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32021R2282*.
44. Fontrier, A.-M., E. Visintin, and P. Kanavos, *Similarities and Differences in Health Technology Assessment Systems and Implications for Coverage Decisions: Evidence from 32 Countries.* PharmacoEconomics - Open, 2022. 6(3): p. 315-328.
45. Angelis, A., A. Lange, and P. Kanavos, *Using health technology assessment to assess the value of new medicines: results of a systematic review and expert consultation across eight European countries.* Eur J Health Econ, 2018. 19(1): p. 123-152.
46. *EUPATI Toolbox. Health Technology Assessment process: Fundamentals.* [cited 2023 May]; Available from: https://toolbox.eupati.eu/resources/health-technology-assessment-process-fundamentals/.
47. *MP Europe website. Health technology assessment (HTA).* [cited 2023 Jan]; Available from: https://www.mpeurope.org/what-we-do/publications/qas/health-technology-assessment-hta/.
48. CRD, *Systematic Reviews. CRD's guidance for undertaking reviews in health care*. Centre for Reviews and Dissemination, University of York, 2008.
49. Drummond, M.F., et al., *Methods for the economic evaluation of health care programmes*. Oxford university press, 2015.
50. CADTH, *Canada's Drug and Health Technology Agency. Procedures for the CADTH Common Drug Review and Interim Plasma Protein Product Review. V 4.0*. 2020, CADTH. p. 1-141.
51. CADTH, *Canada's Drug and Health Technology Agency. CADTH METHODS AND GUIDELINES. Guidelines for the Economic Evaluation of Health Technologies: Canada. v 1.0 (4th Edition)*. 2017, CADTH. p. 1-76.
52. CADTH. *Canada's Drug and Health Technology Agency. Proposal to Revise Category 1 Requirements for CADTH Single Technology Reviews*. 2019 [cited 2023 Jan];

Available from: https://www.cadth.ca/cadth-pharmaceutical-reviews-update-issue-9.
53. C2H, *Center for Outcomes Research and Economic Evaluation for Health, National Institute of Public Health (C2H)Guideline for Preparing Cost-Effectiveness Evaluation to the Central Social Insurance Medical Council. Version 2.0*.
54. EUnetHTA, *EUnetHTA JA3WP6B2-2 Authoring Team. Process of information retrieval for systematic reviews and health technology assessments on clinical effectiveness. Methodological Guidelines. Diemen (The Netherlands): EUnetHTA; 2019. Available from https://www.eunethta.eu/*.
55. EUnetHTA. *EUnetHTA JA2WP 7. LEVELS OF EVIDENCE: Internal validity of randomised controlled trials. Adapted version*. 2015; Available from: https://www.eunethta.eu/.
56. Moher, D., et al., *Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement.* Systematic Reviews, 2015. 4(1): p. 1.
57. Thokala, P., J.J. Carlson, and M. Drummond, *HTA'd in the USA: A Comparison of ICER in the United States with NICE in England and Wales.* Journal of Managed Care & Specialty Pharmacy, 2020. 26(9): p. 1162-1170.
58. van der Knaap, L.M., et al., *Combining Campbell Standards and the Realist Evaluation Approach: The Best of Two Worlds?* American Journal of Evaluation, 2008. 29(1): p. 48-57.
59. Baumeister, A., et al., *The quality of systematic reviews and other synthesis in the time of COVID-19.* Epidemiology and Infection, 2021. 149, e182: p. 1-8.
60. Yu, Y., et al., *Assessment of the quality of systematic reviews on COVID-19: A comparative study of previous coronavirus outbreaks.* Journal of medical virology, 2020. 92(7): p. 883-890.
61. WHO. *Rapid Reviews to Strengthen Health Policy and Systems: A Practical Guide*. 2017 [cited 2023 Jan]; Available from: https://ahpsr.who.int/publications/i/item/2017-08-10-rapid-reviews-to-strengthen-health-policy-and-systems-a-practical-guide.
62. Plüddemann, A., et al., *Redefining rapid reviews: a flexible framework for restricted systematic reviews.* BMJ Evidence-Based Medicine, 2018. 23(6): p. 201.
63. Pandor, A., et al., *Delphi consensus reached to produce a decision tool for SelecTing Approaches for Rapid Reviews (STARR).* J Clin Epidemiol, 2019. 114: p. 22-29.
64. Tricco, A.C., et al., *An international survey and modified Delphi approach revealed numerous rapid review methods.* J Clin Epidemiol, 2016. 70: p. 61-7.
65. Tricco, A.C., et al., *Rapid reviews and the methodological rigor of evidence synthesis: a JBI position statement.* JBI Evidence Synthesis, 2022. 20(4): p. 944-949.
66. Serralde-Zúñiga, A.E., et al., *Fluoxetine for adults who are overweight or obese.* Cochrane Database Syst Rev, 2019. 10; Art. No.: CD011688: p. 1-158.
67. Madsen, B.K., et al., *Melatonin for preoperative and postoperative anxiety in adults.* Cochrane Database Syst Rev, 2020. 12; Art. No.: CD009861: p. 1-101.
68. Dushianthan, A., et al., *Immunonutrition for acute respiratory distress syndrome (ARDS) in adults.* Cochrane Database Syst Rev, 2019. 1; Art. No.: CD012041: p. 1-57.
69. Lumley, T., *Random effects (DerSimonian-Laird) meta analysis. Package rmeta version 3.0. p. 1-17. 2018 Available from: https://cran.r-project.org/web/packages/rmeta/rmeta.pdf*.
70. Schwarzer, G., *General Package for Meta-Analysis. Version 6.2-1, p. 1-218. 2023 Available from: https://cran.r-project.org/web/packages/meta/meta.pdf*.
71. Levine, L.R., S. Rosenblatt, and J. Bosomworth, *Use of a serotonin re-uptake inhibitor, fluoxetine, in the treatment of obesity.* Int J Obes, 1987. 11 Suppl 3: p. 185-90.

72. Wurtman, J.J., et al., *Dexfenfluramine, Fluoxetine, and Weight Loss Among Female Carbohydrate Cravers.* Neuropsychopharmacology, 1993. 9: p. 201-210.
73. Pedrinola, F.R., Wagner L; Lima, Nicolay; Medeiros Neto, Geraldo, *Fluoxetine induced weigth loss in overweight subjects: a double-blind placebo-controlled trial.* Arquivos Brasileiros de Endocrinologia e Metabologia, 1993. 37: p. 31-3.
74. Jain, N., et al., *Effect of oral melatonin on patients' anxiety scores and dose requirement of propofol during bispectral index-guided induction of general anesthesia.* The Indian Anaesthetists Forum, 2019. 20: p. 16.
75. Marzban, S., et al., *Comparison of the effect of Melatonin and Gabapentin on pain and anexiety in patients undergoing cataract surgery with Phacoemulsification with topical anesthesia.* Anesthesiology and Pain, 2016. 7(3): p. 1-10.
76. Hosseini, V.S., et al., *The Efficacy of Melatonin, Clonidine and Gabapentin in Reducing Preoperative Anxiety and Postoperative Pain in Patients Undergoing Laparoscopic Cholecystectomy: A Randomized Clinical Trial.* Arch Anesth & Crit Care., 2015. 1(4): p. 120-125.
77. Tsafnat, G., et al., *Systematic review automation technologies.* Syst Rev, 2014. 3: p. 74.
78. Borah, R., et al., *Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry.* BMJ Open, 2017. 7(2): p. e012545.
79. OpenAI. *ChatGPT*. [cited 2023 Mar]; Available from: https://openai.com/blog/chatgpt.
80. Mahesh, B., *Machine Learning Algorithms - A Review.* International Journal of Science and Research, 2020. 9(1): p. 381-386.
81. Brown, T., et al., *Language models are few-shot learners.* Advances in neural information processing systems, 2020. 33: p. 1877-1901.
82. Ouyang, L., et al., *Training language models to follow instructions with human feedback.* Advances in Neural Information Processing Systems, 2022. 35: p. 27730-27744.
83. De Angelis, L., et al., *ChatGPT and the rise of large language models: the new AI-driven infodemic threat in public health.* Front Public Health, 2023. 11: p. 1166120.
84. Qureshi, R., et al., *Are ChatGPT and large language models "the answer" to bringing us closer to systematic review automation?* Systematic Reviews, 2023. 12(1): p. 72.
85. Wolfe, C.R. *Specialized LLMs: ChatGPT, LaMDA, Galactica, Codex, Sparrow, and More*. 2023 [cited 2023 May]; Available from: https://towardsdatascience.com/specialized-llms-chatgpt-lamda-galactica-codex-sparrow-and-more-ccccdd9f666f.
86. Cierco Jimenez, R., Lee, T., Rosillo, N. et al. , *Machine learning computational tools to assist the performance of systematic reviews: A mapping review. .* BMC Med Res Methodol 22, 322 (2022).
87. Thomas, J., Graziosi, S., Brunton, J., Ghouze, Z., O'Driscoll, P., & Bond, M. Koryakina A, *EPPI-Reviewer: advanced software for systematic reviews, maps and evidence synthesis.* EPPI Centre, UCL Social Research Institute, University College London, 2022.
88. Borden, J., *SysRev software. 2023 [Software] Available from: https://sysrev.com/*.
89. Mourad Ouzzani, H.H., Zbys Fedorowicz, Ahmed Elmagarmid, *Rayyan — a web and mobile app for systematic reviews. Systematic Reviews 5, 210 (2016). https://doi.org/10.1186/s13643-016-0384-4*.
90. Munn, Z., Aromataris, E., Tufanaru, C., Stern, C., Porritt, K., Farrow, J., Lockwood, C., Stephenson, M., Moola, S., Lizarondo, L. and McArthur, A., *SUMARI software. 2018 [Software] Available from: https://sumari.jbi.global/*.

91. A-Evidence ApS, *Silvi.ai software. 2018 [Software] Available from: https://www.silvi.ai/.*
92. Evidence Prime Group, *Evidence Prime software. 2023 [Software] Available from:https://www.evidenceprime.com/.*
93. Wagner, G., R. Lukyanenko, and G. Paré, *Artificial intelligence and the conduct of literature reviews.* Journal of Information Technology, 2022. 37(2): p. 209-226.
94. King, R.D., et al., *The automation of science.* Science, 2009. 324(5923): p. 85-9.
95. Library Research Guides. *Systematic Reviews. Information on how to conduct systematic reviews in the health sciences.* 2023 [cited 2023 May]; Available from: https://guides.lib.umich.edu/c.php?g=283340&p=2126706.
96. Hemingway P. and Breteton N., *What is a systematic review? .* Hayward Medical Communications, a division of Hayward Group Ltd., 2009. 2: p. 1-8.
97. Li L., S.H.E., Atun R., Tudor Car L., *Search strategies to identify observational studies in MEDLINE and Embase (Review). Cochrane Database Syst Rev. 2019 Mar 12;3(3):MR000041, p. 1-35.*
98. Univeristy of Exeter Library. *Searching Online: Effective Library Research*. 2023 [cited 2023 Mar]; Available from: https://libguides.exeter.ac.uk/searching#:~:text=Identify%20your%20search%20terms%20Map%20out%20your%20keywords,Experts%20%2F%20Key%20writers%20Terminology%20change%20over%20time.
99. University of Leeds. *Literature searching explained. Develop a search strategy.* 2023 [cited 2023 Mar]; Available from: https://library.leeds.ac.uk/info/1404/literature_searching/14/literature_searching_explained/4.
100. Doust JA, P.E., Sanders S, Glasziou PP, *Identifying studies for systematic reviews of diagnostic tests was diFicult due to the poor sensitivity and precision of methodologic filters and the lack of information in the abstract.* Journal of Clinical Epidemiology, 2005. 58: p. 444-449.
101. Fraser C, M.A., Burr J. , *Identifying observational studies of surgical interventions in MEDLINE and EMBASE.* BMC Medical Research Methodology, 2006. 6: p. 1-9.
102. Wright K., G.J., Lefebvre C., on behalf of the ISSG. *The InterTASC Information Specialists' Sub-Group Search Filter Resource (SFR). HLG Conference 2016.* [cited 2023 Mar]; Available from: https://cdn.ymaws.com/www.cilip.org.uk/resource/group/12a7a632-bba2-4306-9f70-df0047d097d7/annual_conference/hlg_conference_2016/posters/poster_2_-_the_intertasc_inf.pdf.
103. Harvard Countway Library. *Systematic Reviews and Meta Analysis. A resource for finding data sources, filters, and standards to support systematic searches of the biomedical literature.* 2023 [cited 2023 Mar]; Available from: https://guides.library.harvard.edu/c.php?g=309982&p=2070466.
104. Relevo, R., *Chapter 4: effective search strategies for systematic reviews of medical tests.* Journal of General Internal Medicine, 2012. 27: p. 28-32.
105. Beynon R., L.M.M., McDonald S. et al. , *Search strategies to identify diagnostic accuracy studies in MEDLINE and EMBASE.* Cochrane Database of Systematic Reviews, 2013. 9: p. 1-101.
106. Hausner E, M.M., Richter B et al. , *Study filters for non-randomized studies of interventions consistently lacked sensitivity upon external validation.* BMC Medical Research Methodology, 2018. 18(1): p. 171.
107. Waffenschmidt S, N.-R.T., Hobson N et al. , *Development and validation of study filters for identifying controlled non-randomized studies in PubMed and Ovid MEDLINE.* Research Synthesis Methods, 2020. 11(5): p. 617-626.

108. David Gough, J.T., Sandy Oliver, *Clarifying differences between review designs and methods.* Systematic Reviews, 2012. 1:28: p. 1-9.
109. Dang, A., *Real-World Evidence: A Primer.* Pharmaceutical Medicine, 2023. 37: p. 25-36.
110. Gilmartin-Thomas Julia FM, L.D., Hopper I. , *Observational studies and their utility for practice.* Australian Presciber, 2018. 41: p. 82-85.
111. SIGN. *Healthcare Improvement Scotland. SIGN Search filters*. 2023 [cited 2023 May]; Available from: https://www.sign.ac.uk/what-we-do/methodology/search-filters/.
112. BMJ. *Search strategies. Evidence-based medicine (EBM) toolkit. BMJ Best Practice*. 2023 [cited 2023 May]; Available from: https://bestpractice.bmj.com/info/toolkit/learn-ebm/study-design-search-filters/.
113. Ovid. *Expert Searches. Ovid Tools & Resources Portal*. 2023 [cited 2023 May]; Available from: https://tools.ovid.com/ovidtools/expertsearches.html.
114. *ISSG Search Filter Resource [updated 10 June 2023]*. 2006 June 2023]; Available from: https://sites.google.com/a/york.ac.uk/issg-search-filters-resource/home
115. CADTH. *Search filters database.* 2023 [cited 2023 May]; Available from: https://searchfilters.cadth.ca/.
116. Mann, C.J., *Observational research methods. Research design II: cohort, cross sectional, and case-control studies.* Emerg Med J, 2003. 20(1): p. 54-60.
117. Morgan, S.J., et al., *Case Study Observational Research: A Framework for Conducting Case Study Research Where Observation Data Are the Focus.* Qual Health Res, 2017. 27(7): p. 1060-1068.
118. Chi, C.C., et al., *Safety of topical corticosteroids in pregnancy.* Cochrane Database Syst Rev, 2015. 2015(10): p. Cd007346.
119. Law, C.C.Y., et al., *Risk of Postoperative Infectious Complications From Medical Therapies in Inflammatory Bowel Disease: A Systematic Review and Meta-Analysis.* Inflamm Bowel Dis, 2020. 26(12): p. 1796-1807.
120. Taylor-Rowan, M., et al., *Anticholinergic burden (prognostic factor) for prediction of dementia or cognitive decline in older adults with no known cognitive syndrome.* Cochrane Database Syst Rev, 2021. 5(5): p. Cd013540.
121. Vinceti, M., et al., *Selenium for preventing cancer.* Cochrane Database Syst Rev, 2018. 1(1): p. Cd005195.
122. Iheozor-Ejiofor, Z., et al., *Water fluoridation for the prevention of dental caries.* Cochrane Database Syst Rev, 2015. 2015(6): p. Cd010856.
123. Kyrgiou, M., et al., *Obstetric outcomes after conservative treatment for cervical intraepithelial lesions and early invasive disease.* Cochrane Database Syst Rev, 2017. 11(11): p. Cd012847.
124. Lansbury, L.E., et al., *Corticosteroids as Adjunctive Therapy in the Treatment of Influenza: An Updated Cochrane Systematic Review and Meta-analysis.* Crit Care Med, 2020. 48(2): p. e98-e106.
125. Smith, R., et al., *Accuracy of measures for antiretroviral adherence in people living with HIV.* Cochrane Database Syst Rev, 2022. 7(7): p. Cd013080.
126. Mahé, A., et al., *The cosmetic use of skin-lightening products during pregnancy in Dakar, Senegal: a common and potentially hazardous practice.* Trans R Soc Trop Med Hyg, 2007. 101(2): p. 183-7.
127. Shah, R.C., et al., *Cognitive decline in older persons initiating anticholinergic medications.* PLoS One, 2013. 8(5): p. e64111.
128. Peleg, I., S. Morris, and C.G. Hames, *Is serum selenium a risk factor for cancer?* Med Oncol Tumor Pharmacother, 1985. 2(3): p. 157-63.

129. Menkes, M.S., et al., *Serum beta-carotene, vitamins A and E, selenium, and the risk of lung cancer.* N Engl J Med, 1986. 315(20): p. 1250-4.
130. Al-Alousi, W., et al., *Enamel mottling in a fluoride and in a non-fluoride community. A study.* British dental journal, 1975. 138(2): p. 56-60.
131. Herrera, M.T.A., et al., *Well water fluoride, dental fluorosis, and bone fractures in the Guadiana Valley of Mexico.* Fluoride, 2001. 34: p. 139-149.
132. de Crousaz, P., *[Observations on enamel opacities in Switzerland in relation to water or salt fluoridation].* Schweizerische Monatsschrift fur Zahnheilkunde = Revue mensuelle suisse d'odonto-stomatologie, 1982. 92(4): p. 332-344.
133. Forrest, J.R. and P.M. James, *A blind study of enamel opacities and dental caries prevalence after eight years of fluoridation of water.* British dental journal, 1965. 119(7): p. 319-322.
134. Franzolin Sde, O., et al., *Epidemiology of fluorosis and dental caries according to different types of water supplies.* Cien Saude Colet, 2010. 15 Suppl 1: p. 1841-7.
135. Goward, P.E., *Mottling on deciduous incisor teeth. A study of 5-year-old Yorkshire children from districts with and without fluoridation.* Br Dent J, 1982. 153(10): p. 367-9.
136. Haavikko, K. and A. Helle, *The prevalence and distribution of enamel defects in four districts with different fluoride contents in the drinking water.* Proc Finn Dent Soc, 1974. 70(5): p. 178-85.
137. Künzel, W. and F.S. Pádron, *Caries and dental fluorosis in Cuban children.* Caries Res, 1976. 10(2): p. 104-12.
138. Leverett, D., *Prevalence of dental fluorosis in fluoridated and nonfluoridated communities--a preliminary investigation.* J Public Health Dent, 1986. 46(4): p. 184-7.
139. Louw, A.J., S.R. Grobler, and W.K.T.J. van, *Degree of fluorosis in areas of South Africa with differing levels of fluoride in drinking water.* Gen Dent, 2002. 50(4): p. 352-6.
140. Mandinic, Z., et al., *Fluoride in drinking water and dental fluorosis.* Sci Total Environ, 2010. 408(17): p. 3507-12.
141. Marya, C.M., et al., *Exposure to high-fluoride drinking water and risk of dental caries and dental fluorosis in Haryana, India.* Asia Pac J Public Health, 2014. 26(3): p. 295-303.
142. McGrady, M.G., et al., *Dental fluorosis in populations from Chiang Mai, Thailand with different fluoride exposures - paper 1: assessing fluorosis risk, predictors of fluorosis and the potential role of food preparation.* BMC Oral Health, 2012. 12: p. 16.
143. Ray, S.K., et al., *Prevalence of dental fluorosis in relation to fluoride in drinking water in two villages of Varanasi (U.P.).* Indian J Public Health, 1982. 26(3): p. 173-8.
144. Russell, A.L. and E. Elvove, *Domestic water and dental caries. VII. A study of the fluoride-dental caries relationship in an adult population.* Public Health Rep (1896), 1951. 66(43): p. 1389-1401.
145. Saravanan, S., et al., *Prevalence of dental fluorosis among primary school children in rural areas of chidambaram taluk, cuddalore district, Tamil Nadu, India.* Indian J Community Med, 2008. 33(3): p. 146-50.
146. Scheinin, A., et al., *Prevalence of dental caries and dental health Iin relation to variable concentration of fluorides in drinking water; a clinical study on finnish school-children.* Acta Odontol Scand, 1964. 22: p. 229-54.
147. Segreto, V.A., et al., *A current study of mottled enamel in Texas.* J Am Dent Assoc, 1984. 108(1): p. 56-9.
148. Venkateswarlu, P., D. Narayana Rao, and K. Ranganatha Rao, *Studies in endemic fluorosis: Visakhapatnam and suburban areas; fluorine, mottled enamel and dental caries.* Indian J Med Res, 1952. 40(4): p. 535-48.

149. Wenzel, A., A. Thylstrup, and B. Melsen, *Skeletal development and dental fluorosis in 12--14-year-old Danish girls from a fluoride and a non-fluoride community.* Scand J Dent Res, 1982. 90(2): p. 83-8.
150. Zimmermann, E.R., *Fluoride and nonfluoride enamel opacities.* Public Health Rep (1896), 1954. 69(11): p. 1115-20.
151. Anderson, M.C., D. Horwell, and Z. Broby, *Outcome of Pregnancy after Laser Vaporization Conization.* Journal of Gynecologic Surgery, 1984. 1: p. 35-39.
152. Buller, R.E. and H.W. Jones, 3rd, *Pregnancy following cervical conization.* Am J Obstet Gynecol, 1982. 142(5): p. 506-12.
153. Guo, H.J., R.X. Guo, and Y.L. Liu, *Effects of loop electrosurgical excision procedure or cold knife conization on pregnancy outcomes.* Eur J Gynaecol Oncol, 2013. 34(1): p. 79-82.
154. Kristensen, G.B., *The outcome of pregnancy and preterm delivery after conization of the cervix.* Arch Gynecol, 1985. 236(3): p. 127-30.
155. Larsson, G., et al., *Outcome of pregnancy after conization.* Acta Obstet Gynecol Scand, 1982. 61(5): p. 461-6.
156. Moinian, M. and B. Andersch, *Does cervix conization increase the risk of complications in subsequent pregnancies?* Acta Obstet Gynecol Scand, 1982. 61(2): p. 101-3.
157. Poon, L.C., et al., *Large loop excision of transformation zone and cervical length in the prediction of spontaneous preterm delivery.* Bjog, 2012. 119(6): p. 692-8.
158. Sagot, P., et al., *Obstetrical prognosis for carbon dioxide laser conisation of the uterine cervix.* Eur J Obstet Gynecol Reprod Biol, 1995. 58(1): p. 53-8.
159. Jain, S., et al., *Hospitalized patients with 2009 H1N1 influenza in the United States, April-June 2009.* N Engl J Med, 2009. 361(20): p. 1935-44.
160. Haberer, J.E., et al., *Excellent adherence to antiretrovirals in HIV+ Zambian children is compromised by disrupted routine, HIV nondisclosure, and paradoxical income effects.* PLoS One, 2011. 6(4): p. e18505.
161. Jiamsakul, A., et al., *HIV multi-drug resistance at first-line antiretroviral failure and subsequent virological response in Asia.* J Int AIDS Soc, 2014. 17(1): p. 19053.
162. Okonji, J.A., et al., *CD4, viral load response, and adherence among antiretroviral-naive breast-feeding women receiving triple antiretroviral prophylaxis for prevention of mother-to-child transmission of HIV in Kisumu, Kenya.* J Acquir Immune Defic Syndr, 2012. 61(2): p. 249-57.
163. Acharya, S., *Dental caries, its surface susceptibility and dental fluorosis in South India.* Int Dent J, 2005. 55(6): p. 359-64.
164. Adair, S.M., et al., *Dental caries and fluorosis among children in a rural Georgia area.* Pediatric dentistry, 1999. 21(2): p. 81-85.
165. AlDosari, A.M., E.S. Akpata, and N. Khan, *Associations among dental caries experience, fluorosis, and fluoride exposure from drinking water sources in Saudi Arabia.* J Public Health Dent, 2010. 70(3): p. 220-6.
166. Arif, M., et al., *Assessment of fluoride level in groundwater and prevalence of dental fluorosis in Didwana block of Nagaur district, Central Rajasthan, India.* Int J Occup Environ Med, 2013. 4(4): p. 178-84.
167. Awadia, A.K., et al., *An attempt to explain why Tanzanian children drinking water containing 0.2 or 3.6 mg fluoride per liter exhibit a similar level of dental fluorosis.* Clin Oral Investig, 2000. 4(4): p. 238-44.
168. Awadia, A.K., et al., *Caries experience and caries predictors--a study of Tanzanian children consuming drinking water with different fluoride concentrations.* Clin Oral Investig, 2002. 6(2): p. 98-103.

169. Bao, L., Y. Li, and Y. Zhang, *[Dental caries and fluorosis among 12-year-old children with different fluoride exposure in Heilongjiang province].* Shanghai Kou Qiang Yi Xue, 2007. 16(6): p. 574-7.
170. Berndt, C., et al., *Fluorosis, caries and oral hygiene in schoolchildren on the Ombili Foundation in Namibia.* Oral Health Prev Dent, 2010. 8(3): p. 269-75.
171. Birkeland, J.M., et al., *Severity of dental caries among 12-year-old Sudanese children with different fluoride exposure.* Clin Oral Investig, 2005. 9(1): p. 46-51.
172. Butler, W.J., V. Segreto, and E. Collins, *Prevalence of dental mottling in school-aged lifetime residents of 16 Texas communities.* American Journal of Public Health, 1985. 75(12): p. 1408-1412.
173. Brothwell, D.J. and H. Limeback, *Fluorosis risk in grade 2 students residing in a rural area with widely varying natural fluoride.* Community Dent Oral Epidemiol, 1999. 27(2): p. 130-6.
174. Clarkson, J.J. and D.M. O'Mullane, *Prevalence of enamel defects/fluorosis in fluoridated and non-fluoridated areas in Ireland.* Community dentistry and oral epidemiology, 1992. 20(4): p. 196-199.
175. Correia Sampaio, F., et al., *Dental fluorosis and nutritional status of 6- to 11-year-old children living in rural areas of Paraiba, Brazil.* Caries research, 1999. 33(1): p. 66-73.
176. Cypriano, S., et al., *[Oral health of schoolchildren residing in areas with or without water fluoridation in Sorocaba, São Paulo State, Brazil].* Cad Saude Publica, 2003. 19(4): p. 1063-71.
177. Driscoll, W.S., et al., *Prevalence of dental caries and dental fluorosis in areas with negligible, optimal, and above-optimal fluoride concentrations in drinking water.* Journal of the American Dental Association (1939), 1986. 113(1): p. 29-33.
178. Ekanayake, L. and W. van der Hoek, *Dental caries and developmental defects of enamel in relation to fluoride levels in drinking water in an arid area of Sri Lanka.* Caries Res, 2002. 36(6): p. 398-404.
179. Eklund, S.A., et al., *High-fluoride drinking water, fluorosis, and dental caries in adults.* Journal of the American Dental Association (1939), 1987. 114(3): p. 324-328.
180. Ellwood, R.P. and D.M. O'Mullane, *Dental enamel opacities in three groups with varying levels of fluoride in their drinking water.* Caries research, 1995. 29(2): p. 137-142.
181. Ellwood, R.P. and D. O'Mullane, *The association between developmental enamel defects and caries in populations with and without fluoride in their drinking water.* Journal of public health dentistry, 1996. 56(2): p. 76-80.
182. Ermiş, R.B., F. Koray, and B.G. Akdeniz, *Dental caries and fluorosis in low- and high-fluoride areas in Turkey.* Quintessence Int, 2003. 34(5): p. 354-60.
183. Firempong, C., et al., *Soluble fluoride levels in drinking water-a major risk factor of dental fluorosis among children in Bongo community of Ghana.* Ghana Med J, 2013. 47(1): p. 16-23.
184. García-Pérez, A., M.E. Irigoyen-Camacho, and A. Borges-Yáñez, *Fluorosis and dental caries in Mexican schoolchildren residing in areas with different water fluoride concentrations and receiving fluoridated salt.* Caries Res, 2013. 47(4): p. 299-308.
185. Grobleri, S.R., A.J. Louw, and T.J. van Kotze, *Dental fluorosis and caries experience in relation to three different drinking water fluoride levels in South Africa.* Int J Paediatr Dent, 2001. 11(5): p. 372-9.
186. Hardwick, J.L., J. Teasdale, and G. Bloodworth, *Caries increments over 4 years in children aged 12 at the start of water fluoridation.* British dental journal, 1982. 153(6): p. 217-222.

187. Heintze, S.D., J.R. Bastos, and R. Bastos, *Urinary fluoride levels and prevalence of dental fluorosis in three Brazilian cities with different fluoride concentrations in the drinking water.* Community dentistry and oral epidemiology, 1998. 26(5): p. 316-323.
188. Heller, K.E., S.A. Eklund, and B.A. Burt, *Dental caries and dental fluorosis at varying water fluoride concentrations.* Journal of public health dentistry, 1997. 57(3): p. 136-143.
189. Hong, C.Y., et al., *Prevalence of mottled enamel after 12 years of water fluoridation in Chung-hsing New Village.* Journal of the Formosan Medical Association = Taiwan yi zhi, 1990. 89(3): p. 225-230.
190. Ibrahim, Y.E., A.A. Affan, and K. Bjorvatn, *Prevalence of dental fluorosis in Sudanese children from two villages with 0.25 and 2.56 ppm fluoride in the drinking water.* International journal of paediatric dentistry / the British Paedodontic Society [and] the International Association of Dentistry for Children, 1995. 5(4): p. 223-229.
191. Ismail, A.I., et al., *Prevalence of dental caries and dental fluorosis in students, 11-17 years of age, in fluoridated and non-fluoridated cities in Quebec.* Caries research, 1990. 24(4): p. 290-297.
192. Machiulskiene, V., et al., *Prevalence and extent of dental caries, dental fluorosis, and developmental enamel defects in Lithuanian teenage populations with different fluoride exposures.* Eur J Oral Sci, 2009. 117(2): p. 154-60.
193. Masztalerz, A., et al., *Fluorine and the dentition. [German].* Fortschritte der Kieferorthopadie, 1990. 51(4): p. 234-237.
194. Mella, S., et al., *Prevalence of dental fluorosis in Chile: a pilot study. [Spanish].* Revista medica de Chile, 1992. 120(8): p. 866-871.
195. Riordan, P.J. and J.A. Banks, *Dental fluorosis and fluoride exposure in Western Australia.* Journal of dental research, 1991. 70(7): p. 1022-1028.
196. Riordan, P.J., *Dental fluorosis decline after changes to supplement and toothpaste regimens.* Community Dent Oral Epidemiol, 2002. 30(3): p. 233-40.
197. Rugg-Gunn, A.J., S.M. al-Mohammadi, and T.J. Butler, *Effects of fluoride level in drinking water, nutritional status, and socio-economic status on the prevalence of developmental defects of dental enamel in permanent teeth in Saudi 14-year-old boys.* Caries research, 1997. 31(4): p. 259-267.
198. Rwenyonyi, C., et al., *Altitude as a risk indicator of dental fluorosis in children residing in areas with 0.5 and 2.5 mg fluoride per litre in drinking water.* Caries research, 1999. 33(4): p. 267-274.
199. Selwitz, R.H., et al., *Dental caries and dental fluorosis among schoolchildren who were lifelong residents of communities having either low or optimal levels of fluoride in drinking water.* J Public Health Dent, 1998. 58(1): p. 28-35.
200. Stephen, K.W., et al., *A blind caries and fluorosis prevalence study of school-children in naturally fluoridated and nonfluoridated townships of Morayshire, Scotland.* Community Dent Oral Epidemiol, 2002. 30(1): p. 70-9.
201. Szpunar, S.M. and B.A. Burt, *Dental caries, fluorosis, and fluoride exposure in Michigan schoolchildren.* Journal of Dental Research, 1988. 67(5): p. 802-6.
202. Tabari, E.D., et al., *Dental fluorosis in permanent incisor teeth in relation to water fluoridation, social deprivation and toothpaste use in infancy.* Br Dent J, 2000. 189(4): p. 216-20.
203. Tsutsui, A., M. Yagi, and A.M. Horowitz, *The prevalence of dental caries and fluorosis in Japanese communities with up to 1.4 ppm of naturally occurring fluoride.* J Public Health Dent, 2000. 60(3): p. 147-53.

204. Vignarajah, S., *Dental caries experience and enamel opacities in children residing in urban and rural areas of Antigua with different levels of natural fluoride in drinking water.* Community dental health, 1993. 10(2): p. 159-166.
205. Villa, A.E., S. Guerrero, and J. Villalobos, *Estimation of optimal concentration of fluoride in drinking water under conditions prevailing in Chile.* Community dentistry and oral epidemiology, 1998. 26(4): p. 249-255.
206. Warnakulasuriya, K.A., et al., *Determining optimal levels of fluoride in drinking water for hot, dry climates--a case study in Sri Lanka.* Community dentistry and oral epidemiology, 1992. 20(6): p. 364-367.
207. Zheng, C.L., *A survey of dental caries in Guangzhou after 18 years of community water fluoridation. [Chinese].* Zhonghua yu fang yi xue za zhi [Chinese journal of preventive medicine], 1986. 20(2): p. 79-82.
208. Braet, P.G., J.M. Peel, and D.W. Fenton, *A case controlled study of the outcome of pregnancy following loop diathermy excision of the transformation zone.* Journal of Obstetrics and Gynaecology, 1994. 14(2): p. 79-82.
209. Haffenden, D.K., et al., *Pregnancy following large loop excision of the transformation zone.* Br J Obstet Gynaecol, 1993. 100(11): p. 1059-60.
210. Delgado-Rodríguez, M., et al., *Different prognosis in hospitalized patients with influenza one season after the pandemic H1N1 influenza of 2009-2010 in Spain.* Influenza Other Respir Viruses, 2013. 7(6): p. 1336-42.
211. Kinikar, A.A., et al., *Predictors of mortality in hospitalized children with pandemic H1N1 influenza 2009 in Pune, India.* Indian J Pediatr, 2012. 79(4): p. 459-66.
212. Mady, A., et al., *Clinical experience with severe 2009 H1N1 influenza in the intensive care unit at King Saud Medical City, Saudi Arabia.* J Infect Public Health, 2012. 5(1): p. 52-6.
213. Ayoub, F., et al., *Evaluating the impact of vedolizumab on postoperative complications in inflammatory bowel disease patients.* Diseases of the Colon and Rectum, 2018. 61(5): p. e141.
214. Ferrante, M., et al., *Perioperative use of vedolizumab is not associated with short-term postoperative infectious complications in patients with ulcerative colitis undergoing (procto)colectomy with ileal pouch-anal anastomosis.* Gastroenterology, 2017. 152(5 Supplement 1): p. S581-S582.
215. Guasch, M., et al., *Prevalence and risk factors for postoperative septic complications in Crohn's disease.* Journal of Crohn's and Colitis, 2016. 10(Supplement 1): p. S350.
216. Nguyen, G.C., A. Elnahas, and T.D. Jackson, *The impact of preoperative steroid use on short-term outcomes following surgery for inflammatory bowel disease.* J Crohns Colitis, 2014. 8(12): p. 1661-7.
217. Schluender, S.J., et al., *Does infliximab influence surgical morbidity of ileal pouch-anal anastomosis in patients with ulcerative colitis?* Dis Colon Rectum, 2007. 50(11): p. 1747-53.
218. Shaib, Y.H., et al., *Surgical outcomes among inflammatory bowel disease patients undergoing colectomy: Results from a national database.* Gastroenterology, 2017. 152(5 Supplement 1): p. S363-S364.
219. Uchino, M., et al., *Infliximab administration prior to surgery does not increase surgical site infections in patients with ulcerative colitis.* International Journal of Colorectal Disease, 2013. 28(9): p. 1295-1306.
220. Wilson, M.Z., et al., *Organ space infection following ileocolectomy for Crohn's disease: a National Surgical Quality Improvement Project study.* Am J Surg, 2014. 208(5): p. 749-755.
221. Avong, Y.K., et al., *Adherence to Anti-Retroviral Therapy in North Central Nigeria.* Curr HIV Res, 2015. 13(4): p. 268-78.

222. Duarte, H.A., et al., *Relationship between viral load and behavioral measures of adherence to antiretroviral therapy in children living with human immunodeficiency virus in Latin America.* Braz J Infect Dis, 2015. 19(3): p. 263-71.
223. Ekstrand, M.L., et al., *Developing useful highly active antiretroviral therapy adherence measures for India: the Prerana study.* J Acquir Immune Defic Syndr, 2010. 53(3): p. 415-6.
224. Oette, M., et al., *Predictors of short-term success of antiretroviral therapy in HIV infection.* J Antimicrob Chemother, 2006. 58(1): p. 147-53.
225. Parienti, J.J., et al., *Average adherence to boosted protease inhibitor therapy, rather than the pattern of missed doses, as a predictor of HIV RNA replication.* Clin Infect Dis, 2010. 50(8): p. 1192-7.
226. Fex, G., B. Pettersson, and B. Akesson, *Low plasma selenium as a risk factor for cancer death in middle-aged men.* Nutr Cancer, 1987. 10(4): p. 221-9.
227. Heinen, M.M., et al., *Intake of antioxidant nutrients and the risk of skin cancer.* Eur J Cancer, 2007. 43(18): p. 2707-16.
228. Li, H., et al., *Plasma selenium levels associated with subsequent risk of prostate cancer.* American Journal of Urology Review, 2005. 3: p. 28-34.
229. Schober, S.E., et al., *Serologic precursors of cancer. I. Prediagnostic serum nutrients and colon cancer risk.* Am J Epidemiol, 1987. 126(6): p. 1033-41.
230. Sophie Beale, S.D., Julie Glanville, Carol Lefebvre, Dianne Wright, Rachael McCool, Danielle Varley, Charles Boachie, Cynthia Fraser, Jenny Harbour, Lynne Smith, *Choosing and using methodological search filters: searchers' views.* Health Info Libr J. 2014 Jun, 2006. 31(2): p. 133-47.
231. Leeflang, M.M., et al., *Use of methodological search filters to identify diagnostic accuracy studies can lead to the omission of relevant studies.* J Clin Epidemiol, 2006. 59(3): p. 234-40.
232. Wang, Z., et al., *Error rates of human reviewers during abstract screening in systematic reviews.* PLoS One, 2020. 15(1): p. e0227742.