Quality of evidence of the efficacy of therapeutic interventions on patient-important outcomes in Cochrane's systematic reviews' abstracts: A survey

Christine Maynié-François, Benoît Tudrej, Elodie Tawil, Florian Naudet, Caroline Huas, Denis Pouchain, Hélène Vaillant-Roussel, Nemat Jaafari, Rémy Boussageon

THERAPIES

HEADING: Clinical pharmacology

Quality of evidence of the efficacy of therapeutic interventions on patient-important outcomes in Cochrane's systematic reviews' abstracts: a survey

Quality of evidence and patient-important outcomes

Christine Maynié-François[a,b,*], Benoît Tudrej[c], Elodie Tawil[c], Florian Naudet[d,e], Caroline Huas[f,g], Denis Pouchain[h], Hélène Vaillant-Roussel[i], Nemat Jaafari[j], Rémy Boussageon[a,b]

[a] *Collège universitaire de médecine générale, université Claude Bernard Lyon 1, université de Lyon, 69008 Lyon, France*

[b] *Laboratoire de biométrie et biologie évolutive, équipe évaluation et modélisation des effets thérapeutiques, UMR CNRS 5558, 69008 Lyon, France*

[c] *Département de médecine générale, université de Poitiers, 86000 Poitiers, France*

[d] *University of Rennes 1, 35000 Rennes, France*

[e] *Clinical investigation center (INSERM 1414) and adult psychiatry department, Rennes university hospital, 35000 Rennes, France*

[f] *Fondation santé des étudiants de France, 75014 Paris, France*

[g] *CESP, INSERM U1018, université Paris-Sud, Université Paris-Saclay, UVSQ, 94800 Villejuif, France*

[h] *Département de médecine générale, université François Rabelais, 37000 Tours, France*

[i] *Clermont Auvergne university, faculty of medicine, UPU ACCePPT, department of general practice, 63000 Clermont-Ferrand, France*

[j] *Université de Poitiers, unité de recherche clinique Pierre-Deniker, centre hospitalier Henri-Laborit, 86021 Poitiers, France*

Received June 18, 2020; accepted September 1st, 2020

Corresponding author. Collège universitaire de médecine générale, 8 avenue Rockefeller, 69008 Lyon, France.

E-mail adress: christine.maynie-francois@univ-lyon1.fr (C. Maynié-François)

Summary

Objective. The objective of this study was to evaluate the proportion of therapeutics that have proven their efficacy on patient-important outcomes with a high quality of evidence among Cochrane systematic reviews (SRs). *Methods.* We surveyed a random sample of 400 SRs' abstracts published between September 2012 and December 2015, which compared therapeutic interventions with at least a placebo or no intervention control. The primary endpoint was the proportion of SRs with a statistically significant efficacy on a patient-important outcome and with a high quality of evidence. *Results*. Among the 400 abstracts surveyed, 32 (8%) found efficacy on a patient-important outcome with a high quality of evidence. Half of the 400 SRs (50.2%) evaluated a pharmacological intervention and 12% of these found efficacy of the intervention on a patient-important outcome with a reported high quality of evidence.

Conclusion. Based on an analysis of 400 abstracts of SRs from the Cochrane Collaboration, we found that there is a low number of therapeutic interventions which have proven their efficacy on patient-important outcomes with a high quality of evidence.

Abbreviations

EBM: evidence based medicine

Keywords: Systematic review; Evidence-based medicine; GRADE approach; Patient-important outcomeGRADE: grades of recommendation, assessment, development, and evaluation criteria

MA: meta analysis

SMD: standardized mean difference

PIO: patient-important outcomes

RCTs: randomized controlled trials

SRs: systematic reviews

WMD: weighted mean differences

Introduction

When therapeutic interventions can be tested, they must have their efficacy proven in high-quality studies, such as randomized controlled trials (RCTs), or systematic reviews (SRs) with or without meta-analysis (MA) of these high-quality RCTs [1,2]. However, evidence of efficacy is not always established in clinical practice. In 2013, the BMJ Clinical Evidence team reported that of 3000 treatments, half had an « unknown » efficacy: either not studied or studied but with no evidence of efficacy [3]. Only 11 % are considered effective with a good level of evidence (as defined by the quality of studies, RCTs, or RSs/MAs of RCTs). A 2007 study on 1016 Cochrane systematic reviews found that further evidence was needed in 96% of SRs to conclude about efficacy; the authors did not adjudicate on efficacy of interventions in 49 % of SRs [4]. A recent review found that only 13.5 % of the Cochrane SRs were based on high quality evidence according to the grades of recommendation, assessment, development, and evaluation (GRADE) criteria [5].

To be of interest for patients, treatments also have to demonstrate an efficacy on patient-important outcomes (PIO) [6] such as: a clinically relevant reduction of a given symptom, a lower absolute risk of disease, disability, or mortality, or a quality of life improvement. A recent study found that contrary to randomized controlled trials, 80% of the SRs considered the effect of treatments on patient-important outcomes, especially in Cochrane reviews (95 %) [7].

The objectives of the Cochrane collaboration are to bring high-quality and up-to-date information on the efficacy of healthcare interventions, to help healthcare professionals and patients in their decisions, and to promote evidence based medicine (EBM) [8]. Cochrane develops, updates and publishes systematic reviews with meta-analyses. Access to abstracts is open online. Cochrane systematic reviews are considered more valid than others [9] because of a better control of biases [10], a higher level of update [11] and a rigorous method they systematically publish [12]. Cochrane collaboration highly recommends the evaluation of treatments on PIO [13].

The aim of our study was to determine the proportion of therapeutic interventions evaluated by the Cochrane systematic reviews with 1) good quality evidence of their efficacy on 2) patient-important outcomes.

Methods

We conducted a review on a random sample of Cochrane open access abstracts.

The protocol was recorded on PROSPERO (CRD42016048606).

Inclusion and exclusion criteria

We considered all open access abstracts of systematic reviews on Cochrane Library website, published between September 2012 and December 2015. We considered abstracts only for feasibility reasons, and because they are an essential mode for disseminating research results to practicing physicians [14]. Moreover, abstracts are supposed to contain all the data needed for this study, namely a description of the main outcome, most important results, and authors' conclusions regarding the quality of evidence in their paper.

We included all systematic reviews which evaluated a therapeutic intervention, with placebo or the absence of intervention as one of the controls, to demonstrate the specific efficacy of the intervention.

We excluded reviews studying the efficacy of a given therapeutic intervention versus another intervention with no placebo or no intervention control: our aim was to study the evaluation of interventions' efficacy, not therapeutic strategies.

We excluded reviews on diagnostic or intervention methods, protocols, reviews of Cochrane magazines ("overviews"), dental reviews and reviews with no placebo control. Dental reviews were excluded based on the lack of expertise of the team in dentistry.

Extraction of data and sample building

A pre-defined number of 400 abstracts was established to obtain an accuracy of at least 5% under the extreme assumption of 50% of therapeutic interventions with good quality evidence of their efficacy on at least one patient-important outcome.

We first extracted all SRs' titles published within the pre-defined period. On this first dataset, we used the ALEA Excel® function to mix titles and randomly sample SRs. We then selected the first 400 abstracts which met inclusion criteria. Eligible titles and abstracts were independently selected by two authors (ET, RB). A third author (BT) was consulted in case of a disagreement.

Primary endpoint

The primary endpoint was the proportion of SRs with a statistically significant effect on a PIO with a high level of evidence.

Efficacy was considered when the intervention showed a statistically significant effect on at least one outcome with no consideration for the effect size or its clinical relevance.

A PIO was defined as any patient centered outcome, such as mortality, pain, disability, functional status, clinically active disease, or quality of life. Non patient centered outcomes such as biology, imaging etc., were not considered as PIO (noted N-PIO).

The level of evidence was extracted from the conclusion given by the authors in the abstract: high, moderate, low, and uncertain. When the level of evidence was not stated, we considered the report of the risk of bias or the quality of evidence. We considered the level of evidence as high when the abstract reported either a low risk of bias or a high quality of evidence. If none of these elements (level of evidence, quality of evidence, risk of bias) was presented, the level of evidence of the results was considered as "not reported in the abstract".

We chose a high standard primary endpoint: not only does it includes a clinically meaningful outcome (PIO), but it also considers the level of evidence. This choice was made on purpose to stick to standards of therapeutic decision: evidence of efficacy is needed when any treatment is discussed, and the level of this evidence is expected to be high. Only high-quality studies can give this type of high-quality evidence.

Secondary endpoints

The secondary endpoints included: the proportion of SRs with PIO with statistically significant effects whatever the level of evidence, the proportion and level of evidence of SRs with PIO and a non-statistically significant effect, the proportion and level of evidence of SRs with N-PIO and a non-statistically significant conclusion, the proportion and level of evidence of SRs with N-PIO and a statistically significant conclusion, and the number of RCTs in conclusive meta-analyses (whatever the outcome).

Other variables of interest

We categorized therapeutic interventions as: "Drug", "Pharmacological Treatments (including Vaccines, Gene Therapies)", "Surgical Treatments", "Alternative Medicine and Dietary Supplements", "Care Procedures: Anesthesia, Nursing, Radiotherapy, Skincare", "Psychotherapies and Behavioral

Therapies", "Education, communication strategies", "Rehabilitation (physical, rehabilitation, occupational therapy)", and "Other".

Statistical analysis

We performed a descriptive analysis and reported the average for the quantitative variables as well as the percentages for the qualitative variables. The accuracy of the estimates was estimated by calculating their 95% confidence interval.

Primary and secondary endpoints were described on the total of included abstract, and also for each type of therapeutic intervention.

Results

We identified 3051 SRs published between September 2012 and December 2015 (Fig. 1). After randomization as described in the methods, papers were screened to include the first 400 meeting with inclusion and exclusion criteria. Along the process, we excluded 178 papers (Fig. 1).

Characteristics of included systematic reviews

Among all 400 SRs, 55 were empty, meaning that no RCT was included in the review. The mean number of RCT included per RS was 14 [1-108]. Two thirds of SRs (n = 266) included one or more meta-analyses. Half of the SRs (n = 201) were about a pharmacological intervention. The second most common type of intervention was medical procedures (n = 61) [Table 1]. One in five abstracts (n = 72) did not mention the level of evidence [Table 2].

Primary endpoint

Among the 400 selected systematic reviews abstracts, 32 (8%; 95% CI = 5.7 – 11.1) found conclusive results on a PIO with a high level of evidence (Table 2).

Secondary endpoints

On the 400 selected abstracts, 297 (74.2%; 95% CI = 69.7 - 78.3) had at least one PIO (Table 2). Among the 32 SRs with a high level of evidence, two papers mentioned that the level of evidence was based on the GRADE tool.

Among the 186 SRs conclusive on a PIO, 153 (82.3%; CI 95% = 76.1 - 87.1) included one or more meta-analysis. The average number of RCTs included in the 186 SRs conclusive on a PIO was 17 [1-107] and 9 SRs (4.8%) included only one RCT. The level of evidence was not provided in 28 of the 186 SRs (15%).

Of the 400 selected abstracts, 111 (27.8%; 95% CI = 23.6 - 32.3) with at least one PIO were not conclusive.

Among SRs with no PIO, 32/400 (8%; 95% CI = 5.7 - 11.1) were conclusive and 16 SRs on the total of 400 were inconclusive (4%; 95% CI = 2.5 - 6.4).

Finally, among the 32 SRs with a statistically significant result on an N-PIO, two had a high level of evidence.

Characteristics of systematic reviews based on the type of intervention

Among the 201 SR evaluating pharmacological interventions, 107 (53.2%) had at least one PIO, but only 25 (12.4%) had a PIO and mentioned a high level of evidence; 54 (26.9%) had at least one PIO but were not statistically significant (Table 1).

Discussion

Main findings

The main objective of this study was to determine the proportion of Cochrane systematic reviews (SR) demonstrating the efficacy of therapeutics evaluated on at least one patient-important outcome with a reported high level of evidence.

Our choice to consider only patient-important outcomes was unusual compared to other SR studies. The primary objective of medical and clinical research should be to be useful to patients [6]. In comparison, other systematic reviews often considered the primary outcome regardless of its nature (PIO or N-PIO) [15,16]. It is interesting to note that although Fleming et al [5] considered the main outcome, the proportion they found of significant studies with GRADE criteria on main outcome (8.6%) was similar to the proportion we found of significant studies with high level of evidence on PIO (8%). This similarity is consistent with existing literature [7].

On a sample of 400 SRs' abstracts evaluating several therapeutics, 186 (46.5%) found a statistically significant efficacy on a PIO, but only 32 (8%) reported a high level of evidence for this result in the abstract.

This difference is of importance: given the reported level of evidence (high-quality or not), the conclusions drawn from the reported efficacy may vary and have consequences on clinical care.

This can be illustrated by the case of corticoids and their potential benefit in meningitis. In 2007, a Cochrane meta-analysis [17] including 18 RCTs found that corticoids in meningitis lowered global mortality (RR = 0.83, 95% CI = 0.71 - 0.99) and the risk of subsequent severe hearing loss (RR 0.65, 95%CI = 0.47 - 0.91). In 2015, a revised meta-analysis including new RCTs found that the benefit on global mortality became non-significant (RR 0.90, 95% CI = 0.80 - 1.01) while the benefit on hearing loss persisted (RR 0.67, 95% CI = 0.51 - 0.88). However, a secondary analysis [18] restricted to the four high-quality RCTs found non-significant results both on mortality (RR 1.00, 95 % CI = 0.88 - 1.14) and severe hearing loss (RR 0.90, 95% CI = 0.73 - 1.12).

Another example can be found in the Cochrane review on the use of antidepressants for the treatment of people with co-occurring depression and alcohol dependence [19]. The abstract states that there is low-quality evidence on the efficacy of antidepressants versus placebo to reduce the severity of depression evaluated with interviewer-rated scales at the end of trial (14 studies, 1074 participants, standardized mean difference [SMD] -0.27, 95% CI = -0.49 to -0.04). When studies with a high risk of bias are excluded, the difference of efficacy between antidepressants and placebo becomes non-significant (SMD -0.17, 95% CI = -0.39 to 0.04).

Limitations and strengths

The main limitation of the study lies in the choice to analyze only free access abstracts. There was no access to the complete data, and the level of evidence was noted as reported by the authors in the abstract. The level of evidence was unreported in 72 of the abstracts, and most notably in 29 of the abstracts with a significant result on a PIO. This evaluation is therefore subjected to more selection biases than Fleming's et al [5]. However, open access abstracts may be easy to read by clinicians in day-to-day practice. As stated in the methods section, we considered abstracts only because they are often what practicing physicians read in the first place or at all [14]. The similarity between our results based on abstracts only and Fleming's based on full papers may be of interest for clinical readers.

In this study, we did not evaluate the clinical relevance of statistically significant results (by considering for example the minimal important difference [20]), nor did we choose to include such a criterion in the primary endpoint. It is probable that with a clinical relevance inclusion criterion, the number of included SRs would be even lower. A recent meta-analysis on the effects of opioids on non-cancer chronic pain [21] found that, by taking into account the minimal important difference (defined as a 1.00 cm difference on a 10 cm visual analog pain scale), opioids had a limited clinical efficacy, even if their efficacy was statistically significant (weighted mean difference [WMD], -0.69 cm [95% CI = -0.82 - -0.56 cm]).

The findings of this study are consistent with the existing literature on the quality of evidence from Cochrane SRs [3,5]. In 2016, Fleming et al found that 45% of Cochrane SRs used the GRADE scale to assess the quality of evidence, and 8.6% had a significant result on the major outcome with a high level of evidence according to GRADE [5], which is similar to our main result (8% of SRs with significant results on a PIO with reported high level of evidence). In a sample of SRs not using GRADE, 44% had at least one significant result regardless of the outcome. This study adds to Fleming's by selecting clinical outcomes and by the description of several other parameters such as the number of empty or inconclusive SRs.

Finally, there is a lack of evidence for some therapeutic interventions. Our study found 55 (14%) empty SRs, meaning no RCT was available for the studied intervention. This number reached 40% for reviews in surgery, where RCT versus placebo or no treatment can be more difficult or even not possible.

In this review of a random sample of Cochrane's systematic reviews' abstracts, we found a low proportion of reviews on therapeutics with a statistically significant efficacy on at least one patient-important outcome with a high level of evidence (8%). This can have important consequences in clinical practice since good-quality evidence on the efficacy of therapeutics is found to be scarce.

Prasad and colleagues in 2013 found that among 363 articles testing standard of care, 40% reversed that practice [22]. In a recent review, among 3017 randomized controlled trials from three leading medical journals (the Journal of American Medical Association, the Lancet, and the New England Journal of Medicine) [23], the authors identified 396 (13%) medical reversals, 80% of which were confirmed by a systematic review. As we previously discussed, considering only high-quality evidence can change the result of a meta-analysis [17-19]. Our findings that only 32 (8%) of the 400 abstracts surveyed found efficacy on a patient-important outcome with high quality of evidence suggests

that up to 92% of reviewed medical practices may have few if no therapeutic interest and may be reversed.

Disclosure of interest

Authors have no conflict of interest to report.

References

[1] Guyatt G, Cairns J, Churchill D, Cook D, Haynes B, Hirsh J, et al. Evidence-based medicine. A new approach to teaching the practice of medicine. JAMA 1992;268:2420-5.

[2] Howick J. The philosophy of evidence based medicine. BMJ Books. 2011. Print ISBN:9781405196673. Online ISBN:9781444342673. DOI:10.1002/9781444342673.

[3] Attia A. Clinical evidence. Adaptation of international evidence based clinical practice guidelines: The ADAPTE process. Middle East Fertility Society Journal 2013;18:123–6. https://www.sciencedirect.com/science/article/pii/S1110569013000435. [Accesses September 11, 2020].

[4] El Dib RP, Atallah AN, Andriolo RB. Mapping the Cochrane evidence for decision making in health care. J Eval Clin Pract 2007;13:689-92.

[5] Fleming PS, Koletsi D, Ioannidis JPA, Pandis N. High quality of the evidence for medical and other health-related interventions was uncommon in Cochrane Systematic Reviews. J Clin Epidemiol 2016;78:34-42. J Clin Epidemiol 2016 Oct;78:34-42.

[6] Gandhi GY, Murad MH, Fujiyoshi A, Mullan RJ, Flynn DN, Elamin MB, et al. Patient-important outcomes in registered diabetes trials. JAMA 2008;299:2543-9.

[7] Ameur H, Ravaud P, Fayard F, Riveros C, Dechartres A. Systematic reviews of therapeutic interventions frequently consider patient-important outcomes. J Clin Epidemiol 2017;84:70-7.

[8] Levin A. The Cochrane Collaboration. Ann Intern Med 2001;135:309-12.

[9] Teich ST, Lang LA, Demko CA. Characteristics of the Cochrane Oral Health Group systematic reviews. J Dent Educ 2015;79:5-15.

[10] Jørgensen A, Hilden J, Gøtzsche P. Cochrane reviews compared with industry supported meta-analyses and other meta-analyses of the same drugs: systematic review. BMJ 2006;333(7572):782.

[11] Jadad AR, Cook DJ, Jones A, et al. Methodology and reports of systematic reviews and meta-analyses: a comparison of Cochrane reviews with articles published in paper-based journals. JAMA 1998;15:278-80.

[12] Shea B, Boers M, Grimshaw M, Hamel C, Bouteret LM. Does updating improve the methodological and reporting quality of systematic reviews? BMC Med Res Methodol 2006;6:27-34.

[13] Higgins JPT, Deeks JJ, Altman DG. Special topics in statistics. In: Higgins JPT, Green S, eds. Cochrane handbook for systematic reviews of interventions. Version 5.1.0. Cochrane Collaboration, 2011. http://handbook-5-1.cochrane.org/. [Accessed September 11, 2020].

[14] Barry HC, Ebell MH, Shaughnessy AF, Slawson DC, Nietzke F. Family physicians' use of medical abstracts to guide decision making: style or substance? J Am Board Fam Pract 2001;14(6):437-42.

[15] Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and reporting characteristics of systematic reviews. Plos Med 2007;4.e78.

[16] Davey J, Turner RM, Clarke MJ, Higgins JPT. Characteristics of meta-analyses and their component studies in the Cochrane database of systematic reviews: a cross-sectional, descriptive analysis. BMC Med Res Methodol 2011; 11:160.

[17] Van de Beek D, de Gans J, McIntyre P. Corticosteroids for acute bacterial meningitis. Cochrane Database Syst Rev 2007;(1):CD004405.

[18] Brouwer MC, McIntyre P, Prasad K, et al. Corticosteroids for acute bacterial meningitis. Cochrane Database Syst Rev 2015;(9):CD004405.

[19] Agabio R, Trogu E, Pani PP. Antidepressants for the treatment of people with co-occurring depression and alcohol dependence. Cochrane Database Syst Rev 2018;4:CD008581.

[20] Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials 1989;10:407-15.

[21] Busse JW, Wang L, Kamaleldin M, Craigie S, Riva JJ, Montoya L, et al. Opioids for chronic noncancer pain: a systematic review and meta-analysis. JAMA 2018;320:2448-60.

[22] Prasad V, Vandross A, Toomey C, Cheung M, Rho J, Quinn S, et al. A decade of reversal: an analysis of 146 contradicted medical practices. Mayo Clin Proc 2013 Aug;88(8):790-8.

[23] Herrera-Perez D, Haslam A, Crain T, Gill J, Livingston C, Kaestner V, et al. A comprehensive review of randomized clinical trials in three medical journals reveals 396 medical reversals. eLife 2019;8:e45183.

Table 1. Distribution of the findings of systematic reviews according to the type of intervention

	Statistically significant				Non-significant		Empty
Intervention	**PIO n (%)**	**PIO high level of evidence n (%)**	**PIO moderate or low level of evidence n (%)**	**N-PIO n (%)**	**PIO n (%)**	**N-PIO n (%)**	
Total N = 400	186 (46.5)	32 (8)	125 (31.2)	32 (8)	111 (27.7)	16 (4)	55 (13.8)
Pharmacological N = 201	107 (53.2)	25 (12.4)	65 (32.3)	12 (6)	54 (26.9)	8 (4)	20 (9.9)
Surgical N =25	7 (28)	0	5 (20)	1 (4)	7 (28)	0	10 (40)
Care procedure N = 61	24 (39.3)	4 (6.6)	17 (27.9)	4 (6.5)	17 (27.9)	2 (3.3)	14 (23)
Alternative medicine N = 47	24 (51.1)	1 (2.1)	20 (42.5)	5 (10.6)	11 (23.4)	11 (23.4)	2 (4.5)
Re-education N = 16	5 (31.3)	0	4 (25)	3 (18.6)	7 (43.8)	0	1 (6.3)
Education N = 15	4 (26.7)	0	4 (26.7)	5 (33.3)	4 (26.7)	1 (6.7)	1 (6.7)
Psychotherapy N = 11	6 (54.5)	0	5 (45.4)	0	3 (27.3)	0	2 (18.2)
Other N = 24	9 (37.5)	2 (8.3)	5 (20.8)	2 (8.3)	8 (33.3)	0	5 (20.8)

N-PIO: non patient-important outcome; PIO: patient important outcome.

We did not report in this table the results of the 29 abstracts with a significant result on a PIO but with no mention of the level of evidence.

Table 2. Distribution of conclusions of 400 systematic reviews analyzed

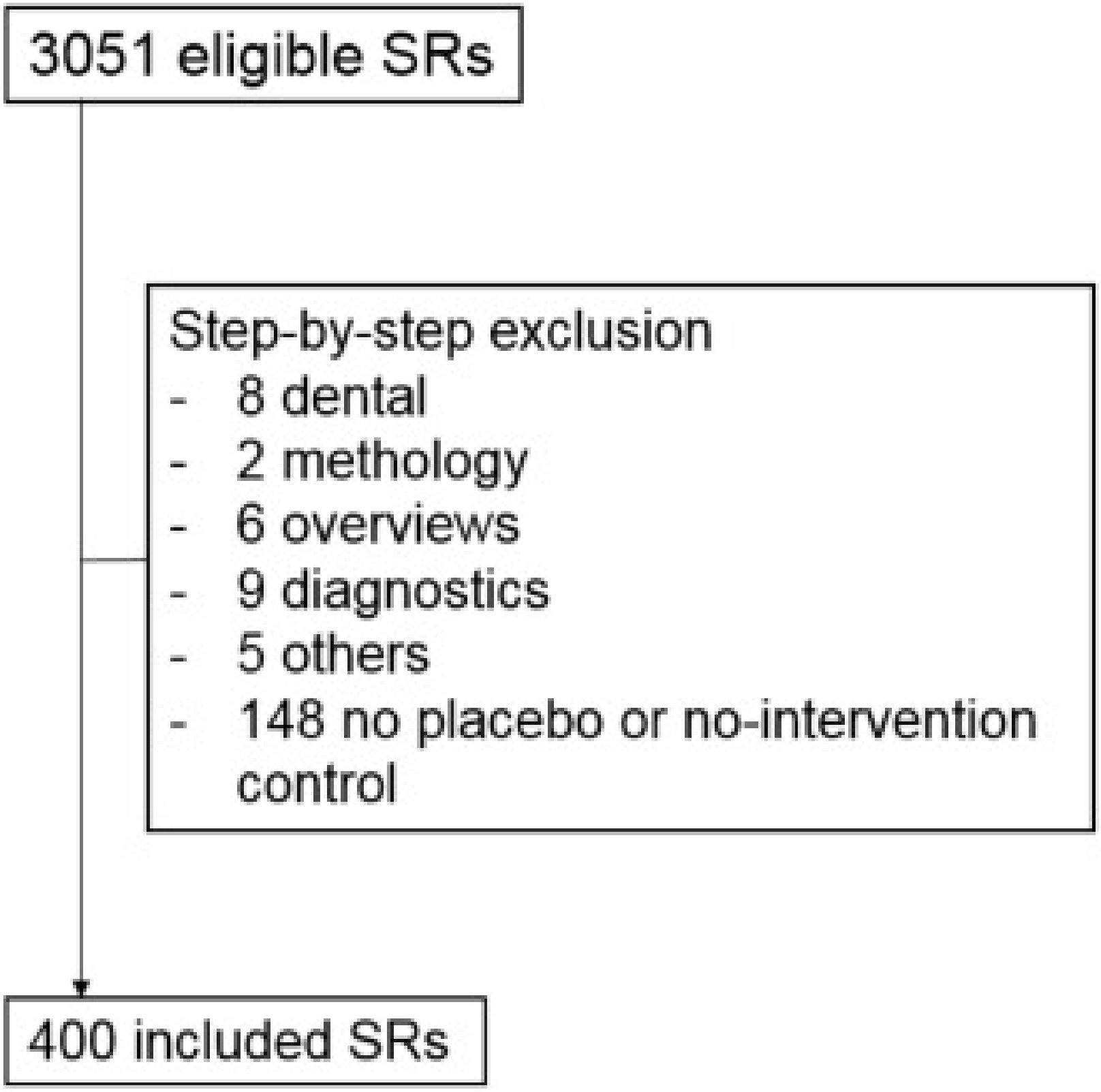
		Included SRs N = 400				
Outcome, n(%)		**PIO** 297 (74.2)		**N-PIO** 48 (12)		**Empty** 55 (13.8)
Statistical significance, n(%)		**Significant** 186 (46.5)	**Non significant** 111 (27.8)	**Significant** 32 (8)	**Non significant** 16 (4)	-
Level of evidence	**High, n(%)**	32 (8)	12 (3)	2 (0.5)	2 (0.5)	-
	Moderate or low, n(%)	125 (31.3)	72 (18)	19 (4.8)	9 (2.3)	-
	Unreported, n(%)	29 (7.3)	27 (6.8)	11 (2.8)	5 (1.3)	-

N-PIO: non-patient-important outcome; PIO: patient-important outcome; SRs: systematic reviews.

All percentages are calculated with the total number of included SRs, 400, as the denominator.

Figures

Figure 1. Flow chart of the included and excluded systematic reviews



SR: systematic review