

定性的システマティックレビューの手順と作法

前川賢治

Procedures and rules for conducting qualitative systematic reviews

Kenji Maekawa, DDS, PhD

抄 録

システマティックレビュー (SR) は、その統合のプロセスが定性的にのみ行われるか、定量的評価も加えられるかにより大別されるが、本稿では定性的 SR に焦点をあてる。どちらの SR も、あるクリニカルクエスチョンに対するその時点での結論を導き出すものであり、過去の経験的エビデンス収集のための作法は同一である。しかし、定性的 SR では統計学的統合は伴わずに、各一次研究の質的評価、批判的吟味を実施し、各エビデンスのプロファイルを示したうえで結論を導く。本稿では、定性的 SR を進めるにあたり、PRISMA フローダイアグラムの各ステップにおいて配慮、注意すべき点について、実例も含めて解説を行う。

キーワード

定性的システマティックレビュー、エビデンステーブル、質的評価、PRISMA フローダイアグラム

ABSTRACT

Systematic reviews (SRs) are broadly classified into two categories, qualitative and quantitative ones based on the process of data integration. This article focuses on the qualitative SRs. Since both SRs draw conclusions for a given clinical question at that point, the rules for collecting past empirical evidences are identical. However, qualitative SRs involve only qualitative evaluation and critical review of each primary study, without statistical integration. Then, certain conclusions are drawn based on the explicit profile of each piece of evidence. This article explains the points to consider and attentions to pay in each step of PRISMA flow diagram, when processing with qualitative SRs by indicating actual examples.

Key words:

Qualitative systematic review, Evidence table, Qualitative assessment, PRISMA flow diagram

I. 緒 言

システマティックレビュー (SR) とは、コクラン共同計画 (現コクラン) の 1992 年の定義によると「ある特定の課題に対して、全ての経験的エビデンスをあらかじめ定めた基準で網羅的に収集、評価、統合するもの」とされている¹⁾。SR は、収集した一次研究の結果 (エビデンス) の統合にあたり、得られた各エビデンスに対してバイアスの関与等に関する質的評

価、批判的吟味のみを実施し、各エビデンスのプロファイルを示したうえで一定の結論を導き出す定性的 SR と、統計学的手法による量的な評価を加えた定量的 SR に大別される。しかし、どちらの SR もあるクリニカルクエスチョンに対するその時点での結論を導き出すものであり、過去に実施された一次研究を収集する作法は同一である (図 1)。そして、その作法、手順は、PRISMA 声明²⁾ に定められたシステマティックレビューまたはメタアナリシスを報告する際の項目チェックリストに準拠することが、現在では望ましい

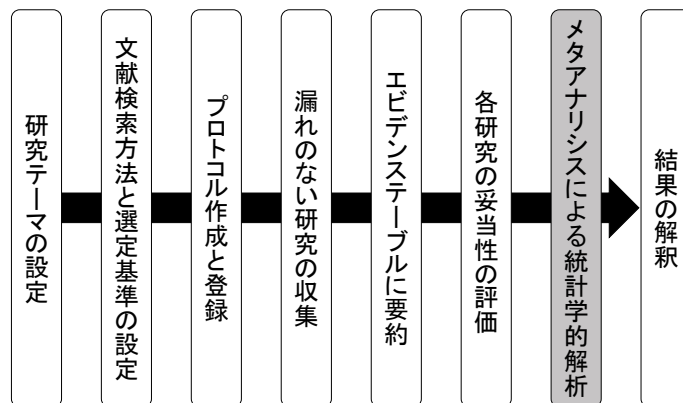


図1 システマティックレビューの基本的な進め方

システマティックレビューを進めるうえでの基本的なステップを示す。本稿で取り上げる定性的システマティックレビューでは、図中の「メタアナリシスによる統計学的解析」のステップは含まず、それまでのステップから得られた結果を解釈して結論をまとめる。

状況となっている。何故なら、最近ではSRの投稿に際し、PRISMA 声明に従って実施したことを本文内へ明文化することや、PRISMA チェックリストの提出を求める学術雑誌が増加しているからである。それ故、この声明はランダム化比較試験を始めとする介入研究の評価に重点を置いているものの、観察研究を対象としたSRにおいてもチェックリスト内で該当する項目については従ったうえで進めておくべきである。実際の手順としては、まず明確なPI(E)CO ならびに文献の包含、除外基準の設定が重要となる。そして検索のステップにおいては、各種特色のある検索データベースから複数選択し、網羅的・体系的に文献収集するための検索式の設定にも配慮が必要である。ここまでのプロトコルが作成された時点で、世界的なSRレジストりに登録したうえでのレビュー開始が推奨されている。最初に収集された文献は相当数であることも少なくなく、一次スクリーニングから始まる各ステップを経て、選択基準を満たす文献のみを採用する。定性的SRではアウトカムとして各一次研究からなるエビデンステーブルを提示し、上述のとおりバイアスの関与等に関する質的評価、批判的吟味を実施したうえで一定の結論を導き出す。本稿では、これら定性的SRの各ステップにおいて配慮、注意すべき点について、筆者のこれまでの経験も含めて解説したい。

II. 研究テーマの設定と注意点

SRで取り扱う研究テーマは、通常、漠然と思いついた臨床上の疑問や、生命科学の疑問を基盤とし、事前調査による検討を通して決定することが基本である

う。本稿では例として臨床上の問題を扱うSRの進め方について説明するが、まずは漠然と想起したクリニカルクエスチョンを、より具体化するためにPE(I)CO形式に展開、構造化してみる。すなわち次の4項目にあてはめてみる。① Patient/Problem (患者、対象/問題) どんな対象に、② Exposure/Intervention (暴露因子/介入) 何があると、何をすると、③ Comparison (比較) 何と比べて、④ Outcome (結果) どうなるか? である。理解を得やすいように咀嚼機能と認知機能の関係为例としてクリニカルクエスチョンを構造化してみると、P: 成人において、E: 咀嚼機能の低下は、C: 咀嚼機能が健全な場合に比較して、O: 認知機能の低下、認知症の発症頻度を高めるか? などとなる。このように構造化したクリニカルクエスチョンを基盤として、着想したテーマに類似した先行するSRの存在の有無を文献検索データベースから確認する。通常はSRのタイトル、アブストラクトに目を通すことから開始するが、その際、着想した研究テーマのPE(I)COがすべて同じ、もしくは非常に近いSRが既出の場合はテーマの再考も必要となると思われるが、一項目でも異なれば十分貴重なSRとなることもありうる。また、時間的要因もテーマの決定には影響を与えうる。着想したテーマに関するSRが既出であっても、出版からある程度の期間が経過しているとともに、その間に着想したSRの対象となる一次研究がさらに一定数報告されている場合は、最新のデータを基にした新たなSRとしてまとめることで十分価値あるものとなりえる。特に直近の一次研究であれば、より妥当性の高い研究デザインにより実施されたものが増加していることも少なくなく、以前の報告よりも質の高い、有益

なSRとなりうるからである。そして、このような調査を通して着想したテーマに基づくSRの新規性、価値を検討した後は、後述の事前登録サイト上で現在進行中の類似のレビューの存在の有無も確認し、その後の労力が無駄にならないようにすることも重要である。

ここでSRのテーマ設定時の注意点について、筆者の経験を例に挙げて述べておきたい。質の高い、そして外的妥当性の高い結論が得られるSRを実施するために必要な要件として以下のようなものがあげられる。まずは対象となる一次研究の数である。エビデンスの統合のためには一定数の一次研究の存在は必須である。特にアウトカムを主要評価項目として検討し、交絡の影響を可及的に調整した結果を示す一次研究の数が豊富であることが理想である。また、報告元に偏りがないことも重要である。いくら一次研究の数が豊富であっても、それらの多くが同一施設から報告されている状況であれば、それらのデータを統合した結果は決して外的妥当性の高いものであるとは言えないからである。これらの点に関して、筆者の経験を紹介しておきたい。以前、筆者は認知症と口腔顔面痛の関係に関するSRを企画した。文献検索データベースを用いた事前調査により、類似のSRの報告はないことを確認したうえで、文献検索式を立てて実際に一次研究の検索を行ったところ、10数編の一次研究の候補を抽出した。しかし、それらの多くは認知症や認知機能の低下した研究対象者における口腔内の状況として、齦蝕や歯周病を主要評価項目として評価を行ったものであり、口腔顔面痛はあくまで副次的な評価項目であった。また、研究実施施設にも偏りがあり、同じ施設から複数の論文が報告されていたことから、これらの一次研究をまとめて得た結論の外的妥当性も高いとは言えない状況にあると判断し、SRではなくナラティブレビューとして報告した³⁾。このように、着想したレビューのテーマと類似の既報や進行中のものがなくとも、十分な妥当性と信頼性を有するSRにできそうかという点に関しては、事前の論文検索や調査も非常に重要となる。

III. 一次研究の選択基準と文献検索

構造化したクリニカルクエスションと、事前に実施した各調査に基づいてSRのテーマを決定した後は、実際の一次研究の検索方法の検討に進む。まずは一次研究の選択基準の設定が必要となるが、その包含基準、除外基準に含まれる項目には、一次研究が記された論

文言語、研究対象、研究デザイン、観察研究においては暴露因子、介入研究においては介入に関する規定、アウトカムの評価方法など系統的検索のための詳細な基準を含む。その後、これらの選択基準を満たす一次研究を漏れなく収集するステップに進むが、その際に考慮すべき事項としては、まず、1) 多数存在する文献検索のオンライン・データベースの中から、研究内容、学問分野に適したものを複数使用すること、2) 検索が再現できる検索式の詳細な検討と記述、3) ハンドサーチを含めたオフラインツールの活用などがあげられる。そして、検索式の設定と実際の検索は、リサーチライブラリアンなどのエレクトリックサーチを適切に実施できる検索専門家に依頼することが理想ではあるが、各学会が診療ガイドラインを作成するために行うSRとは異なり、個人、研究施設単独で論文作成するためのSRにおいては、費用面からもそのような対応は困難であることも少なくないと思われ、現実的にはレビューチーム内からの複数名で実施することがほとんどであろう。

検索式の設定については、基本的にはPICO、PECOをベースとしたリサーチクエスションから、対象(P)、介入、暴露(I,E)、研究デザインを組み合わせ、適切な検索語(Text term)を吟味したうえで用いて検索する。また、研究対象、論文が発表された期間や記載言語等で絞り込むフィルターをかけることも可能である。注意点として、検索式を厳しく設定することで、適合率は上昇するものの、再現率が低下する、すなわち漏れが大きくなることをあげておきたい。したがって、可能な限り最低限の条件設定(しぼり込み)で広く拾えるように設定することが勧められる。

また、シソーラスが網羅的検索で大きな力を発揮するもことも知られているが、それだけに頼ると取りこぼしも多くなる。このシソーラスとは同義語、類義語などを分類、整理した語彙集であるが、文献データベースのものとしてはPubMedのMedical Subject Heading (MeSH)⁴⁾はよく知られるものの一つである。このMeSHとは米国国立図書館が作成するシソーラスであり、各文献の主題・内容を表す統制されたキーワードで階層構造をもつ。一つの文献には10~20語のMeSH termが付与されている。このMeSHを用いることで、よりノイズの少ない的確な文献を探ることができると言われている。確かに筆者もMeSH検索により抽出される文献の適合率が高いことはこれまでも多く感じてきた。その一方で、抽出されるべき文献が漏れていることを経験したことも少なくない。したがって実際の検索式の設定には、Text

表 1 主な文献検索オンライン・データベースと特徴

データベース名	特徴
MEDLINE	米国国立医学図書館 (NLM) が管理し, PubMed を通して無料で利用可能 MeSH (シソーラス) に紐付いた網羅的な検索が可能 収載雑誌は NLM の基準を満たした米国を中心とした先進国に偏っている傾向もある
EMBASE	Elsevier 社が提供するデータベースで薬学に強い 収載雑誌の違いにより MEDLINE ではヒットしない文献をみつけることができる EMTREE というシソーラスが用いられる
Cochrane Library	コクラン共同計画が管理するデータベース 7つのデータベースから成り, その中の CENTRAL には, ランダム化比較試験, 比較臨床試験を中心とした臨床試験, 学会抄録, 非公開の臨床試験が含まれる
CINAHL	看護学, 関連医療, 生体臨床医学, 健康科学など関連する 17 分野の情報を収録
Web of Science	クラリベイト・アナリティクスにより提供されているデータベース 自然科学, 社会科学, 人文科学の主要雑誌がカバーされる
Scopus	Elsevier 社から提供されている全分野 (科学, 技術, 医学, 社会科学, 人文科学) のデータベース 検索機能対象には特許データベースも組み込まれている
Science Direct	Elsevier 社から提供され, 生化学, 遺伝学, 分子生物学, 化学, 臨床医学, エンジニアリング, 獣医学など 24 の学問分野にわたるコンテンツを含むデータベース
医中誌	日本国内の医学関連分野 (医学, 歯学, 薬学, 看護学, 獣医学分野) の学術誌に掲載された論文, 会議録を収録するデータベース

term も含めた検索式をたて, レビューチーム内で実施する予備的検索で, ヒットする文献の適合率と再現率を確認しながら最終的な検索式を設定するのが現実的であるように思う。

また, 実際の一次研究の検索に用いるオンラインデータベースは, 表 1 に示すように医学系, 生物学系に絞っても数多く存在する。通常はこれらの中から SR が対象とする研究内容や, 対象学問領域などからフィットするデータベースを複数選択するが, レビューチーム内の研究者が所属する施設で使用が可能なもの, すなわち利用可能性の制限を受けたうえで選ばざるを得ないこともある。

IV. レビュープロトコルの事前登録

SR のテーマと一次研究の選択基準, 文献検索に用いるデータベース, 検索式を含むプロトコルが完成するとプロトコル・レジストリへの事前登録が可能となる。プロトコルをレジストリに登録することで, これからレビューを開始することを全世界に通知し, 他の研究者が同じ SR を始めることを防止できる。逆に考えると, SR の構想段階でこれらのレジストリ内のデータベースを確認し, 非常に類似したプロトコルが掲載されていた場合は, その時点でプロトコルの再考, もしくは SR 自体を中止することでそれ以降の無駄な努力を防ぐことができる。また, レビューの事前登録により, 研究および選択的報告バイアスを取り除くことができる⁵⁾。この選択的報告バイアスとは, 研究成果

の都合の良い部分のみを抜粋して報告することなどを指すが, プロトコルの事前登録により SR で取り扱うアウトカムが事前に明確化されることで, 後に出版された SR 内のアウトカムと相違が見られる場合は, それらのバイアスの存在が明らかにできる。

プロトコル・レジストリの中には, 社会的介入に特化した SR を扱う Campbell Collaboration (<https://www.campbellcollaboration.org>) や, 医療介入に特化した SR を扱う Cochrane Collaboration (<https://www.cochrane.org>) などがあるが, PROSPERO (<https://www.crd.york.ac.uk/prosperto/>) はすべての SR を対象としていることから, 登録数は非常に多い。この PROSPERO はイングランドのヨーク大学が提供しているレジストリであるが, 幅広い研究領域, そして研究デザインについても介入, 観察, レビューのレビューなど幅広いタイプの SR を受け入れる。上述のように, このレジストリに登録しておくことで他の研究者は類似のレビューを開始できないが, 登録されているもの一向に進行しないレビューもあることが問題点としてあげられる。また, 審査に時間がかかりすぎることもよく耳にしてきた。この審査に関しては, Covid-19 パンデミックが始まって以降, 現時点 (2022 年 12 月) においても登録の遅延を避けるため, 申請のあったプロトコルは自動的に実施される基本審査のみを経て自動公開されている。このようにして公開されたプロトコルはその旨追記されて公開されているが, 公開された情報が正確であるかは保証しない旨も伝えている。

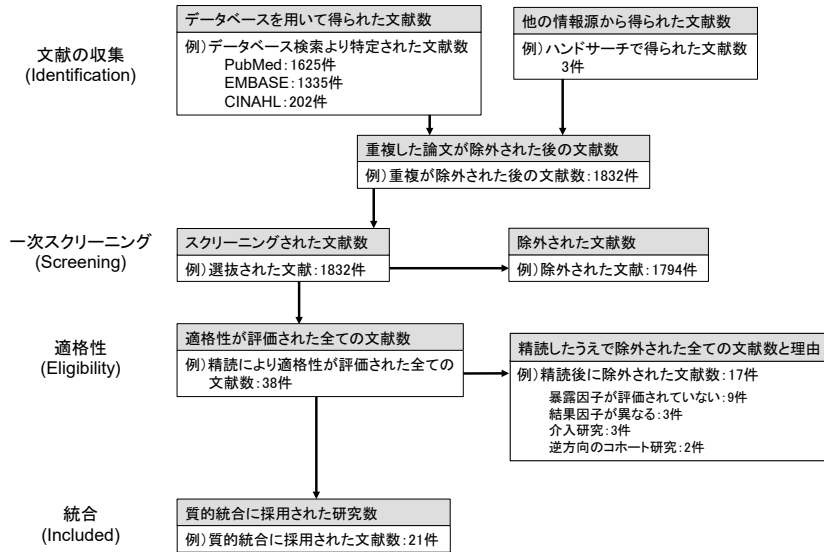


図2 定性的システマティックレビューにおけるPRISMAフローダイアグラムの例

- 研究デザイン(横断研究, コホート研究などに分けて作成)
- 対象サンプル(属性, 特性など)
- 対象者数
- 追跡期間(コホート研究)
- 暴露因子(観察研究の場合: 測定, 評価方法など)
- 介入因子(介入研究の場合: その内容)
- アウトカム(測定, 評価方法など)
- 考慮された交絡因子
- 研究結果
- その他

図3 エビデンステーブルに含まれる内容

最終対象となった一次研究から, このような内容を抽出して表を設計, 作成する。

V. 一次研究の検索と選別

プロトコル・レジストリへの事前登録を終えた後は, 実際に文献検索とそれらの選別を進めていく。手順としては, まず設定した検索式を用い, 採用した各データベースで特定された文献をリストする。各データベースで特定された文献のうち, 重複分を除外したうえで一次スクリーニングに進む。この一次スクリーニングでは, 各文献の論文タイトルとアブストラクトの内容から明らかに対象外となる文献を削除する。そしてこの一次スクリーニングで抽出された文献に対してはSRチーム内の複数名で精読を行い, 採用の適格性を判断する二次スクリーニングへと進む。このプロセスでは, 通常二名の担当者が独立して精読を行い, 採用の可否に関する意見が別れた場合は, 両者の合議も

しくは第三者の判断で決定する。また, 厳密な検索式で実施した系統的検索でヒットしなかった選択基準を満たす論文が, ハンドサーチでピックアップされることもあり, その場合は追加する。このハンドサーチで見いだされる文献は, 特別な手法を用いて抽出されるものではなく, 二次スクリーニングにおける各論文精読中に, それらの論文内で引用されているものの中から見つかるものも多い。

上記のSRへの採用論文抽出のプロセスを図としてまとめたものが, PRISMAフローダイアグラム(フローチャート)と呼ばれるものであり, 通常, SRの結果の章で最初に目にするものである(図2)。複数の担当者による精読で選択基準を満たさなかった一次研究に関しては, それぞれ理由を付して除外し, 最終的に採用された研究数を表示する。

VI. 最終対象となった一次研究からの情報抽出

最終対象となる一次研究が決定した後には, 各研究から情報を抽出したエビデンステーブルを設計, 作成する。通常は研究デザインごと(横断研究, コホート研究など)に分けて作成される。このエビデンステーブル内に抽出, 表示されることが多い項目を図3に示すが, 研究対象に関する情報, コホート研究などであれば対象の追跡期間, 観察研究であれば暴露因子, 介入研究であれば介入因子, そして考慮された交絡因子などの情報を提供したうえで主な研究結果を短くまとめて記載した表となる。もう一つの表としてSR論文内に掲載されるものの中に, 各研究のバイアスリス

表 2 観察研究に用いられる主なバイアスリスク評価ツール

ツール名	発表年	作成者	項目数	特色
Downs and Black instrument ⁶⁾	1998	Downs SH and Black N	29 項目	ケースコントロール研究に不適な項目が含まれる疫学の専門家向き。実施に時間がかかる。
Newcastle-Ottawa scale ⁷⁾	2008	Wells GA et al.	症例対照研究用：7 項目 コホート研究用：8 項目	コクラン共同計画推奨 信頼性に難あり
Risk of Bias Assessment tool for Non-randomized Studies (RoBANS) ⁸⁾	2013	健康保険審査評価院 (韓国)	6 項目	非無作為化研究のバイアスリスク評価 種々のツールを外観 (英国国立医療技術評価機構 (NICE), 米国予防医療専門委員会 (USPSTF) など)
RTI item bank ⁹⁾	2013	医療研究品質機構 (米国)	17 項目	バイアス領域ごとの評価は不可能
CASP checklist ¹⁰⁾	2014	Critical Appraisal Skills Programme (英国)	コホート研究用：12 項目 横断研究用：10 項目	もともとはコホート研究用に開発。サンプルの追跡に関する項目を削除し、横断研究用にも使用が可能

クなどを評価するツールを用いて各一次研究の質的評価を行った結果を示すものがある。バイアスリスクを評価するツールは、介入研究用 (ランダム化, 非ランダム化), 観察研究用ともにさまざまなものが存在して使用されている (表 2)。観察研究の質的評価を行うツールのなかで、有名なものとして Newcastle-Ottawa scale⁷⁾ があげられる。このツールはケースコントロール研究, コホート研究の両デザインの質的評価が可能であり, それぞれのデザインに対応して対象の選択 (4 項目), 比較性 (1 項目), 暴露 (3 項目) に関する 8 項目から構成されている。このツールが開発されて以降, 非常に多くの SR で用いられているが, 最近では信頼性が低く, 偏った結果の研究を同定できないという意見もあり, 改訂の必要も指摘されている。また, 比較的新しいツールとしては, 2013 年に発表されたものとして The Risk of Bias Assessment Tool for Non-randomized Studies (RoBANS)⁸⁾ があるが, これも同様にコホート, ケースコントロール研究に用いることが可能であり, 対象者の選択, 交絡因子, 暴露の測定, アウトカムに対して評価がなされる。さらに, イギリスの健康支援活動団体である Critical Appraisal Skills Program が作成したチェックリスト (CASP check list)¹⁰⁾ は, もともとはコホート研究用に開発されたものであるが, 研究対象者のフォローアップに関する評価項目を削除して, 横断研究用に改変されたものも存在する。これらのツールを用いたバイアスへのリスク評価も, 一次研究の採用決定などと同様に複数の独立した評価者が実施する。この評価にあたっては, 事前に十分なキャリブレーションを行い, 評価者間での不一致が生じた際の解決方法についても決定しておく必要がある。

VII. 結果の統合

ここまでのステップを踏んで採択した一次研究の内容, 結果, 研究の質, バイアスへのリスク評価を考慮に入れて統合し, クリニカルクエスチョンに対する答えを導き出す。この部分は論文内の考察に記載されるが, 定性的 SR の場合は, 定量的なものとは異なり, 採択した結果を総体的に質的にまとめて記載する。その項目立ては, SR で取り上げた主題に沿って, ある程度の自由度をもって良いと考えるが, 少なくとも PRISMA チェックリストの内容は含むべきと考える。また, 先行研究結果と一致するかといった議論や, 得られた結果に関連するメカニズムの考察, 研究の限界や将来展望などが述べられるべきである。

VIII. まとめ

本稿では, 定性的 SR の進め方に関する作法と実際の手順について説明させていただいた。紙面の都合から, 多数種存在するバイアスに関する記載やその扱いなどは割愛させていただいたが, 実際に論文としてまとめる際にはそれらに関する考察も大変重要となることを追記させていただく。本稿が, これから SR を企画される方のお役に立てれば幸いである。

文 献

- 1) 唐 文涛, 小島原典子, 河合富士美, 津谷喜一郎. 診療ガイドラインとシステムティックレビュー. 薬理と治療 2014; 42: 189-97.
- 2) Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic

- reviews and meta-analyses: the PRISMA statement. *Ann Intern Med* 2009; 151(4): 264-9, W64.
- 3) 前川賢治, 窪木拓男. 認知機能低下, 認知症と口腔顔面痛—文献レビュー. *日口腔顔面痛会誌* 2021; 13: 1-10.
 - 4) 坪内政義. Medical subject Headings: MeSH. *医図書館* 2001; 48: 64-5.
 - 5) Williams RJ, Tse T, Harlan WR, Zarin DA. Registration of observational studies: is it time? *CMAJ* 2010; 182: 1638-42.
 - 6) Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomized and non-randomized studies of health care interventions. *J Epidemiol Community Health* 1998; 52: 377-84.
 - 7) Wells GA, Shea B, O'Connell D, Peterson J, Welch V, Losos M, Tugwell P. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomized studies in meta-analyses 2021. https://www.ohri.ca/programs/clinical_epidemiology/oxford.asp : Accessed on December 12th, 2022.
 - 8) Kim SY, Park JE, Lee YJ, Seo HJ, Sheen SS, Hahn S et al. Testing a tool for assessing the risk of bias for nonrandomized studies showed moderate reliability and promising validity. *J Clin Epidemiol* 2013; 66(4): 408-14.
 - 9) Viswanathan M, Berkman ND. Development of the RTI item bank on risk of bias and precision of observational studies. *J Clin Epidemiol* 2012; 65(2): 163-78.
 - 10) Critical Appraisal Skills Programme. CASP checklists. <https://casp-uk.net/casp-tools-checklists/>: Accessed on December 12th, 2022.
-

著者連絡先：前川 賢治

〒540-0008 大阪府大阪市中央区大手前1丁目5-17 大阪歯科大学欠損歯列補綴咬合学講座
Tel: 06-6910-1043
Fax: 06-6910-1044
E-mail: maekawa-k@cc.osaka-dent.ac.jp