

「AMSTAR 2：ヘルスケア介入のランダム化または非ランダム化研究あるいは両方を含むシステマティック・レビューのための厳密な評価ツール」の紹介と解説



上岡 洋晴 (東京農業大学大学院環境共生学専攻)
折笠 秀樹 (富山大学大学院医学薬学研究部)
津谷 喜一郎 (東京有明医療大学保健医療学部)

はじめに

昨 2017 年 9 月に、Shea BJ らによる “AMSTAR 2: a critical appraisal tool for systematic reviews that include randomized or non-randomised studies of healthcare interventions, or both” が *BMJ* に open access として掲載された¹⁾。2007 年 2 月にほぼ同じ著者らによる AMSTAR (A Measurement Tool to Access systematic Review) が *BMC Medical Research Methodology* に掲載²⁾されて 10 年後である。掲載誌はより格の高い *BMJ* になっている。わたしたちは 2007 年の AMSTAR 全文の日本語訳を時間をかけて作成し『臨床研究と疫学研究のための国際ルール集 Part 2』(ライフサイエンス出版, 2016) に収載した³⁾。また、これを「水中運動・温泉療法」⁴⁾、「音楽療法」⁵⁾、「ピラティス運動」⁶⁾、「機能性表示食品」⁷⁾などの領域のシステマティック・レビュー (systematic review: SR) において用いてきた。

AMSTAR 2 (2017) は、AMSTAR (2007) より分量が多くやや読みにくい箇所もある。また、読者はすでに AMSTAR 2 (2017) のサブタイトルが AMSTAR (2007) とは異なっていることに気づかれたであろう。そこで本稿では、AMSTAR 2 (2017) の根幹となる 16 のチェック項目の解説と表を主として後方に配置し、1. AMSTAR (2007) 開発の背景、2. その内容と課題、3. AMSTAR 2 (2017) の開

発方法、4. AMSTAR 2 チェックリスト、5. AMSTAR 2 の特徴—の構成とした。

1. AMSTAR (2007) 開発の背景

「エビデンスに基づいた医療」(evidence-based medicine: EBM) において、種々の介入の最適利用のために、医師や診療ガイドライン (clinical practice guidelines: CPGs) 作成者は、理想的には、適切な方法かつすべての関連する主要な研究を包んで実施された SR に基づいて意思決定すべきである⁸⁾。意思決定に SR を用いようとする興味は大量の SR を作成させてきた。しかし、それらはしばしば冗長で、不必要で、誤認をもたらし、矛盾が生じていることを認識する必要がある。それによって正しい方向への軌道修正がなされるべきである⁹⁾。研究の全体像を把握し、CPGs を作成する研究者は、実施された SR のデザインや実施方法における潜在的な限界について注意深く捉えるべきであり、取り扱う SR の方法論の質について評価し、報告すべきである¹⁰⁾。

適切な SR の計画・実施・報告という一連の方法のチェックリストの「報告ガイドライン」(reporting guidelines) として、2009 年に PRISMA¹¹⁾ が公表された。SR により 2 次的にエビデンスを「つくる」局面における、SR の報告方法の 27 項目からなるチェックリストであり、論文として投稿する際の必須項目を包括している。2015 年には、SR における

ネットワークメタ分析の方法論に焦点を絞った PRISMA-NMA¹²⁾, SR を実施する際のプロトコールを作成するための PRISMA-P¹³⁾ が相次いで公表された。

研究のプロトコールから論文まで, これらのチェックリストに基づいて漏れなく記載されていればよい。しかし, さまざまな分野で膨大な SR が生産されるなかで, そうではないものも相当存在し, 前述のように SR の質が疑問視されている。

こうした背景もありエビデンスを「つかう」局面での「SR の質を評価するためのチェックリスト」が開発された。2007 年の AMSTAR, 2016 年の ROBIS¹⁴⁾, 同年のコクランによる “Methodological Expectations of Cochrane Intervention”¹⁵⁾ である。そのなかでも, AMSTAR は介入効果に関連する SR を包括的に質評価するツールとして広く用いられてきた。

2. AMSTAR (2007) の内容と課題

AMSTAR は, 研究選択やデータ抽出, データの統合など, SR の各ステージにおける方法の適正性をカバーする 11 項目から構成され, “Yes” (はい), “No” (いいえ), “can’t answer” (回答不能), “not applicable” (非該当) で評価を行うものであった (表 1)^{2,3)}。このうち「非該当」とは, たとえば「メタ分析が適正な方法でなされたか?」という設問に際して, メタ分析を含まない定性的な SR の場合に選択する。AMSTAR は対象となる SR を 11 点満点で総合評価することに特徴がある。その他の優位性として, SR に対する質評価が 10-20 分という短時間で行える簡便性がある¹⁶⁾。さらに本 2018 年には, AMSTAR と ROBIS の測定者間信頼性 (評価者間の一致度) はともに高値で, ほぼ同等であるとする報告がなされている¹⁷⁾。

しかし, 総合得点という意味で意義がある一方で, 逆に個々の項目についての重要な強さと弱さが明瞭にならないという限界も指摘された^{16,18)}。また, 一部の項目がつねに同じ方法で解釈されるとは

限らないことや, 方法論的な質ではなく報告の質を評価する項目も含まれているという批判もなされた^{19,20)}。こうした批判やさらなる発展性を考慮して, AMSTAR の改訂版となる AMSTAR 2 が 2017 年に開発されたことになる。

3. AMSTAR 2 (2017) の開発方法

AMSTAR 2 の開発に際して, もととの AMSTAR 開発者に加え, 非ランダム化ヘルスケア介入研究 (Non-Randomised Studies of healthcare Interventions: NRSI), 評価ツールの開発, バイオ統計, 研究デザインの専門家からなるエキスパート・グループが設立された。カナダのオタワにおいて, より価値のあるアップデートを図るための議論が 1 日なされた。その議論では, AMSTAR 利用者による調査結果, 2015 年と 2016 年のコクラン会議 (The Cochrane Colloquium) における AMSTAR ワークショップの参加記録, AMSTAR ウェブサイトからのフィードバック, 公表された批評が共有された。

エキスパート・グループは, AMSTAR の改訂にあたって SR の実施のすべての側面に焦点をあてることに加え, 非ランダム化研究も含めるべきだと考えた。また, 改訂版は, レビューを実施する者にとって教育用ツールとして, かつ詳細なチェックリストとしても機能すべきだとも考えた。

ノミナル・グループ (Nominal Group) 法¹⁹⁾により, グループメンバーは独立して自身のアイデアを出すことが求められ, それらは名前を伏せられ誰が記載したかわからないように順番が入れ替えられた形 (round-robin format) で示された。メンバーはブラインド下で審査し, 順位づけを行った。それらは統計学的に取りまとめられ, 以下の 7 点の変更が同意された (ただし, すべてが重要であるため優先順位はない)。

- ・回答カテゴリーを単純化すること。
- ・リサーチ・クエスションの定義を PICO の枠組み (参加者, 介入, コントロール群, アウトカム) に

訳注:

一般的に, 自由に意見を出し合うブレイン・ストーミング (ここではチェック項目) と, 複数の会議を設け無記名で投票し統計結果を議論するというステップを繰り返してコンセンサスを形成するデルフィー・コンセンサスを組み合わせ, 科学的な合意形成法のこと。ガイドラインの作成やチェックリストの開発などに用いられる。

表 1 AMSTAR：システマティック・レビューの方法論的な質を評価するために作成された測定ツール (2007)³⁾

<p>1. デザインは“あらかじめ (a priori)” 提示されたか？ 研究の疑問および組入れ基準 (inclusion criteria) はレビューの実施前に規定する。</p>	<p><input type="checkbox"/> はい <input type="checkbox"/> いいえ <input type="checkbox"/> 回答不能 <input type="checkbox"/> 非該当</p>
<p>2. 研究の選択とデータの抽出は複数で行われたか？ 少なくとも 2 人のデータ抽出者 (data extractors) が存在し、不一致があった場合は合意された方法により取り扱う。</p>	<p><input type="checkbox"/> はい <input type="checkbox"/> いいえ <input type="checkbox"/> 回答不能 <input type="checkbox"/> 非該当</p>
<p>3. 包括的な文献検索がなされたか？ 少なくとも 2 つの電子的なデータ源を検索する。報告では検索に含めた年と用いたデータベース (たとえば Central, EMBASE, MEDLINE) を記載する。キーワードおよび/または MeSH タームを明示しなければならず、可能なら検索式を提示する。検索は、最新の研究、レビュー、教科書、特殊なデータベース (specialized register)、特定の研究分野における専門家の意見を参考にし、見つかった研究における引用文献をさらに検討して補完する。</p>	<p><input type="checkbox"/> はい <input type="checkbox"/> いいえ <input type="checkbox"/> 回答不能 <input type="checkbox"/> 非該当</p>
<p>4. 出版形態 (status of publication) (灰色文献 grey literature など) は組入れ基準に用いられたか？ 著者らは、出版のタイプ (publication type) に関係なく報告を検索したことを明示する。また、出版形態、言語等に基づいて何らかの報告を (システマティック・レビューから) 除外したか否かについても明示する。</p>	<p><input type="checkbox"/> はい <input type="checkbox"/> いいえ <input type="checkbox"/> 回答不能 <input type="checkbox"/> 非該当</p>
<p>5. (組入れられたおよび除外された) 研究のリストは提示されたか？ 組入れられたおよび除外された研究のリストを提示する。</p>	<p><input type="checkbox"/> はい <input type="checkbox"/> いいえ <input type="checkbox"/> 回答不能 <input type="checkbox"/> 非該当</p>
<p>6. 組入れられた研究の特性 (characteristics) は提示されたか？ 表のような集約された形式で、参加者、介入、アウトカムに関する元の研究からのデータを提示する。解析されたすべての研究における特性の範囲 (ranges of characteristics)、たとえば年齢、人種、性、関連する社会経済的なデータ、疾患の状態、罹患期間、重症度、その他の疾患などを報告する。</p>	<p><input type="checkbox"/> はい <input type="checkbox"/> いいえ <input type="checkbox"/> 回答不能 <input type="checkbox"/> 非該当</p>
<p>7. 組入れられた研究の科学的な質は評価され、記録されていたか？ “あらかじめ決められた” 評価の方法を提示する (たとえば有効性試験 effectiveness study に対して、ランダム化二重盲検プラセボ対照試験のみを組入れること、あるいは組入れ基準として割付けの隠蔽 allocation concealment を採用した場合など); 他のタイプの研究に対しては、代わりの項目が関連するであろう。</p>	<p><input type="checkbox"/> はい <input type="checkbox"/> いいえ <input type="checkbox"/> 回答不能 <input type="checkbox"/> 非該当</p>
<p>8. 組入れられた研究の科学的な質は、結論を導く際に適正に利用されたか？ 方法論的な厳密さや科学的な質に関する結果は、解析とレビューの結論において考慮し、推奨を導く際に明確に示す。</p>	<p><input type="checkbox"/> はい <input type="checkbox"/> いいえ <input type="checkbox"/> 回答不能 <input type="checkbox"/> 非該当</p>
<p>9. 研究結果を併合する (combine) のに用いられた方法は適正だったか？ 統合された結果 (pooled results) に対し、研究が併合可能であったことを保証し、その均質性 (homogeneity) を評価するための検定を行う (均質性に対するカイ 2 乗検定, I^2)。異質性 (heterogeneity) が存在する場合、変量効果モデル (random effects model) を用いるか、および/または併合することの臨床的妥当性 (併合が理にかなっているか) を考慮する。</p>	<p><input type="checkbox"/> はい <input type="checkbox"/> いいえ <input type="checkbox"/> 回答不能 <input type="checkbox"/> 非該当</p>
<p>10. 出版バイアスの可能性は評価されたか？ 出版バイアスの評価には、グラフによる補助 (たとえば漏斗状プロット funnel plot, その他の利用可能な検定)、および/または統計学的な検定 (たとえば Egger 回帰検定) の組み合わせを用いる。</p>	<p><input type="checkbox"/> はい <input type="checkbox"/> いいえ <input type="checkbox"/> 回答不能 <input type="checkbox"/> 非該当</p>
<p>11. 利益相反は明示されたか？ 可能性のある資金源 (potential sources of support) は、システマティック・レビューと、レビューされた研究どちらにおいても明示する。</p>	<p><input type="checkbox"/> はい <input type="checkbox"/> いいえ <input type="checkbox"/> 回答不能 <input type="checkbox"/> 非該当</p>

表 2 AMSTAR 2:ヘルスケア介入のランダム化または非ランダム化研究あるいは両方を含むシステマティック・レビューのための厳密な評価ツール¹⁾

項目 1 レビューの際、リサーチ・クエスチョンと組入れ基準に PICO の要素を含めたか？		
Yes の場合： <input type="checkbox"/> 参加者 <input type="checkbox"/> 介入 <input type="checkbox"/> コントロール群 <input type="checkbox"/> アウトカム	選択項目（推奨） <input type="checkbox"/> フォローアップ期間	<input type="checkbox"/> Yes <input type="checkbox"/> No
項目 2 レビューの報告に、レビュー実施前にその方法を決定したことを明示し、重大な計画変更があれば正当な理由を示したか？		
部分的な Yes の場合： 著者は以下のすべてを含むプロトコルや基準があることを明示する： <input type="checkbox"/> レビュー・クエスチョン <input type="checkbox"/> 検索戦略 <input type="checkbox"/> 組入れ/除外基準 <input type="checkbox"/> バイアスリスク評価	Yes の場合： 部分的な Yes の場合に加え、プロトコルが事前に登録され、また特定されるべき： <input type="checkbox"/> 必要に応じて、メタ分析/併合の計画、および <input type="checkbox"/> 異質性の原因を探る計画 <input type="checkbox"/> プロトコルからの何らかの逸脱に対する妥当性	<input type="checkbox"/> Yes <input type="checkbox"/> Partial Yes <input type="checkbox"/> No
項目 3 レビュー著者は、レビューに組入れる際の研究デザインの選択基準について説明したか？		
Yes の場合、そのレビューは以下の 1 つを満たすべきである： <input type="checkbox"/> RCT のみを組入れたことの説明 <input type="checkbox"/> または、NRSI*のみを組入れたことの説明 <input type="checkbox"/> または、RCT と NRSI の両方を組入れたことの説明		
*NRSI：非ランダム化ヘルスケア介入研究（Non-Randomised Studies of healthcare Interventions）		
項目 4 レビュー著者は、網羅的な文献検索の手法を用いたか？		
部分的な Yes の場合（以下のすべて）： <input type="checkbox"/> 少なくとも 2 つのデータベースで検索した（リサーチ・クエスチョンに関して） <input type="checkbox"/> キーワードと検索戦略を示した <input type="checkbox"/> 出版形態の制限の正当性を示した（例：言語）	Yes の場合、次も必要（以下のすべて）： <input type="checkbox"/> 検索した参考文献リスト/組入れた研究の文献一覧 <input type="checkbox"/> 検索した試験/研究登録データベース <input type="checkbox"/> 検索で組入れた/相談したその分野の専門家 <input type="checkbox"/> 関連する場合は、検索した灰色文献 <input type="checkbox"/> レビュー完成の 24 ヶ月以内に検索を実施した	<input type="checkbox"/> Yes <input type="checkbox"/> Partial Yes <input type="checkbox"/> No
項目 5 レビュー著者は、複数人で研究選択を行ったか？		
Yes の場合、以下のいずれか 1 つを含む： <input type="checkbox"/> 少なくとも 2 人のレビュアーが、基準を満たす研究の選択について独立して合意し、組入れる研究について合意形成した <input type="checkbox"/> または、2 人のレビュアーが基準を満たす研究のサンプルを選択し、良好な一致度（少なくとも 80%）に至り、残りは 1 人のレビュアーによって選択された		
項目 6 レビュー著者は、複数人でデータ抽出を行ったか？		
Yes の場合、以下のいずれか 1 つを含む： <input type="checkbox"/> 少なくとも 2 人のレビュアーが、組入れられた研究から抽出するデータについて合意形成した <input type="checkbox"/> または、2 人のレビュアーが基準を満たす研究のサンプルからデータを抽出し、良好な一致度（少なくとも 80%）に至り、残りは 1 人のレビュアーによって抽出された		

表 2 つづき

項目 7 レビュー著者は、除外した研究のリストを提示し、その除外が正当である理由を示したか？

部分的な Yes の場合： <input type="checkbox"/> 除外されたがフルペーパーとして十分に読み込まれたすべての潜在的に関連する研究のリストを提供した	Yes の場合，次も含まなければならない： <input type="checkbox"/> 潜在的に関連する研究についてレビューから除外したことの正当性を示した	<input type="checkbox"/> Yes <input type="checkbox"/> Partial Yes <input type="checkbox"/> No
---	--	---

項目 8 レビュー著者は、組入れた研究を十分詳細に記述したか？

部分的な Yes の場合（以下のすべて）： <input type="checkbox"/> 参加者の記載 <input type="checkbox"/> 介入の記載 <input type="checkbox"/> 比較対照の記載 <input type="checkbox"/> アウトカムの記載 <input type="checkbox"/> 研究デザインの記載	Yes の場合，次も必要（以下のすべて）： <input type="checkbox"/> 詳細な参加者の記載 <input type="checkbox"/> 詳細な介入の記載（関連する投与量も含む） <input type="checkbox"/> 詳細な比較対照の記載（関連する投与量も含む） <input type="checkbox"/> 研究の場所の記載 <input type="checkbox"/> フォローアップ期間の記載	<input type="checkbox"/> Yes <input type="checkbox"/> Partial Yes <input type="checkbox"/> No
---	---	---

項目 9 レビュー著者は、レビューに組入れた各研究のバイアスリスク評価に際し、満足のいく手法を用いたか？

RCT		
部分的な Yes の場合，以下の状態のバイアスリスクが評価されていなければならない： <input type="checkbox"/> 隠蔽されていない割振り，および <input type="checkbox"/> アウトカムを評価する際の，参加者と評価者のブラインディングの欠如(すべての死亡原因のような客観的なアウトカムの場合には不要)	Yes の場合，以下の状態のバイアスリスクも評価されていなければならない： <input type="checkbox"/> 本当にランダムではない割振り手順，および <input type="checkbox"/> ある特定のアウトカムについて多重測定や多重解析から報告された結果の選択	<input type="checkbox"/> Yes <input type="checkbox"/> Partial Yes <input type="checkbox"/> No <input type="checkbox"/> Includes only NRSI
NRSI		
部分的な Yes の場合，以下の状態のバイアスリスクが評価されていなければならない： <input type="checkbox"/> 交絡因子，および <input type="checkbox"/> 選択バイアス	Yes の場合，以下のバイアスリスクも評価されていなければならない： <input type="checkbox"/> 曝露とアウトカムを確認するために使用された方法，および <input type="checkbox"/> ある特定のアウトカムについて多重測定や多重解析から報告された結果の選択	<input type="checkbox"/> Yes <input type="checkbox"/> Partial Yes <input type="checkbox"/> No <input type="checkbox"/> Includes only RCT

項目 10 レビュー著者は、レビューに組入れた研究に対する資金源を報告したか？

Yes の場合： <input type="checkbox"/> レビューに組入れられた個々の研究の資金源が報告されなければならない（注：レビュアーがこの情報を探したが，著者によって報告されていない場合は，そのように報告する）	<input type="checkbox"/> Yes <input type="checkbox"/> No
---	---

項目 11 メタ分析が行われた場合、レビュー著者は、結果の統計的併合において適正な方法を用いたか？

RCT		
Yes の場合： <input type="checkbox"/> 著者はメタ分析でデータ併合を正しく行った <input type="checkbox"/> そして，研究結果の併合に適した重みづけ法を用い，存在する場合には異質性を調整した <input type="checkbox"/> そして，異質性の原因について調べた		<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> No meta-analysis conducted

表 2 つづき

NRSI

Yes の場合： Yes

著者はメタ分析でデータ併合を正しく行った No

そして、研究結果の併合に適した重みづけ法を用い、存在する場合には異質性を調整した No meta-analysis conducted

そして、生データを併合するのではなく、交絡因子を調整した NRSI からの効果量推定値を統計的に併合した、または調整された効果量推定値が利用できなかった場合に生データを併合することの正当性を示した

そして、RCT と NRSI の両方をレビューに組入れた場合、それぞれに分けて効果量を報告した

項目 12 メタ分析が行われた場合、レビュー著者は、各研究のバイアスリスクがメタ分析や他の統合結果に与える影響について評価したか？

Yes の場合： Yes

低バイアスリスクの RCT のみを用いた No

または、プールされた効果量がさまざまなバイアスリスクでの RCT または NRSI に基づく場合、要約効果量におけるバイアスリスクの影響を調べる分析を行った No meta-analysis conducted

項目 13 レビュー著者は、レビュー結果の解釈や考察をする際にバイアスリスクを考慮したか？

Yes の場合： Yes

低バイアスリスクの RCT のみを用いた No

または、中から高バイアスリスクの RCT や NRSI をレビューに組入れた場合、結果におけるバイアスリスクの影響を考察した

項目 14 レビュー著者は、レビュー結果にみられるいかなる異質性に対しても、満足のいく説明や考察を行ったか？

Yes の場合： Yes

結果において有意な異質性はなかった No

または、異質性があった場合、結果における異質性の原因を調べ、レビュー結果に与える異質性の影響を考察した

項目 15 量的な統合がなされた場合、レビュー著者は出版バイアス（小規模研究バイアス）について十分な検討、およびレビュー結果に与える影響の可能性を考察したか？

Yes の場合： Yes

出版バイアスに対するグラフィック的または統計的検定を行い、出版バイアスの影響の程度を考察した No

No meta-analysis conducted

項目 16 レビュー著者は、レビュー実施に際して受領した財源を含め、考えられる利益相反について報告したか？

Yes の場合： Yes

レビュー著者は利益相反がないことを報告した、または No

レビュー著者は資金源を記述し、潜在的な利益相反にどのように対処したかを記載した

適合させること。

- ・SRの組入れに関し、異なる研究デザイン(ランダム化や非ランダム化)を選択することの妥当性をレビュー著者に求めること。
- ・レビューからの研究の除外についてより詳細な理由を求めること。
- ・レビュー著者は組入れた研究(ランダム化や非ランダム化かどうか)に関するバイアスリスク(bias risk)を十分に詳細に評価したかどうかを明らかにすること。
- ・組入れられた研究のバイアスリスクは、結果の統計的な統合時に適切に考慮されたかどうかを明らかにすること。
- ・組入れられた研究のバイアスリスクは、レビューの結果の解釈、考察時に適切に考慮されたかどうかを明らかにすること。

各草案項目に対して説明が作成され、また表現方法が洗練化された。

AMSATR 2 (2017) は、もともとの10の領域を保持し、AMSTAR (2007) の11項目から全体で16項目となった。そして、AMSTAR (2007) よりも回答カテゴリーが単純化され、より包括的な利用者ガイドを含み、重要な部分における弱点に基づく全体的な評価を有している。

4. AMSTAR 2 (2017) のチェック項目

チェック項目を正しく理解するために、“AMSTAR 2 guidance document” (<https://amstar.ca/docs/AMSTAR%202-Guidance-document.pdf>) があるので、これを参考にするとよい。表2にすべての項目を示す。またBox 1に示す7つの重要項目は項目数にアンダーラインを付した。

項目1 レビューの際、リサーチ・クエスションと組入れ基準にPICOの要素を含めたか？

PICO(参加者、介入、コントロール、アウトカム)はリサーチ・クエスションに対する共通した項目であり、Yesと評価する場合は必ず4つが含まれていなければならない。必要に応じて、フォローアップする期間も追加としての選択項目になる場合がある。たとえば介入効果が数年後に期待されるよ

うな場合である。

項目2 レビューの報告に、レビュー実施前にその方法を決定したことを明示し、重大な計画変更があれば正当な理由を示したか？

SRは観察的な研究形式であり、リサーチ・クエスションやレビューの方法は開始前に決定すべきである。著者はそれを証明する必要がある、PROSPEROやオープン・ジャーナル(たとえば*BMJ Open*)、あるいは研究室や倫理委員会への日付入りの文書提出が用いられる。これは、最低限その報告のなかに記載すべきである。Yesと評価する場合は、実際に事前に書かれたプロトコル(登録された)との照合をしなければならず、もしプロトコルからの逸脱がある場合には、著者がそれについて正当性を含めて述べているかを評価すべきである。明らかに説明されていない場合には、評価のグレードを下げることになる。

項目3 レビュー著者は、レビューに組入れる際の研究デザインの選択基準について説明したか？

一般的なルールでは、SRはRCTに限定されるが、何らかの理由でRCTだけでは不十分な場合もある。たとえば、RCTが存在しない、利用可能なRCTにおいて必要なアウトカムがない(害のRCTなど)、検出力が不十分、参加者が制限されている、代表的ではないコントロールや介入の治療方法の場合など、さまざまな理由が考えられる。一方、NRSIだけにすることは、その理由が必要である。両方を用いる場合でも理由が必要である。両方を用いることは、ある介入の有効性や害の完全な形を得るのに妥当かもしれない。後述するが、その場合には、2つの研究デザインについて、それぞれ独立して内容を評価・メタ分析を実施することが推奨される。

項目4 レビュー著者は、網羅的な文献検索の手法を用いたか？

少なくとも2つ以上のデータベースが必要であり、調べた年と名称(たとえば、CENTRAL、EMBASE、MEDLINE)を記載する。キーワードやMeSH項目が報告され、十分な検索戦略(search

strategy) になっているべきである。検索には、補助的に公表されているレビュー、特定の登録、特定の研究分野の専門家、そして発見された研究からの参考文献リストを見ることも含まれる。ときには、結果を明らかにしたり、更新や訂正箇所を特定化するためにもとの研究者に連絡することも必要とされる。

すべての関連する言語での公表物が対象になるべきだが、言語における制限がある場合にはその正当性を述べる。灰色文献 (grey literature) は、ときには政策やプログラム評価の報告を有していて重要であり、ウェブサイト (たとえば、政府、非政府または健康・科学技術の機関) からのみ入手できる。著者は試験登録、学会抄録、学位論文、個人的なウェブサイト (たとえば、大学、リサーチ・ゲート) に掲載されている未公表報告のような適切な情報源を検索すべきである。加えて、薬剤介入試験はピアレビュー雑誌に公表されないかもしれないが、主宰企業や研究者から直接入手しうる。Yes の場合には、評価者はすべての関連する検索の手続きがレビュー著者によって実施されていると捉えるべきである。

項目 5 レビュー著者は、複数人で研究選択を行ったか？

最良の実施には、SR に組入れるための研究の適正性を決定する 2 人のレビュー著者を必要とする。回答の選択肢において、組入れる研究のサンプルの選択で 2 人の測定者間の一致度をみる方法が示されている。組入れる研究の一致度が 0.8 以上 (κ 係数) の強さを示すことが推奨される。

項目 6 レビュー著者は、複数人でデータ抽出を行ったか？

項目 5 と同じように、2 人の独立したレビュー著者によってデータ抽出が行われるべきである。

項目 7 レビュー著者は、除外した研究のリストを提示し、その除外が正当である理由を示したか？

関連する研究の正当性に欠く除外は、レビュー結果にバイアスをもたらすかもしれない。レビュー著者には、除外した文献の完全なリストを提供することが求められる。

項目 8 レビュー著者は、組入れた研究を十分詳細に記述したか？

PICO や研究デザイン、分析、場所の情報は研究結果を応用する際に重要な情報となり、どの研究が適しているか、自身の治療や政策に関連しているかなどを判断する材料になる。記述語も介入効果の異質性を研究するためのフレームワークを提供する (たとえば、投与量、年齢幅、臨床背景)。

項目 9 レビュー著者は、レビューに組入れた各研究のバイアスリスク評価に際し、満足のいく手法を用いたか？

いずれの SR においても、評価の必須項目である。鍵となる点は、結果の要約や解釈の際にバイアスリスクを考慮したかどうかである。RCT 限定のレビューの際は、コクラン・ハンドブック (バイアスリスク評価) を参考にすることが推奨される。もし、レビュー著者がスタンダードなバイアスリスク評価ツールを用いていないなら、深刻な方法論的な欠陥を発見しうることを明らかにしなければならない。“Newcastle Ottawa Scale”, “SIGN”, “Mixed Methods Appraisal Tool (MMAT)” なども一般的である。

AMSTAR 2 の開発においては、RCT のためのコクランの Risk of Bias が引用されている。NRSI では、コクランの “Risk Of Bias In Non-randomized Studies of Interventions (ROBINS-I tool)” (以前は A Cochrane Risk Of Bias Assessment Tool: for Non-Randomized Studies of Interventions (ACROBAT-NRSI) と称された) が推奨される。

項目 10 レビュー著者は、レビューに組入れた研究に対する資金源を報告したか？

商業ベースの主宰者 (sponsor) の研究結果は、そうでない独立した資金で実施された研究結果よりも肯定的な傾向になるとする報告がなされている。この情報を入れることは価値があり、商業的に資金提供 (fund) されたものと、そうでない研究結果を分けて分析できるかもしれない。

項目 11 メタ分析が行われた場合、レビュー著者は、結果の統計的併合において適正な方法を用いたか？

レビュー著者は、プロトコールに基づいてメタ分析を実施したかを明確に示すべきである。RCTの場合には、固定効果モデル (fixed effect model) を用いたのか変量効果モデル (random effect model) を用いたのか説明すべきである。NRSI の場合には、何百、何千という多くの参加者がいる場合があり、もし参加者の少ない RCT と一緒に分析されると、NRSI のデータに誘導されてしまう可能性がある。したがって、別々に分析をする必要がある。NRSI が RCT と同じくらい低バイアスであることは稀であり、NRSI の効果量 (effect size) と信頼区間 (confidence interval: CI) にはとくに注意を払うべきである。

異質性 (heterogeneity) はメタ分析において重要な問題であり、とくに NSRI ではいっそう重要である。一般的な異質性の原因 (異なるコントロール (control), アウトカムのベースラインリスクの分散や参加者の他の特性, 異なる介入 (たとえば, 量反応, 状況/場所, 臨床経験), 異なるアウトカムの定義) に加えて, 参加者の出身地, データの完全性, データ管理や解析の方法における異質性を考慮することが重要である。交絡因子 (confounding factor) を調整した結果は, 生データを調整しないで得た結果とまったく異なるかもしれない。一般的に, NRSI の結果を併合する場合には, 交絡因子等の十分な調整をしなければならない。

項目 12 メタ分析が行われた場合、レビュー著者は、各研究のバイアスリスクがメタ分析や他の統合結果に与える影響について評価したか？

レビュー著者が、質の高い RCT だけを選択する場合には、結果における潜在的なバイアスの影響の議論はほとんどないかもしれないが、多様な質の RCT を含む場合には、回帰分析や、低バイアスリスクの研究のみで効果サイズを評価することによって、この影響を評価すべきである。NRSI においては、低から中のバイアスリスクあるいは低バイアスリスク (もしあれば) だけで効果サイズを評価すべ

きである。メタ分析を行わない場合には、個々の研究結果におけるバイアスリスクの影響を記述すべきである。

項目 13 レビュー著者は、レビュー結果の解釈や考察をする際にバイアスリスクを考慮したか？

メタ分析が実施されなかった場合でも、レビュー著者は結果の解釈にバイアスリスクの影響についての考察を含めるべきである。とりわけ、RCT であっても多様なバイアスリスクを含む場合や、NRSI を組み入れている場合には重要である。

項目 14 レビュー著者は、レビュー結果にみられるいかなる異質性に対しても、満足のいく説明や考察を行ったか？

RCT よりも NRSI の結果において、異質性の潜在的原因は多い。PICO 要素と項目 9 で示したバイアスの領域のどちらも、異質性の重要な潜在的な原因と考えるべきである。レビュー著者はこれらの可能性を探り、結論や何らかの推奨をするうえで異質性の影響を考察しなければならない。

項目 15 量的な統合がなされた場合、レビュー著者は出版バイアス (小規模研究バイアス) について十分な検討、およびレビュー結果に与える影響の可能性を考察したか？

この項目はとても重要な問題だが、レビュー著者や評価者が完全に解決するのは難しいかもしれない。鍵は、レビュー著者が出版バイアスを明らかにするために、深く集中的な検索を行う努力をしたか、結果に関する解釈や考察でそれについての影響を示したか、そして感度分析を実施したかである。

項目 16 レビュー著者は、レビュー実施に際して受領した財源を含め、考えられる利益相反 (conflict of interest: COI) について報告したか？

既得権益によって資金提供された個々の研究は、独立した研究よりも介入の方向に有利な結果を生み出す可能性があることは前述したが、これは SR にも該当する。著者は直接的な資金源を含め、関連する潜在的な利益相反事項を報告しなければならない。

5. AMSTAR 2 (2017) の特徴

AMSTAR 2 は、全体のスコア (得点) を記述するためにデザインされていないことが強調されている。高い得点の SR でも、不十分な文献検索や、SR で組入れられた個々の研究のバイアスリスクの評価を失敗していることなど、特定の領域における重大な弱点が隠されているかもしれないからである。SR の総合評価に際しては、重要な領域の欠落を考慮することが重要であり、そうした部分の弱点によって SR の信頼性が大きく失われる可能性があることを認識しなければならない。

また領域ごとの項目は、“Yes” (はい) という肯定的な結果で回答するように枠付け (framing) がなされた。AMSTAR (2007) では “can’t answer” (回答不能), “not applicable” (非該当) があつたが、これらの選択肢は削除された。AMSTAR 2 (2017) で組入れられたすべての領域は、現代のヘルスケア介入の SR に関連していることが確信されるためである。ある項目に関して評価のための情報が無い (Yes にチェックされない) 場合、その項目は “No” (いいえ) として判断される。また、部分的にはスタンダードに沿っているような場合の価値を認め、“Partial Yes” (部分的に Yes) という選択肢が設けられた。

レビューの妥当性と結論に影響を与えるとくに重大な領域が **Box 1** にまとめられている。そのうち 2 つはバイアスリスクに関するものであり、「バイアスリスクは十分に評価されたか」と、「バイアスリスクはどのようにレビューの結果に影響したか」である。AMSTAR 2 (2017) は、非ランダム化研究も組入れる SR の評価にも用いられることが予想されるため、これは重要である。

個々の項目における不十分さを問題とすること、全体の総合得点を算出しないことが再度強調されたうえで、**Box 2** に弱点を解釈する枠組みが提供されている。また、これはあくまで報告であり、評価者自身が熟慮のもとで、どの項目がもっとも重要かを決すべきだとされている。(項目ごとの得点や比重など、基準は一切示されていない。)

AMSTAR 2 (2017) を用いた実際に要する時間、つまり評価にかかる時間は、論文そのものを読み込む時間を除外して 15-32 分間だったとされる (AMSTAR のおよそ 2 倍)。

Box 1 AMSTAR 2 における重大領域¹⁾

- ・項目 2 レビュー開始前のプロトコールの登録
- ・項目 4 文献検索の適切性
- ・項目 7 除外された個々の研究の妥当性
- ・項目 9 レビューに組入れられた個々の研究のバイアスリスク
- ・項目 11 メタ分析方法の適正性
- ・項目 13 レビューの結果を解釈する際のバイアスリスクの考慮
- ・項目 15 出版バイアスの存在とその影響の評価

Box 2 レビューの結果における全体的な信頼性評価¹⁾

高 (High): 重大な弱点はない、または 1 つだけ重大ではない弱点がある

その SR は、関心のある疑問に対する利用可能な研究結果の正確かつ包括的な要約を提供している。

中 (Moderate): 2 つ以上の重大でない弱点がある*

その SR は、2 つ以上の重大でない弱点はあるが、重大な弱点はない。それは、組入れられた利用可能な研究結果の正確な要約を提供しているかもしれない。

低 (Low): 重大でない弱点の有無にかかわらず、1 つだけ重大な弱点がある

その SR は、重大な欠陥があり、関心のある疑問に対する利用可能な研究結果の正確かつ包括的な要約を提供していないかもしれない。

極めて低 (Critically low): 重大でない弱点の有無にかかわらず、2 つ以上の重大な弱点がある

その SR は、重大な欠陥があり、利用可能な研究の正確かつ包括的な要約を提供していると信頼すべきではない。

*重要ではない複数の弱点はレビューの信頼性を減じるかもしれない、全体評価を中から低へと下げるかもしれない。

おわりに

AMSTAR (2007) は確かに短時間で評価ができる簡便性を有し、また日本人特有かもしれないが合計得点 (11 点満点) で個々の SR を示せることが魅力的であった。しかし、AMSTAR 2 (2017) のなかで強調されているように、重要な領域における不備があるかないかが重要であり、合計点が高いことによってそれがマスクされうる危険性があることは指摘のとおりである。

2013 年末までで SR の総数は約 2 万件、そのうちコクランレビューが約 5,500 件と報告されている²¹⁾。また、International Prospective Register of Systematic Reviews (PRSPERO) において、2011 年 2 月の立ち上げ以降、SR の事前登録が 2016 年に 2 万件 (107 ヶ国) に到達したことが報じられている²²⁾。このように膨大な数の SR が計画・実施されている実情を踏まえると、こうして各研究分野で大量生産される SR においてますますその質が問われることになる。

ここで SR が広く使われる 2 つの領域の日本の現状について述べておこう。

第 1 は、医療技術評価 (health technology assessment: HTA) の領域である。広義の費用対効果分析 (cost-effectiveness study: CEA) を基本とする。日本では 1990 年代から HTA の動きがあったが、2016 年からオプジーボ[®] などの超高額薬の登場で、急速に社会的な認知が進んだ。しかし HTA はけっして「費用」だけを論じるのではなくつねに「効果」(effectiveness) とのバランスを研究するものであり、この効果について 1 つのスタディではなく、SR によるエビデンスが必要となるのである^{23,24)}。

第 2 は、診療ガイドライン (CPGs) の領域である。公益財団法人・日本医療機能評価機構 Minds ガイドラインセンター (<https://minds.jcqh.or.jp/>、Minds は Medical Information Network Distribution Service の略称) は、2004 年からそのウェブサイト に収録する CPGs を AGREE (2001) に基づき選定することを開始し、2009 年から AGREE II (2009) に変更した。CPGs 作成にあたって、SR のあり方には、1) 「テーラメード SR」(tailor-made systematic)、2) 「既存の SR」(ready-made systematic review)、3) 「SR のレビュー」(review of systematic review)、の 3 つがある²¹⁾。2015 年には「推奨作成における既存のシステマティック・レビューの活用」²⁴⁾ が、『Mind 診療ガイドライン作成マニュアル ver. 2.0』(2014) の特別寄稿として翌 2015 年 12 月に Minds website に掲載された²⁴⁾。

AMSTAR 2 の Box 2 には、レビューの結果における全体的な信頼性評価が示されている。なかなか手厳しくもあり、1 つでも重大な弱点があったら、その SR の質は “Low” (低) となり、さらに 2 つ重大

な弱点があったら “Critically low” (極めて低) となる。おそらく、現時点において、日本を含めて世界においても、この評価で中から “High” (高) の SR は少ないのではないだろうか。

AMSTAR 2 は、SR の質を高めるためのポイントを知らしめてくれているとも捉えられる。そこで、今後、SR の実施予定者は、計画段階当初から 16 のチェック要素をすべてカバーできるようにすればよい。この AMSTAR 2 を十分に役立てつつ、PRISMA や PRISMA-P、PRISMA-NMA などとあわせて、SR の計画・実施・報告をすることが、実施者の務めであると考えられる。

謝辞：

本稿作成に協力を得た、京都大学大学院医学研究科社会健康医学系専攻健康情報学教授・中山健夫氏に謝意を表す。

附記：

本稿は、日本学術振興会科学研究費助成事業「平成 30 年度基盤研究 C、研究代表者：上岡洋晴、研究課題：システマティック・レビューによる「機能性表示食品」の有効性評価方法に関する研究 (課題番号 17K00885)」の一環として執筆した。

参考文献：

- 1) Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, Moher D, Tugwell V, Kristjansson E, Henry DA. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomized or non-randomised studies of healthcare interventions, or both. *BMJ* 2017; 358: j4008 doi: 10.1136/bmj.j4008
- 2) Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol* 2007; 7: 10.
- 3) 折笠秀樹, 上岡洋晴, 津谷喜一郎 (訳). AMSTAR: システマティック・レビューの方法論的な質を評価するための測定ツールの開発. In: 中山健夫, 津谷喜一郎. 臨床研究と疫学研究のための国際ルール集 Part 2. ライフサイエンス出版 2016 p.148-55.
- 4) Kamioka H, Tsutani K, Okuizumi H, Mutoh Y, Ohta M, Handa S, et al. Effectiveness of aquatic exercise and balneotherapy: a summary of systematic reviews based on randomized controlled trials of water immersion therapies. *J Epidemiol* 2010; 20: 2-12.
- 5) Kamioka H, Tsutani K, Yamada M, Park S, Okuizumi H, Tsuruoka K, et al. Effectiveness of music therapy: a summary of systematic reviews based on randomized controlled trials of music interventions. *Patient Preference and Adherence* 2014; 8: 727-54.

- 6) Kamioka H, Tsutani K, Katsumata Y, Yoshizaki T, Okuizumi H, Okada S, et al. Effectiveness of Pilates exercise: a quality evaluation and summary of systematic reviews based on randomized controlled trials. *Complementary Therapies in Medicine* 2016; 25: 1-19.
- 7) Kamioka H, Tsutani K, Hideki O, Yoshizaki T, Kitayuguchi J, Shimada M, et al. Quality of systematic reviews of the Foods with Function Claims registered at the Consumer Affairs Agency Web site in Japan: a prospective systematic review. *Nutrition Research* 2017; 40: 21-31.
- 8) Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, et al.: GRADE Working Group. Grading quality of evidence and strength of recommendations. *BMJ* 2004; 328: 1490-3. [津谷喜一郎, 中山健夫, 島村治子 訳, エビデンスの質とお勧め度のグレーディング. *薬理と治療* 2005; 35: 1241-54]
- 9) Ioannidis JP. The mass production of redundant, misleading, and conflicted systematic reviews and meta-analysis. *Milbank Q* 2016; 94: 485-514.
- 10) Pollock M, Fernandes RM, Becker LA, Featherstone R, Hartling L. What guidance is available for researchers conducting overviews of reviews of healthcare interventions? A scoping review and qualitative metasummary. *Sys Rev* 2016; 5: 190.
- 11) Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Ann Intern Med* 2009; 151: W65-94.
- 12) Hutton B, Salanti G, Caldwell DM, Chaimani A, Schmid CH, Cameron C, et al. The PRISMA Extension Statement for Reporting of Systematic Reviews Incorporating Network Meta-analyses of Health Care Interventions: Checklist and Explanations. *Ann Intern Med* 2015; 162: 777-84.
- 13) Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015 statement. *Syst Rev* 2015; 4: 1. doi: 10.1186/2046-4053-4-1
- 14) Whiting P, Savovic J, Haggins JP, Caldwell DM, Reeves BC, Shea B, et al. ROBIS: a new tool to assess risk of bias in systematic reviews was developed. *J Clin Epidemiol* 2016; 69: 225-34.
- 15) Haggins JPT, Lasserson T, Chandler J, Tovey D, Churchill R. Methodological expectations of Cochrane intervention reviews. London UK: Cochrane; 2016.
- 16) Pieper D, Buechter RB, Li L, Prediger B, Eikermann M. Systematic review found AMSTAR, but not R(evised)-AMSTAR, to have good measurement properties. *J Clin Epidemiol* 2015; 68: 574-83.
- 17) Banzi R, Cinquini M, Gonzalez-Lorenzo M, Pecoraro V, Capobussi B, Minozzi S. Quality assessment versus risk of bias in systematic reviews: AMSTAR and ROBIS had similar reliability but differed in their construct and applicability. *J Clin Epidemiol* 2018; 99: 24-32.
- 18) Burda BU, Holmer HK, Norris SL. Limitations of a measurement tool to assess systematic reviews (AMSTAR) and suggestions for improvement. *Syst Rev* 2016; 5: 58.
- 19) Wegewitz U, Weikert B, Fishta A, Jacobs A, Pieper D. Resuming the discussion of AMSTAR: what can (should) be made better? *BMC Med Res Methodol* 2016; 16: 111.
- 20) Faggion CM Jr. Critical appraisal of AMSTAR: challenges, limitations, and potential solutions from the perspective of an assessor. *BMC Med Res Methodol* 2015; 15: 63.
- 21) 唐文涛, 河合富士美, 小島原典子, 津谷喜一郎. 診療ガイドラインとシステマティック・レビュー. *薬理と治療* 2014; 42: 189-97.
- 22) International Prospective Register Of Systematic Reviews. PROSPERO REGISTRATIONS REACH 20,000. [https://www.crd.york.ac.uk/PROSPERO/#aboutpage Accessed 1 Sep 2018]
- 23) 唐文涛, 津谷喜一郎. 医療技術評価とシステマティック・レビュー. *医薬ジャーナル* 2014; 50: 81-5.
- 24) 小嶋雅代, 中山健夫. 特別寄稿 推奨作成における既存のシステマティックレビューの活用 (2015.12.15 掲載). In: *Minds 診療ガイドライン作成マニュアル ver. 2.0.2014* [http://minds4.jcqh.or.jp/minds/guideline/pdf/special_articles4.pdf Accessed 1 Sep 2018]

* * *

本稿は, Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, Moher D, Tugwell V, Kristjansson E, Henry DA. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomized or non-randomised studies of healthcare interventions, or both. *BMJ* 2017; 358: j4008 doi: 10.1136/bmj.j4008 に基づき, 上岡洋晴, 折笠秀樹, 津谷喜一郎が一部の訳とともに紹介解説したものである。上記は Creative Commons Attribution (CC BY 4.0) license のもとで配布されているオープンアクセス論文であり, 原論文が適切に引用されているかぎり, 営利目的を含む配布, 再編集, 翻案, 創作が許可されている。