

REVIEW

機能性表示食品制度において届出された システマティック・レビューの報告の質

—PRISMA 2020に基づいての評価—

Reporting Quality of Systematic Reviews Submitted for Foods with Functional Claims in Japan

—Evaluation Based on PRISMA 2020—

上岡 洋晴^{1)*} 北湯口 純²⁾ 折笠 秀樹³⁾ 津谷 喜一郎⁴⁾

ABSTRACTS

Purpose This study aimed to clarify whether systematic reviews (SRs) published on the Consumer Affairs Agency (CAA) website after April 2024 have been appropriately described using PRISMA 2020, a guideline checklist for SR reporting published in 2021.

Methods All submitted papers based on SRs published on the CAA website from April 1, 2024 to September 30, 2024 were eligible. As of September 2024, there is a grace/transition period for use of the new PRISMA 2020 edition instead of the old 2009 edition in the guidelines for the functional food labeling system, therefore, SRs based on the old edition were excluded and only SRs based on the new edition were subject to evaluation. The quality assessment was based on the CAA verification project report, Verification of Scientific Evidence on Effectiveness of the System of “Foods with Function Claims”: Assessment of the Submitted Systematic Literature Reviews. For each item, quality was assessed as “implemented/described”, “non-implemented/no description”, or “insufficient or unclear description”; “non-implemented/no description” and “insufficient or unclear description” were collectively considered as “deficiencies.” To increase the usefulness of this study, we listed the frequency of use of literature search databases (DBs), clinical trial DBs, and other DBs.

Results Forty SRs were included, of which eight (20%) had protocol study registrations. Six SRs (15%) conducted MAs, and the remaining 34 (85%) were qualitative SRs. Among the PRISMA 2020 checklist items, 11 were commonly lacking in more than 40% of the SRs surveyed. The average total number of DBs used in this study was 4.9 (range: 2-14). Fifty percent of SRs did not use the CAA’s “Foods with Function Claims Notification Search.”

Conclusions Particular attention should be paid to common deficiencies during the planning stages of the review, and those deficiencies should be adequately and thoroughly documented without omissions. In this study, the DB used in the search contained many reports with a small number of types, so the number of DBs should be increased to enable future comprehensive searches.

¹⁾ 東京農業大学大学院環境共生学専攻 ²⁾ 雲南市立身体教育医学研究所うんなん ³⁾ 統計数理研究所 ⁴⁾ 生存科学研究所

* Corresponding author: h1kamiok@nodai.ac.jp

Hiroharu Kamioka: Faculty of Regional Environment Science, Tokyo University of Agriculture; Jun Kitayuguchi: Physical Education and Medicine Research Center Unnan; Hideki Origasa: Institute of Statistical Mathematic; Kiichiro Tsutani: The Institute of Seizon and Life Sciences

I はじめに

2015年4月に開始された機能性表示食品制度(本制度)は2024年4月で10年目に入った。2025年1月1日時点において、消費者庁のウェブサイトにおける届出数は9,061件と膨大な数(撤回を含む)になっている。機能性のエビデンスとしては8,706件(96%)が研究レビュー、355件(4%)が臨床試験(clinical trial: CT)の結果を用いており、大量の研究レビューが提出され、実際に多くの商品が販売されている。

ここでの「研究レビュー」という名称だが、「機能性表示食品の届出等に関するマニュアル」(マニュアル)¹⁾では、2009年に公表された初版のPRISMAチェックリスト^{2,3)}、もしくは新版の2021年に出たPRISMA 2020チェックリスト^{4,5)}に準拠して作成するように求めていることから、これはシステマティック・レビュー(systematic review: SR)を意味しており、以後はSRと称する。

本制度は企業等の届出者の責任に基づく届出制であることから、当初からエビデンスの質に関して消費者団体やマスコミなどから疑問を呈する声が挙げられていた。アカデミア研究者においても類する見方があった。機能性表示食品は英文で“Foods with Function Claims (FFC)”という表記なので、われわれは「届出SRをさらにSRする」ということで、2015年3月にFFC-SR2という名の研究グループ(ステアリング・オーサー: 上岡洋晴, 津谷喜一郎, 折笠秀樹)を立ち上げて研究を展開してきた。これはSRをはじめ臨床試験(Clinical Trial: CT)の方法論や報告方法の適正性の評価をする研究によって、その後、機能性のエビデンスに関して、より良い届出がなされることを期待してのものだった。

ところで、ここでのSRとCTの質については、消費者庁の公的調査事業やアカデミア研究者の先行研究も合わせて要約すると、大きく分けて「研究方法論(methodology)の質」と「報告(reporting)の質」の2つあり、一部重なる部分もあるが区別して述べる。

まず、レビューの基盤となるCTについては、2017年に消費者庁の検証事業として「機能性表示食品制度における臨床試験及び安全性の評価内容の

実態把握の検証・調査事業報告書」⁶⁾が公表され、サンプルサイズの設定がない、ランダム化の方法が不明確、多重検定(多項目・多時点)、主要・副次アウトカムが不明確などの研究の質の低さの課題が示された。2018年にはTanemuraら⁷⁾はCONSORT 2010チェックリスト^{8,9)}に基づいていない届出が多いことを報告している。2021年にKamiokaら¹⁰⁾はCTのプロトコールの記載内容と、論文中におけるそれが一致していないことを報告している。具体的には、プロトコールにおいて試験食の内容、すなわち機能性関与成分(関与成分)名を伏せたり、アウトカムを大まかな書き方にして主要アウトカムが何か特定できないようなCTが多いことを明らかにしている。

直近として2024年にSomekoら¹¹⁾は、RCTにおいて“Spin(粉飾あるいはねじ曲げ)”が多く、選択的な結果報告によるバイアスのリスクが高いこと、また要約(72%)、要約の結論(81%)、本文の結果(44%)、本文の結論(84%)の結果にこれが存在することを報告している。同年に、Kamiokaら¹²⁾は38編のCTについてa revised tool to assess risk of bias in randomized trials (RoB 2)を用いてバイアスリスクを評価した結果、76%のCTが全体評価として3段階の中でもっともバイアスが高いことを意味する「低い:Low」だったことを報告している。とくに報告結果の選択バイアスが高く、Somekoらの報告と一致していた。これらの問題に対して、上岡ら¹³⁾は2023年時点において、CTの改善ポイントの要約として「研究計画のプロトコールに関して透明性をもって正確に記載するとともに、その段階でバイアスのない方法を適切にセットして実施すること、そしてチェックリストに基づいて丁寧に記載することが必須である」と指摘している。

一方、SRの報告の質については、2016年に消費者庁から「機能性表示食品制度における機能性に関する科学的根拠の検証:届出られた研究レビューの検証事業報告書」¹⁴⁾が公表された。これは初版の「PRISMAチェックリスト」^{2,3)}に従って報告しているかを評価したもので、不備の多い事項や問題点、さらに改善すべき事項が明確にされている。その中では、望ましい書き方の例示もなされている。前述のFFC-SR2は、この2年後となる2018年に、SR

の報告の質が向上したかどうかを明らかにすることを目的として、同一の方法を用いて検証事業前後の比較研究（2015年に届出されたSRと2017-2018年に届出されたSR）を行った¹⁵⁾。その結果、仮説と反して検証事業後の方が、不備率が有意に増加していた。つまり、報告の質は改善していないということだった。

SRにおける研究の質については、FFC-SR2は2007年に公表された初版の「AMSTARチェックリスト」^{16,17)}に基づいて、前述の検証事業とまったく同じ対象のSRについての前後比較を行った。SRとしてすべて良好な方法に基づいていれば11点満点になるが、2015年の検証事業前SRでは6.2 ± 1.8（平均±標準偏差）点で、2017-2018年の検証事業後SRでは5.0 ± 1.9点と、研究の質は有意に低下していた^{18,19)}。

さらに最新の研究として、FFC-SR2は2017年に公表された新版であるAMSTAR 2^{20,21)}を用いて、2022年4月1日から同年10月31日までに消費者庁ウェブサイトにて公表された記号番号HシリーズのSR 519編のうち、ランダム抽出した40編のSRを評価した²²⁾。その結果、全体の信頼性評価（4段階）では、「高 (High)」が0編 (0%)、「中 (Moderate)」が0編 (0%)、「低 (Low)」が2編 (5%)、「Critically low」が38編 (95%)であった。とくに信頼性を下げていたのは、「プロトコルの事前登録の欠如」「包括的な検索戦略の不備」「RoB評価の結果を考察や結論に適切に反映していない」ことであった。このように2022年代のSRにおいても質の向上がみられないことが明らかになっている。これらのことは本制度開始からずっと研究の質の低い届出がなされ続けていることを意味している。

そしてこの間、2021年にPRISMA 2020^{4,5)}が公表され、本制度においても2024年からマニュアルに加わり、一定の猶予期間はあるものの、届出SRはこの新版に準拠することが求められている。これまでの届出SRの研究の質と報告の質がともに低い中で、われわれはこの新版において適切に報告されているかどうかに関心があった。

そこで本研究は、2024年4月1日以降に消費者庁ウェブサイトにて公表されたSRの届出を対象として、PRISMA 2020チェックリストを用いて、これ

らに基づいて適切に記載されているかどうかを明らかにすることを目的とした。

II 方法

1. 適格・除外基準

2024年4月1日から同年9月30日までの期間に消費者庁ウェブサイトにて公表されたSRで、PRISMA 2020チェックリストに基づいて届出されたすべてのSRを対象とした（撤回されたSRは除外）。本制度では旧版を用いることの猶予・移行期間であるため、旧版に基づくSRは除外し、新版に基づいて届出されたSRだけを評価の対象とした。同一のSRを機能性の根拠資料として複数の届出がされているような重複がある場合は、最初のSRだけを採用した。1つの届出の中で複数のクレームを表示するために、複数のSRが含まれている場合には、それぞれ別のSRとして採用した。

2. データ抽出源

対象のSRは消費者庁ウェブサイトの機能性表示食品の届出情報からダウンロードした。

3. データ項目と報告の質評価（要約尺度）

PRISMA 2020チェックリストの各項目を用いるが、「#2抄録」部分は省略した。質評価は、前述の消費者庁の検証事業報告書¹⁴⁾に準拠し、「実施・記述あり」「不履行・記述なし」「不十分または不明瞭な記述」で示した。「不履行・記述なし」と「不十分または不明瞭な記述」を合わせて「不備」とした。

一般的なSRでは、論文中に記載がない項目においては著者確認を行う。しかし、本研究においては、届出SR中にすべての項目が適切に記載されているかどうかの程度自体を把握することに注目しているため、本文中に記載がなかった場合（記載はないが実際には適切に実施していたとしても）には著者確認は行わなかった。また、不備が多い事項について、ここでは40%以上のSRで不備のあった項目に関してその内容を要約し、改善のポイントとともに一覧化した。

なお、本研究は、報告の質（履行と明確な記述）だけの形式的な評価を目的としているので、バイア

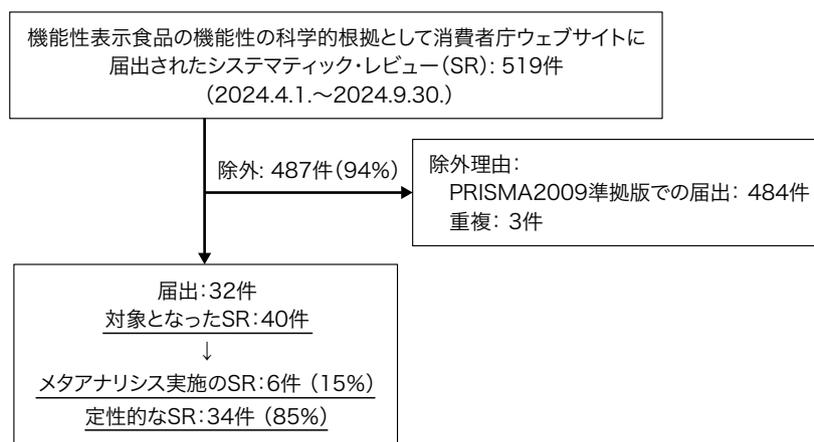


図1 対象としたシステマティック・レビューの選択プロセス

スリスク、非直接性、不精確、非一貫性、出版バイアス、エビデンスの確実性の評価方法が実際に適正だったかどうかについての評価は実施しなかった。また、メタアナリシス(MA)の実施方法の適正性についても評価しなかった。MAに関する記載の評価は、MAを実施したSRのみを行った。報告バイアスについても、狭義な意味で取り扱い、MAを実施したSRのみで評価した。

ところで、プロトコルで設定していなかったが、研究を進める途中で、検索で使用したデータベース(DB)がかなり少ないことに気づいた。そのため、本研究の有用性を高めるために、アドホックにて文献検索DB、臨床試験DB、その他のDBの利用の特徴をリスト化した。

4. エビデンス・テーブル

本研究は、介入としての試験食の機能性を明らかにすることが目的ではないので、構造化抄録は作成しない。

5. 統計解析

基本統計として、各項目と全体の質評価結果と使用したDBの結果について度数(%)と平均値(レンジ)で示した。統計解析ソフトはSPSS Statistics 25.0(IBM Corporation, Armonk, NY, USA)を用いた。

6. 研究登録

本研究は2024年10月1日にプロトコルを確定し、同日にUMIN-CTR(UMIN000055686)に

研究登録を行った。なお、詳細プロトコルは次のURL(https://1drv.ms/b/s!AoQmpnIHE3YUhOxrhcEa_5wP8WHpZw?e=yEGdfO)から閲覧できる。

III 結果

図1は対象となった40編のSR(届出数32)の採用プロセスである。このうち、プロトコル登録をしていたのは8編(20%)だった。MAを実施したSRは6編(15%)、残りの34編(85%)は定性的SRだった。

1. PRISMA 2020における項目別評価(図2, 付録2)

#1 タイトルについては、40編(100%)すべてで不備はなかった。#3 論拠においては、不備が36編(90%)と多かった。先行研究の引用、これには同一の関与成分ですでに届出られたSRを含むが、それらを見逃して記載がなかったり、作用機序、その他の関連研究を十分に踏まえていないこと、またなぜレビューをする必要があるのかの記載が不十分だった。#4 目的においては、不備は12編(30%)だった。

方法の部分では、#5 適格基準においては、不備は33編(83%)で、試験食の摂取量や摂取期間の定義づけがないSRが多かった。ただし、所定の届出資料としての別紙様式(V)-7があるため指定項目の記載はあった。#6 情報源では、不備は22編(55%)で、CTRの記載がないSRが多かった。#7 検索戦略と#8 選択のプロセスでは、40編(100%)すべてで不備はなかった。#9 データの収集プロセスで

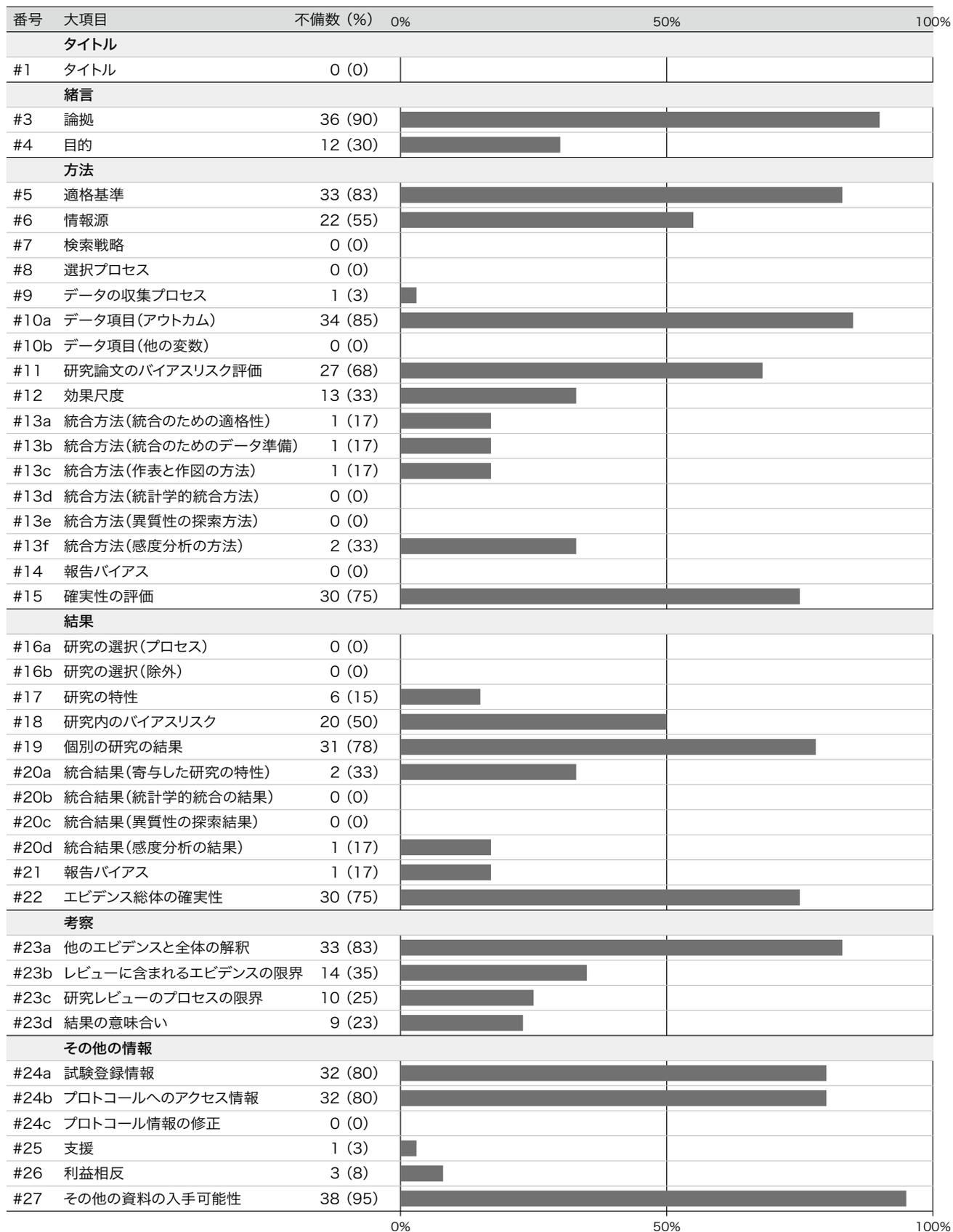


図2 PRISMA 2020 チェックリストに基づく各項目における不備の特徴

は、1編(3%)に不備があった。#10a データ項目(アウトカム)では、34編(85%)に不備があり、とくにアウトカムに関して大まかな括りでの記載や、時間枠(何週目時点なのか、最終時点なのかなど)の記載がないSRが多かった。#10b データ項目(他の変数)については、40編(100%)すべてで不備はなかった。#11 研究論文のバイアスリスク評価では、27編(68%)に不備があり、個々のバイアス項目をどのように評価したのか、具体的には3値での基準(低・中/疑いあり・高)がどのように設定されていたのか不明なSRが多かった。

#12 効果尺度においては、13編(33%)で不備があり、アウトカムにおける両群間の平均値差にするのかなどの記載がないものがあった。#13a MAにおける統合方法(統合のための適格性)においては、6編中1編(17%)に不備があり、統合するのに適格となる研究を決定するプロセスを記載していなかった。#13b 統合方法(統合のための準備)では、1編(17%)に不備があり、欠測値があるのかどうか不明瞭であった。#13c 統合方法(作表と作図の方法)では1編(17%)の不備があり、フォレスト・プロットなどの作図をどのようにするのかの記載がなかった。#13d 統合方法(統計学的統合方法)では、6編(100%)すべてで不備はなかった。

#13e 統合方法(異質性の探索方法)でも、6編(100%)すべてで不備はなかった。#13f 統合方法(感度分析)では、2編(33%)に不備があり、感度分析の記載がない、あるいは不明確だった。#14 報告バイアスについては、6編(100%)すべてで不備はなかった。#15 エビデンスの確実性の評価では、不備は30編(75%)あり、Minds 診療ガイドラインに従ったと記載されているが、5項目(バイアスリスク、非直接性、非一貫性、不精確、報告バイアス)のダウングレードの詳細と、全体の決定ルールが記載されていないSRが多かった。1-2編の定性的なSRで、効果推定量が算出できていないにもかかわらず、エビデンスの確実性が中や高などとするSRが数多く存在していた。

結果の部分において、#16ab 研究結果の選択では、40編(100%)すべてに不備はなかった。#17 研究の特性では、6編(15%)で不備があり、対象となった研究を本文中で引用していないSRがあった。

#18 研究内のバイアスリスクでは、20編(50%)において不備があり、本文中に評価結果を適切に示していないSRがあった。#19 個別の研究結果では、31編(78%)で不備があり、とくに有意確率のみで群間の平均値差が示されていないSRが多かった。#20a 統合結果(寄与した研究の特性)では、2編(33%)に不備があり、研究間の特徴やバイアスリスクの記載が不十分だった。#20b 統合結果(統計学的統合の結果)では、6編(100%)すべてに不備はなかった。#20c 統合結果(異質性の探索結果)でも、6編(100%)すべてに不備はなかった。#20d 統合結果(感度分析の結果)では、1編(17%)で不備があった。#21 報告バイアスでは、1編(17%)で不備があり、その記載がなかった。#22 エビデンス総体の確実性では、30編(75%)で不備があり、#15と同様に、ダウングレードした詳細な理由と全体の決定の流れの記載がないSRが多かった。

考察の部分において、#23a 他のエビデンスとの関連で結果についての全体的な解釈では、33編(83%)に不備があり、結果を繰り返して記載しているだけで深い議論がなされていないSRが多かった。#23b レビューに含まれるエビデンスの限界では、14編(35%)で不備があり、潜在的なバイアスに関しての批判的吟味がなされていないSRが多かった。#23c 実施した研究レビューのプロセスの限界では、10編(25%)で不備があり、SR自体の弱点の記載がなかった。#23d 実践、政策、将来の研究のための結果の意味合いについては、9編(23%)で不備があり、その記載がなかった。ここでの評価は、実践と政策は除いて将来の研究のためだけに限定して行った。

その他の情報において、#24ab 登録とプロトコールでは、ともに32編(80%)で登録をしておらず、その内容も特定することができなかった。#24c プロトコール情報の修正では該当なしだった。#25 支援においては1編(3%)のみ不備があった。#26 利益相反では、3編(8%)で不備があった。#27 データ、コード、その他の資料の入手可能性においては、38編(95%)で届出書類以外、一切の資料を提供していなかった。2編(5%)においては、2021年に公表された検索戦略に特化したチェックリストであるPRISMA-S^{23,24)}に基づき、ピアレビュー結果

表 1 検索で利用したデータベースの特徴

	使用件数	使用割合 (%)
【文献データベース】 平均 3.2 (2-6) 種類		
PubMed	40	100.0
JDream III	34	85.0
Cochrane Central Register of Controlled Trials	23	57.5
医中誌 Web	23	57.5
AGRIS	2	5.0
CiNii Research	2	5.0
Global Index Medicus	2	5.0
CINAHL	1	2.5
Cochrane Database of Systematic Reviews	1	2.5
Embase	1	2.5
ProQuest Central	1	2.5
Western Pacific Region Index Medicus	1	2.5
Global Health Library	0	0.0
Scopus	0	0.0
Web of Science	0	0.0
【研究登録データベース】 平均 0.9 (0-6) 種類		
UMIN-CTR	19	47.5
ICTRP	5	12.5
ClinicalTrials.gov	4	10.0
ISTCTN	2	5.0
PROSPERO	2	5.0
臨床研究情報ポータルサイト	2	5.0
jRCT	1	2.5
【その他のデータベース】 平均 0.7 (0-2) 種類		
消費者庁 機能性表示食品の届出情報	20	50.0
J-Stage	6	15.0
Nature Communications	2	5.0
【合計】 平均 4.9 (2-14) 種類		

の資料が添付されていた。

2. 検索で利用した DB の特徴

DB の特徴を数とともに示した(表 1, 付録 3)。文献 DB では, 平均 3.2 種類(レンジ: 2-6 種類)を使用していた。その中で, 使用頻度が高い順に PubMed (40 編, 100%), JDream III (34 編, 85%), Cochrane Central Register of Controlled Trials (23 編, 57.5%), 医中誌 Web (23 編, 57.5%), AGRIS (2 編, 5%), CiNii Research (2 編, 5%) などとなっていた。

研究登録 DB では, 平均 0.9 種類(レンジ: 0-6 種類)を使用していた。頻度が高い順に, UMIN-CTR

(19 編, 47.5%), ICTRP (5 編, 12.5%), ClinicalTrials.gov (4 編, 10%), PROSPERO (2 編, 5%), 臨床研究情報ポータルサイト (2 編, 5%) などとなっていた。

その他の DB としては, 平均 0.7 種類(レンジ: 0-2 種類)を使用していた。頻度が高い順に, 消費者庁機能性表示食品の届出情報 (20 編, 50%), J-Stage (6 編, 15%), Nature Communications (2 編, 5%) となっていた。

全体を合計してみると, 平均 4.9 種類(レンジ: 2-14 種類)で, そのうち DB が 2-3 種類の SR は 11 編 27.5%, DB が 2-4 種類とすると 21 編 52.5%

ということで、DB 数が少ない届出が多いことが明らかになった。

IV 考察

本研究は、本制度における届出 SR が PRISMA 2020 における要求項目と合致しているかを明らかにした初の報告である。定性的な SR が全体の 85% と多いことに加えて、記載の不備も多いことが明らかになった。その不備については、かなり共通している部分があり、各届出者はこの部分の誤認があると考えられた。また、検索で使用している DB が極端に少ない SR が一定数あることも明確になった。

1. 不備が共通する項目

不備が共通して存在する項目（全体の SR の 40% 以上）とその改善のポイントを表 2 にまとめた。改善のポイントを 1 フレーズで言及するなら、「都合良く簡略化しないこと、丁寧に記載すること」である。論拠に関しては 90% の不備があったが、関与成分に関して、これまでの届出 SR を無視して引用していないことや、作用機序の概要、その他の先行研究を十分にショート・レビューすることが求められる。前述した 2009 年に公表された初版の PRISMA チェックリストに基づく検証事業報告¹⁴⁾と、FFC-SR2 による検証事業前後の比較研究¹⁵⁾を参考値としてみると、それぞれ 49%、70% の不備率であり、問題は不変であることを示している。

適格基準では関与成分の摂取量や摂取期間の明確化を基本として、その基準の曖昧さを残さないことが重要である。情報源においては、網羅的な検索のために重要だが、DB が少ないという基本的な問題が存在している。また、消費者庁の機能性表示食品の届出情報の DB すら検索していない SR が過半数であることから、同一の関与成分の先行研究 (SR) を精査していない、ということを裏付けることにもなっている。データ項目では、先行研究^{6, 10, 11, 13, 14, 19, 22)}が指摘しているように、選択的アウトカムを疑われる書き方、つまり詳細な主要アウトカムでなく大まかな括りでのアウトカムの記載や、時点の設定がない SR が多く、これらの明確化が求められる。バイアスリスクに関しては、個々の項目でどのように評

価されたかを丁寧に記載する必要がある。

そして、結論づけに大きな影響を及ぼすエビデンスの確実性においては、Minds 診療ガイドラインに準拠したとあるだけで、不明瞭なダウングレードや、全体の決定ルールを示さないまま実施している SR が 75% と、深刻な問題になっている。さらには MA を実施していないため、効果推定量が算出されていない定性的 SR でさえも、このように不明瞭なまま、エビデンスの確実性を「高」「中」としている SR が少なくない。本研究で扱う範疇で、提示された確実性の適正性評価は計画していないため言及はここまでにするが、本制度上、マニュアルなどにも反映させ、届出者に徹底すべき最重要事項である。届出者によって、ある程度ダウングレード方式や全体の決定ルールが異なることはやむを得ないが、そうであってもその詳細は記載すべきであることを強調する。

考察においては、結果の繰り返しにとどまっている SR がほとんどであり、関連する先行研究、関与成分の作用機序、害と益のバランスなどに基づく批判的吟味が乏しい。SR はエビデンス・グレーディング最上位の研究手法であり、サイエンスとしての適切な考察が望まれる。

プロトコールの研究登録は依然として課題である。AMSTAR 2^{20, 21)}でも示されているように、研究登録をしていない時点で重大領域としての不備とし、信頼性はすでに 4 段階で下から 2 番目の「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」となる。つまり、その SR の結果は信頼に値しない、もつとえば機能性の届出として不可だと換言してもよいだろう。登録が進まない一番の理由に、競合他社に届出する予定の関与成分・商品を知られてしまうということが考えられる。この点に関してもルール化が示されないと届出者側の自主的な改善は困難かもしれない。

データ、コードなどのその他の資料に関しては、秘匿性を理由に公開していない SR がほとんどだった。しかし、できる限りの生データを示すことによって、反対に本当に SR を行ったという証拠になり、届出者の社会的信頼を高めることにつながる。たとえば GABA などのメジャーともいえる届出 SR の

表 2 共通する不備と改善のポイント*

大項目	番号	共通する不備の特徴	改善のポイント	不備率
論拠	#3	知見の現状や不明な点の記載がない。 関与成分に関する簡略化した作用機序の記載がない。 類似の先行レビュー(届出のこと)の記載がない。 ⇒唐突で脈略がないシステムティック・レビュー(SR)が多い。	関与成分に関する簡潔な作用機序を示すとともに、なぜ、今回のレビューが必要なのか、更新版の場合にはどのような意義があるのかを記載する。その際、関連する主要な臨床試験(CT)やSRの結果を引用する。そのSRは自社や他社の届出されたものも含む。あまりにも数が多い場合には別添でリストアップすることで要約することも可能である。	本研究 90% 2019年 ^a 70% 2016年 ^b 49%
方法				
適格基準	#5	摂取量、摂取期間、性状の記載がない。 関与成分によるが、定量的・定性的同等性の記載がない。 関与成分によるが、基原と抽出方法の記載がない。 発表年・言語の記載がない。 ⇒基準が不明確なSRが多い。	関与成分に関する定量的・定性的な十分な情報を記載する。また、採用するCTの発表年と言語を明確にする。	本研究 83% 2019年 ^a — 2016年 ^b —
情報源	#6	研究登録データベース(DB)を活用していない。 消費者庁機能性表示食品検索サイトを活用していない。 文献DBの数が少なすぎる。 ⇒採用すべき潜在的なCTが存在する可能性がある。	SRはできる限りの網羅的な検索が必要なので、できるだけ多くの文献DBを利用する。国際誌のDBとしてはPubMedだけでは不十分である。また、研究登録DBとして最低限UMIN-CTRでの検索が必須である。届出SRに関する情報を得ることも不可欠なので、その検索も行わなければならない。	本研究 55% 2019年 ^a 58% 2016年 ^b 43%
データ項目 (アウトカム)	#10a	アウトカムの不明確な記載(例:筋力とは大まかな記載であり、実際に測定しているのは握力や脚伸展力などである)。 時間枠、すなわち時点(何週間後なのか)における記載がない。 ⇒アウトカムが不明確なので結果の解釈を難しくしている。	主要アウトカムの詳細やその定義づけ、時点を明確に記載する。当然、プロトコールにおいてである。	本研究 85% 2019年 ^a — 2016年 ^b —
研究論文の バイアス リスク評価	#11	個々の項目を具体的にどのように評価したかの記載がない。 ⇒正しい評価がなされているか疑義が生じる。	バイアスリスクを評価したツールやバージョンを明記するとともに、各バイアス項目の具体的な評価基準を明確にする。たとえば、2値で「あり」「なし」なら明確だが、「低」「中/疑いあり」「高」などのような場合には、その3つの基準があるはずなので詳細を記載する。	本研究 68% 2019年 ^a 80% 2016年 ^b 57%
確実性の評価	#15	多くがMinds診療ガイドライン作成マニュアルに準拠したと記載されているが、5項目(バイアスリスク、非直接性、非一貫性、不精確、報告バイアス)に関するダウングレードの詳細な基準、全体の決定ルールに記載がない。具体的にはどのように-1、-2のダウングレードをするのか詳細をプロトコールで示さなければならない。 1編や2編によるメタアナリシス(MA)を実施していない定性的なSR、すなわち効果推定量が算出されていないにもかかわらず、確実性が高めに評価されているSRが多い。 ⇒確実性は最重要事項だが、著者の意図する評価、すなわち恣意的な結果になっている疑義が生じており、最大の問題点である。	5項目についてのダウングレードで「-1はxxxの場合」「-2はxxxの場合」などのように具体的な基準を記載する。全体においても、4段階「高・中・低・かなり低」はどのように判断するのか詳細を記載しなければならない。 上述に関連するが、とくに不精確、非一貫性、報告バイアスについては厳格なグレードダウンを行うプロトコールが必須である。	本研究 75% 2019年 ^a — 2016年 ^b —

表 2 共通する不備と改善のポイント* (つづき)

大項目	番号	共通する不備の特徴	改善のポイント	不備率
結果				
個別の研究結果	#19	各研究におけるアウトカムの要約統計量を示すが、本文中において有意確率だけで、平均値差の記載のないものが多い。 ⇒主要アウトカムの平均値差の大きさと有意性が重要なので、両方がないと結果を解釈できない。	本文中に主要アウトカム(副次アウトカム)の平均値差と有意確率を合わせて記載する。有意確率はP<0.05ではなく、P=0.042のように数値を記載する。	本研究 78% 2019年 ^a — 2016年 ^b —
エビデンスの 確実性	#22	#15と関連して、ダウングレードした理由と全体の決定ルールに基づく記載がなされていない。 ⇒確実性の適正性に疑義が生じている。	ダウングレードした理由と全体の決定ルールの結果を記載する。	本研究 75% 2019年 ^a — 2016年 ^b —
考察				
考察(解釈)	#23a	結果を繰り返しただけで、考察となっていない設 えのものが多い。批判的吟味がほとんどなされて いない。 ⇒考察が極めてわずかな記載にとどまり、科学的 で適切なSRになっていない。	先行研究(作用機序、他のSR、CTなど)関与成分 の定量的・定性的な同等性、日本人への外挿性、 有害事象を含め益と害のバランス、エビデンスの 確実性を踏まえた議論を十分に記載する必要が ある。	本研究 83% 2019年 ^a 86% 2016年 ^b 77%
その他の情報				
試験登録名と登 録番号の情報、 プロトコルに アクセスできる 場所	#24a #24b	多くのSRにおいてプロトコルのPROSPEROや UMIN-CTRなどへの試験登録がなされていない。 ⇒HARKingやp-hakingなど、PICOとくにIとOの 変更・微修正など、後付けで解析やレビューを進 めてきたのではないかという疑義が生じる。	理想的にはPROSPEROだが、UMIN-CTRなどに 事前にプロトコルを登録する。もし、そのシステ ムではプロトコルの詳細を書ききれない場合に は、その詳細プロトコルを半永久的に誰でも閲 覧できるクラウド上に保存・公開し、そのURLなど を明記してアクセスできるようにする。	本研究 80% 2019年 ^a 98% 2016年 ^b 96%
データ、コード、 資料の入手 可能性	#27	秘匿性を保つために示さないとするSRが多かつ た。また、届出資料における図表をデータだと誤認 しているSRや先行研究リストをもってデータ・資 料としているSRがあった。 ⇒とくに同一の関与成分でのSRが多いような場 合には、できる限りの生データを示さないと本当 にSRを実施したのか、先行のSRを模倣しただけ ではないのかという疑義が生じる。	SR著者2名が独立してバイアスリスクを評価する のが基本だが、その一致度を照合した表や、デー タを抽出してきた元の表など、他の届出者に有利に なるようなことのないデータや資料はあるはずな ので可能な限り公開するようにする。これも詳細プ ロトコルと同様に半永久的に誰でも閲覧できる クラウド上に保存・公開し、そのURLなどを明記し てアクセスできるようにする。	本研究 95% 2019年 ^a — 2016年 ^b —

*付録4の「PRISMA 2020におけるチェックリストの詳細版(細目)」とともに読むことを強く推奨する。

^a Kamioka H, Tsutani K, Origasa H, Yoshizaki T, Shimada M, Kitayuguchi J, et al. Quality of systematic reviews of the Foods with Function Claims in Japan: Comparative before-and after-evaluation of verification reports by the Consumer Affairs Agency. *Nutrients*. 2019;11:1583.の結果を参考値として

^b 消費者庁. 機能性表示食品制度における機能性に関する科学的根拠の検証: 届け出られた研究レビューの検証事業報告書(2016年)を参考値として

http://www.caa.go.jp/foods/pdf/food_with_function_report_0001.pdf

— 基準が大幅に異なるため参考値として表示しない。

多い関与成分では、既存のSRが多いためにSRを行わず、それらをコピー & ペースト、すなわち部分的盗用をして届出しているのではないかと、という疑義が生じている。それを払拭できるのが実際に用いた各種のデータなので、あらためて表示することを推奨する。一方で、資料として添付されて届出の中で、SRの最先端となるピアレビューを実施し、検索を漏れなく適切に行う努力をしているSRもあり、そうした届出の質の格差が今後はますます注目されることになるだろう。

一連の共通する不備とその改善のポイントを、研究計画段階で十分にチェックし、内容を組み入れることでSRの報告の質は向上するし、同時に部分的ではあるが方法の質も高まると確信するので大いに活用がなされることを期待したい。

なお、PRISMA 2020には、解説と詳細を意味する“PRISMA 2020 explanation and elaboration²⁶⁾”があり、正しい理解にはここで示されている小項目や例を把握する必要がある。われわれは小項目を抜粋した形での和訳版と、さらに解説を加えた資料を付録4に示したので合わせて参考にしてもらいたい。

2. 検索で使用したDBの特徴

PubMedの使用は100%で共通していたが、PubMedに掲載されていない国際誌は数多い。とくに新しいオープン・アクセス誌が増加する中で、これらの漏れを防ぐためにはできるだけ多くのDBを駆使することが必須である。しかし、PubMedとJDream IIIの2つだけに象徴されるようにDB数が少ないSRが多かった。一部のDBを除き、使用するには契約が発生し費用負担があることが妨げの原因になっているのかもしれない。

本研究におけるトータルでの使用したDBは平均4.9種類だったが、参考値として2018年のFFC-SR2の研究結果²⁶⁾では平均4.3種類で、本研究とほぼ同じであった。研究登録DBは基本的に無料であることから、報告バイアスを回避するためにも積極的な利用が望まれる。そして、消費者庁の機能性表示食品の届出情報DBの検索は必ず実施すべきであり、先行するSRを介しての後方検索も可能になるし、検索式も参考になるだろう。別の事項になるが、もし、先行SRの検索を全部あるいは部分的に

でも引用したのなら、それを本文中に記載しないと盗用になることに注意しなければならない。

3. 定性的SRの取り扱い

本研究では、MAを実施していない定性的SRが85%だった。参考値として検証事業報告¹⁴⁾では82%、FFC-SR2による検証事業前後の比較研究¹⁵⁾では88%がそれぞれ定性的SRだった。エビデンスの確実性(効果推定量について)で締め括られるように、SRの基本形は統計解析部分を含むMAありのスタイルである。しかし、異質性の問題や研究自体が少ないという理由が主だろうが、依然として定性的なSRで結論づけがなされている。

この背景には、現在のマニュアルの「別紙4 システマティック・レビューの実施手順に係る考え方²⁷⁾」の中で、「前略・・・なお、メタアナリシスの実施に当たっては、論文間の異質性評価に係る知識等、高度な専門性が要求される。SRの実施経験がない者(査読付きSR論文の著者としての実績がない者等)がメタアナリシスを実施することは避けるべきである」という記載がある。このため、あえて定性的SRとして設えていることが考えられる。2015年の本制度開始時には、食品業界は臨床研究(とくにSR)に対する理解が極めて低く、適切な方法論に基づかずMA解析のできるソフトウェアに数値だけ入力して誤った統合結果を用いることを避けるために、旧ガイドラインにこうした文言が入れられたと記憶している。それが現行のマニュアルの別紙としても残っている。本制度開始から10年経過し、少しずつSRに関する理解も進んでいることや、PRISMA 2020も実際にMAとエビデンスの確実性を基本としての記載要求事項となっていることから、早期にこの推奨を削除すべきだと考える。

さらにこの文言が残っていることの弊害がある。基本的にSRは、MAを想定したプロトコルを作成し、異質性の問題などで不可能だった場合には定性的に解析、そして批判的吟味、慎重な結論づけをするように計画すべきところ、定性的SRは最初からMAはしない、という設えで計画されているものが多い。さらに、効果推定量が後発のCTが出たとしても、どれほどぶれることがないかをみるのがエビデンスの確実性であることを理解せず、前述し

た Minds 診療ガイドラインに基づいた、という高い言葉（錦の御旗）だけで、高い自由度をもってこれを評価して届出しているという深刻な問題がある。こうした背景を鑑みて、当該箇所の削除や、MA を想定したプロトコール作成を求める早急な対応が必要だと考える。

4. 研究の限界

本研究にはいくつかの限界がある。第一に、報告の質を明らかにする研究であるため、各チェック項目に関して記載があるかどうかの形式的な評価にとどまり、それぞれの項目が正確に実施されたかどうかに関しては不明である。第二に、2024 年 4 月から 9 月の期間に掲載された SR に限定しているため、その前後分の SR を扱っていないことによる報告バイアスが少なからず存在する。第三に、今回は定性的 SR における MA の部分の記載の適正性については評価していない。最後に、PRISMA 2020 チェックリストは、質評価のために開発されたわけではないので、評価としては実施しにくい部分もある。

V 結論

PRISMA 2020 に準拠して報告したとする SR において、40% 以上の高率で共通して不備のある 11 項目を特定することができた。レビューの計画段階で注意を払い、欠落することなく適切かつ十分な記載が必要である。

検索で使用した DB では種類が少ない届出も多いことや、機能性表示食品の届出情報の検索をしていない SR が多く、網羅的な検索のために DB を増やすべきことが明らかになった。

定性的 SR が依然として多く、適切な手順を踏んだ MA を含んでの結果報告の必要性が明らかになった。

2025 年 4 月 1 日から本制度のマニュアルはさらに格上となり、告示という位置づけになる。これに便乗して本研究の果実がそれに反映されることで、科学的な届出資料としての SR のスタイルが保てるようになること、そして結果的に消費者のより賢明な商品選択に寄与することを期待したい。

【附記】 本論の資金源は、2024 年度日本学術振興会科学研究費基盤 C：研究題目「機能性表示食品制度において届出される機能性の科学的根拠の質向上プログラムの開発 (24K14658)」代表研究者：上岡洋晴」であった。主宰者は上岡洋晴である。

膨大な届出資料・データの収集・整理等を行っていただいた研究アシスタントの石田美千子氏、検索データベースに関しての詳細確認をしていただいた眞喜志まり氏（慶應義塾大学医学部）と佐山暁子氏（聖路加国際大学病院）、データ編集加工を担っていただいた五味達之祐氏（東京都健康長寿医療センター研究所）に感謝の意を表す。

【利益相反】 著者上岡は、ファンケル(株)、カゴメ(株)、丸善製薬(株)、サントリー・ウエルネス(株)、サントリー食品インターナショナル、森下仁丹(株)、富士フィルム(株)、一丸ファルコス(株)、サラシア族植物普及協会、味の素 AGF (株)の機能性表示食品における研究レビューの指導を行って謝金を得た。他の著者においては申告すべき事項はない。

【付録】 本研究における生データや PRISMA 2020 の詳細版（細目）は下の URL からそれぞれ閲覧できる。

【付録 1】 消費者庁届出検索情報一覧

<https://1drv.ms/b/s!AoQmpnIHE3YUhpCfovGULOte1cqyOA?e=kzsi7Q>

【付録 2】 PRISMA 2020 チェックリストに基づく報告の質評価一覧

<https://1drv.ms/b/s!AoQmpnIHE3YUhpZ0807NX4wZoRyxw?e=ngRP9H>

【付録 3】 検索で使用したデータベース一覧

<https://1drv.ms/b/s!AoQmpnIHE3YUhpPZQGnftI5i7Lkq2LQ?e=TU3khZ>

【付録 4】 PRISMA 2020 におけるチェックリストの詳細版（細目）

<https://1drv.ms/b/s!AoQmpnIHE3YUhpCHeC2zpAyCL7DKuw?e=pk83K5>

文 献

- 1) 消費者庁. 機能性表示食品の届出等に関するマニュアル (2024 年 8 月 30 日版).
https://www.caa.go.jp/policies/policy/food_labeling/foods_with_function_claims/assets/food_labeling_cms205_240902_02.pdf
- 2) Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009; 21;6:e1000097.
- 3) 金子善博, 津谷喜一郎, 中山健夫. システマティック・レビューおよびメタアナリシスの報告における望ましい報告項目: PRISMA 声明. In: 中山健夫・津谷喜一郎編著. 臨床試験と疫学研究のための国際ルール集 (Part 2). 東京: ライフサイエンス出版; 2016.p.140-7.
- 4) Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC,

- Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71.
- 5) 上岡洋晴, 金子善博, 津谷喜一郎, 中山健夫, 折笠秀樹. 「PRISMA 2020 声明: レビュー報告のための更新版ガイドライン」の解説と日本語訳. *薬理と治療* 2021;49:831-42.
 - 6) 消費者庁. 機能性表示食品制度における臨床試験及び安全性の評価内容の実態把握の検証・調査事業報告書. https://warp.da.ndl.go.jp/info:ndljp/pid/11257153/www.caa.go.jp/policies/policy/food_labeling/about_foods_with_function_claims/pdf/foods_index_23_171025_0001.pdf
 - 7) Tanemura N, Hamadate N, Urushibara H. Evaluation of randomized controlled trials of Foods with Functional Claims request: The learning outcomes from studies in Japan. *J Funct Foods* 2018; 42: 248-53.
 - 8) Schulz KF, Altman DG, Moher D; CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010; 340: c332.
 - 9) 津谷喜一郎, 元雄良治, 中山健夫. CONSORT 2010 声明ランダム化並行群間比較試験報告のための最新版ガイドライン. *薬理と治療* 2010;939-47.
 - 10) Kamioka H, Origasa H, Kitayuguchi J, Tsutani K. Compliance of clinical trial protocols for Foods with Function Claims (FFC) in Japan: Consistency between clinical trial registrations and published reports. *Nutrients*. 2021;14:81.
 - 11) Someko H, Yamamoto N, Ito T, Suzuki T, Tsuge T, Yabuzaki H, et al. Misleading presentations in functional food trials led by contract research organizations were frequently observed in Japan: Meta-epidemiological study. *J Clin Epidemiol* 2024; 169: 111302.
 - 12) Kamioka H, Kitayuguchi J, Origasa H, Tsutani K. Research quality of clinical trials reported for Foods with Function Claims in Japan, 2023–2024: Evaluation based on a revised tool to assess risk of bias in randomized trials. *Nutrients* 2024, 16, 2744.
 - 13) 上岡洋晴, 津谷喜一郎, 折笠秀樹. 機能性表示食品制度におけるシステマティック・レビューの質の再検証: PRISMA 2020 の導入を目前にして. *薬理と治療* 2023;51:21-30.
 - 14) 消費者庁. 機能性表示食品制度における機能性に関する科学的根拠の検証: 届け出られた研究レビューの検証事業. https://warp.da.ndl.go.jp/info:ndljp/pid/11257153/www.caa.go.jp/policies/policy/food_labeling/about_foods_with_function_claims/pdf/about_food_with_function_report_180416_0001.pdf
 - 15) 上岡洋晴, 津谷喜一郎, 折笠秀樹, 大室弘美, 島田美樹子, 北湯口純ほか. 機能性表示食品制度における届出されたシステマティック・レビューの質: PRISMA 声明チェックリストに基づく前後比較研究. *薬理と治療* 2019;47:357-73.
 - 16) 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 Research Methodol* 2007;15;7:1-7.
 - 17) 折笠秀樹, 上岡洋晴, 津谷喜一郎. 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・津谷喜一郎編著. *臨床試験と疫学研究のための国際ルール集 (Part 2)*. 東京: ライフサイエンス出版;2016.p.148-55.
 - 18) Kamioka H, Tsutani K, Origasa H, Yoshizaki T, Tang W, Shimada M, et al. Quality of systematic reviews of the Foods with Function Claims registered at the Consumer Affairs Agency website in Japan: a prospective systematic review. *Nutrition Research* 2017;40:21-31.
 - 19) Kamioka H, Tsutani K, Origasa H, Yoshizaki T, Shimada M, Kitayuguchi J, et al. Quality of systematic reviews of the Foods with Function Claims in Japan: Comparative before-and after-evaluation of verification reports by the Consumer Affairs Agency. *Nutrients*. 2019;11:1583.
 - 20) Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. 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.
 - 21) 上岡洋晴, 折笠秀樹, 津谷喜一郎. 「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. *薬理と治療* 2018;46:1785-96.
 - 22) Kamioka H, Tsutani K, Origasa H, Yoshizaki T, Shimada M, Kitayuguchi J, et al. A cross-sectional study based on forty systematic reviews of Foods with Function Claims (FFC) in Japan: Quality assessment using AMSTAR 2. *Nutrients* 2023; 15: 2047.
 - 23) Rethlefsen ML, Kirtley S, Waffenschmidt S, Ayala AP, Moher D, Page MJ, et al. PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews. *Syst Rev* 2021;10:39
 - 24) 上岡洋晴, 眞喜志まり, 佐山暁子, 津谷喜一郎, 折笠秀樹. 「PRISMA -S: システマティック・レビュー報告のための PRISMA 声明拡張版」の解説と日本語訳. *薬理と治療* 2021;49:1057-79.
 - 25) Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *BMJ* 2021; 372:n160.
 - 26) 眞喜志まり, 佐山暁子, FFC-SR2 研究グループ. 機能性表示食品制度における届出されたシステマティック・レビューの質: 情報源と検索における現状と課題. *薬理と治療* 2019;47:711-4.
 - 27) 消費者庁. (別紙4) システマティック・レビューの実施手順に係る考え方 (例). https://www.caa.go.jp/policies/policy/food_labeling/foods_with_function_claims/assets/foods_with_function_claims_230929_0010.pdf