

Series

EQUATOR Network から得られる、
質の高い研究報告のための国際ルール —— ⑤

MOOSE：観察研究のメタ解析報告のためのルール

安達 絵美* 中山 健夫*

メタアナリシスとは

メタアナリシスは、複数の臨床研究結果を統合し、統計的手法を用いて、データを量的に統合するものである。医学領域でのメタアナリシスの歴史¹⁾は、1904年の戦時中に腸チフスに対するワクチンの効果を検討したのに端を発する。そのころから、データを統合することの必要性や方法論が検討され始め、1976年にメタアナリシスという言葉が初めて提唱された。1992年にコクラン共同計画が開始され、ランダム化比較試験(randomized controlled trial: RCT)を中心とした臨床試験のシステマティック・レビュー、そしてメタアナリシスが推進されてきた。システマティック・レビューとメタアナリシスの違いは、前者が「明確に作られた疑問に対し、系統的で明示的な方法を用いて、適切な研究を同定、選択、評価を行うことで作成するレビュー」であるのに対して、後者が「複数の研究による結果の定量的な統合」を意味する。従来、メタアナリシスの対象とするのはRCTの研究結果のみとすることが望ましいとされてきたが、近年、観察研究のメタアナリシスの報告も増加傾向にある。

1991年にカナダのGuyattが提唱した根拠に基づく医療(evidence-based medicine: EBM²⁾)の大きな意義のひとつは、疫学的な研究デザインに応じて、因果関係を推論する妥当性に強弱があること、それ

表 1 レベル・オブ・エビデンス (治療の有効性に関して)

I	システマティック・レビュー/メタアナリシス (データ統合型研究)
II	1つ以上のランダム化比較試験による
III	非ランダム化比較試験による
IV	分析疫学的研究 (コホート研究・症例対照研究)
V	記述研究 (症例報告やケース・シリーズ) による
VI	患者データに基づかない、専門委員会や専門家個人の意見

をエビデンス・レベルとして、多くの臨床家にもわかりやすいかたちで提示したことにある²⁾。英国オックスフォード大学EBMセンターでは、エビデンス・レベルの高い順にシステマティック・レビューまたはメタアナリシス、RCT、非ランダム化比較試験、そして観察研究であるコホート研究、症例・対照研究が続く、その下に対照群のない症例集積や専門家の個人的経験に基づいた意見、エビデンスが明示されていない専門委員会の見解としている(表1)。どのような研究デザインが適切かは、研究的疑問(リサーチ・クエスチョン)のカテゴリーによって異なり(表2)、それぞれのクエスチョンに対して望ましい疫学的な臨床研究のデザインと、それらの研究の特性を考慮したエビデンス・レベルが付与されている。介入の有効性が問題となる治療や予防のクエスチョンに関してはRCTのシステマ

* 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

表 2 クエスチョンのカテゴリーと代表的な研究方法

クエスチョン	研究方法
●頻度	横断研究 (有病割合), コホート研究 (罹患率)
●原因・リスク ファクター	コホート研究, 症例対照研究
●診断	比較研究 (横断研究), 検査特性分析
●予後	コホート研究
●治療	介入研究
●コスト	費用効果分析など
●不確定状況下で の意思決定	決断分析

※いずれの研究でも知見の統合はメタアナリシス

ティック・レビューが最上位となり、予後や診断、頻度に関するクエスチョンに関しては、コホート研究などの観察研究のシステマティック・レビューが最上位とされている³⁾。

観察研究を対象としたメタアナリシスには賛否両論ある。観察研究と RCT の両方に出版バイアスや不均一性 (異質性) の問題はあるが、観察研究の重大な限界として「ランダムに割付ができない」ための選択バイアスの影響が accrue に指摘される。未測定の変数因子により、十分な交絡の調整は不可能であり、それらを制御できない研究で推定された個々の推定値にはバイアスに基づく「系統的な誤差 (systematic error)」が不可避である。にもかかわらず、メタアナリシスを行うと、個々の研究に内在するそのような誤差が (読者の) 視野から消えて、誤差のない統合値が得られたかのような印象を与える点は批判の対象とされている^{4,5)}。しかし、その一方で、優れたデザインの観察研究であれば、観察研究でも有効性を過大評価せずに、RCT と類似した結果が得られるとする論文もいくつか発表された^{6~8)}。治療の有効性評価における観察研究について、統計家は RCT と比して劣ると厳しくとらえ、臨床研究者は臨床現場のリアルワールドの中で有用性を再評価しようとする傾向がみられる^{9,10)}。

観察研究の重要性

すべての研究的疑問を RCT で検証することは現実的には不可能である。ランダム配置に伴う倫理的

な制約、厳密な組み入れ・除外基準によって選ばれた対象者での限定的な知見であるという一般化可能性の問題や、費用対効果、適時性の問題がある。介入の有効性は、限定的な条件下で効果を発揮するものであり、実際の臨床現場やリアルワールドでの有効性は、より多様な集団を対象とした観察研究での評価のほうが適切である場合も少なくない。介入的手法が適用できない多様な臨床的・公衆衛生的課題の解決に向けて、観察研究、そして、そのメタアナリシスは現実的で有用な手法のひとつであり、これらの研究デザインでしか答えられない疑問があるかぎり、今後も医学におけるそれらの重要性が減じることはないであろう。

事実、この 20 年間で、欧米における医療業務データやレセプトデータ、患者レジストリなど、膨大な人数の縦断的データを含んだデータベースを用いた観察研究の報告は増加している。それらの研究の多くは、RCT で検証された介入や治療の有効性や安全性がリアルワールドではどうなっているか、という疑問に答えることを目的に実施されてきた。記憶に新しいものでは、2006 年に発表された医療業務データベースを用いた研究¹¹⁾で、高脂血症治療薬であるセリバスタチンナトリウム製剤 (バイコール錠®, セルタ錠®) の服用患者において、副作用の横紋筋融解症による入院のリスクが他の治療薬と比べ 2 倍以上も上昇したことが報告され、セリバスタチンナトリウム製剤は世界的に販売を中止した。有効性を評価した RCT では明らかになっていなかった薬剤の有害事象が、観察研究によってそのリスクの上昇が示され、市場からの撤退を余儀なくされたのである。

日本でも医科レセプトや調剤レセプト用いた研究の発表が増えてきている。2009 年に施行された「高齢者の医療の確保に関する法律」第 16 条の規定に基づき、2010 年からレセプト情報や特定健診・保健指導情報が収集され、2014 年 2 月末時点で、レセプト情報は約 72 億件、特定健診等情報は約 9000 万件が蓄積されている¹²⁾。レセプト情報・特定健診等情報データベース (National Database: NDB) の第三者提供が 2012 年度から運用が開始された。今後、この大規模なデータを用いた観察研究の発表や、複数知見の統合結果の発表が増えることを期待される。

表 3 MOOSE チェックリスト

提案された報告用チェックリスト—観察研究のメタアナリシス著者, 編集者, 査読者用

背景の報告には以下の事項が含まれるべきである

- ・問題の定義
- ・仮説の設定
- ・研究のアウトカムの記述
- ・使用された曝露 exposure もしくは介入 intervention
- ・使用された研究デザイン
- ・研究対象集団 study population

検索ストラテジーの報告には以下の事項が含まれるべきである

- ・検索担当者の適格性 (例: 図書館員および研究者)
- ・検索ストラテジー: 統合 synthesis および検索用語 keyword に含める期間も含めて
- ・すべての入手可能な研究を盛り込む努力—著者との連絡も含む
- ・調査対象となったすべてのデータベースおよび記録
- ・調査のために使用されたソフトウェアの名前とバージョン—特別機能 (explosion など) も含む
- ・ハンドサーチの使用について (例: 入手した論文の参考文献リスト)
- ・特定された文献のリストおよび除外された文献のリスト—除外の妥当性に関する理由も含めて
- ・英語以外で発表された文献の取り扱い
- ・抄録および未発表の研究の取り扱い方法
- ・いかなる形式であれ著者と連絡を取った事実があれば, それについての説明

方法の報告については以下の事項が含まれるべきである

- ・検証すべき仮説を評価するために集められた研究の関連性 relevance および適切さ appropriateness についての記述
- ・データの選択およびコード化に関する論拠 rationale (例: 確固とした臨床的原則, または臨床上の便宜)
- ・データの分類とコード化の方法に関する詳細な記述 (例: 複数評価者, 盲検 blinding, 評価者間信頼性)
- ・交絡因子 confounding の評価 (例: 適切な設定であれば, 研究の症例と対照の比較可能性)
- ・研究の質の評価者に対する盲検, 研究結果の予測因子の層別化 stratification または回帰 regression を含む
- ・不均一性 heterogeneity の評価
- ・統計学的手法について, その再現を可能にする詳細な記述 (例: 固定効果モデルまたはランダム効果モデルについての詳細な記述, 選択されたモデルが研究結果の予測因子となりうるものかどうかについての正当な理由, 用量反応モデル dose-response model, 累積メタアナリシス cumulative meta-analysis)
- ・適切な表およびグラフの提示

結果の報告には以下の事項が含まれるべきである

- ・個々の研究における推定値 estimates および研究全体としての推定値をまとめたグラフ
- ・報告に含まれる各研究についての記述的情報を盛り込んだ表
- ・感度分析の結果 (例: サブグループ解析)
- ・治験における統計学的不確実性 statistical uncertainty の指摘

考察の報告には以下の項目が含まれるべきである

- ・バイアスの量的評価 (例: 出版バイアス)
- ・除外の妥当性に関する説明 (例: 非英語引用文献の除外)
- ・解析に含まれたすべての研究の質的評価

結論の報告には以下の事項が含まれるべきである

- ・観察結果に対する他の説明の考慮
- ・結論一般化 generalization (提示されたデータに対して適切で, また文献レビューの範囲内での一般化)
- ・これからの研究に向けての指針
- ・財源の公表

本稿は, Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA 2000; 283: 2008-12. について, 安達絵美らが解説したものである。

チェックリストの日本語訳は、『臨床研究と疫学研究のための国際ルール集』(中山健夫, 津谷喜一郎編著, ライフサイエンス出版, 2008, p221) からの転載である。チェックリストの掲載にあたって, American Medical Association より許可を得た。

Copyright © 2000 American Medical Association. All rights reserved.

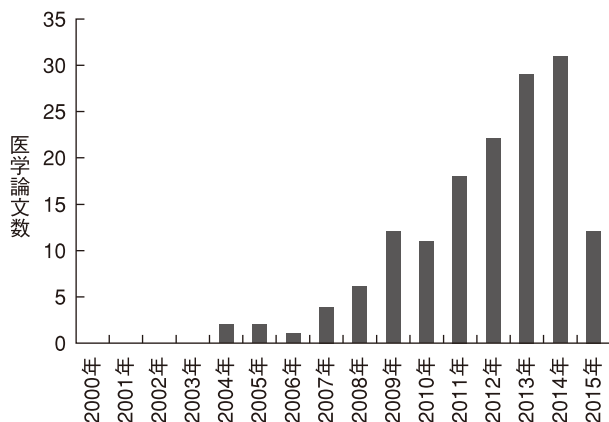


図 1 MOOSE ガイドラインを用いたメタアナリシス

観察研究のメタアナリシス報告に関する執筆ガイドライン：MOOSE 提案

本稿で紹介する MOOSE 提案¹³⁾とは観察研究のメタアナリシスの報告に関する執筆ガイドラインである。1997 年 4 月に米国ジョージア州アトランタで開催されたワークショップにおいて、35 項目に及ぶチェックリスト (表 3) が、論文著者や査読者、編集者、そして読者が利用する目的で作成され、2000 年に MOOSE (Meta-analysis Of Observational Studies in Epidemiology) 提案として発表された。それ以前にも、メタアナリシスの質を向上させるためのガイドラインはいくつか存在していたが、それらは RCT のメタアナリシスに限定したもので、メタアナリシスを行う研究者の視点のみで作成されていた。

MOOSE 提案において観察研究は、「既存のデータベースの活用、横断研究 cross-sectional study、症例集積 case series、ケース・コントロール研究 case-control study、歴史 (過去起点) コホート研究 historical cohort、コホート研究 cohort study の研究デザインによる、病院または有効性に関する研究」と定義されている。MOOSE 提案の作成には、臨床医、臨床試験、統計学、疫学、社会科学、および生物医学雑誌編集など各領域の専門家 27 人が参加した。作成グループは、観察研究のメタアナリシスに関するシステマティック・レビューを実施し (文献データベースとして MEDLINE, Educational Resources Information Center [ERIC], PsycLIT, Current Index to Statistics を利用)、得られた 32 論文のリストに基

づいて各領域の専門家の意見を聴取した。参加者はバイアス、検索と要約、異質性、研究の分類、および統計学的手法について話し合い、推奨事項のチェックリストが作成された。

2000 年 4 月に JAMA より MOOSE 提案が発表されてから 15 年が経つ。この 15 年間で新たに報告された観察研究のメタアナリシスは 3500 件以上に及ぶが、そのなかで MOOSE ガイドラインに基づいて発表されたものは計 132 件あった (5 月 20 日時点, PubMed にて検索)。実際に、初めて MOOSE ガイドラインを用いたメタアナリシスが発表されたのは 2004 年で、普及するのに 4 年かかったが、それ以降は年々増加傾向にある (図 1)。

とはいえ、MOOSE を用いた観察研究のメタアナリシスが 132 件というのは、いささか少ないようにも感じられる。今日、研究報告の質の改善を目的に作成されたガイドラインは MOOSE のほかにも、RCT のための CONSORT、観察研究のための STROBE、システマティック・レビューのための PRISMA (旧称 QUORUM)、診断の精度に関する研究のための STARD、診療ガイドラインのための AGREE、公衆衛生的観察研究のための TREND など、100 以上存在する。これらのガイドラインの遵守は、だれに、どの範囲まで求められているのだろうか。

もちろん、著者、とくに経験の浅い若手研究者にとっては、研究報告の欠落を防ぎ、より明確な論文作成にガイドラインの利用が有用であることはいくまでもない。明確に提示された研究結果は容易に解釈でき、利用しやすい。患者に研究参加の同意を得て、資金を投じて研究していることを考えれば、より正確な報告をすることは、研究者の社会的責任である。

求められる各種執筆ガイドラインの使用

一方、学術雑誌の編集者や査読者は各種ガイドラインをすべて確認するのは、実質的な負担や統計学的解釈の観点から容易ではない¹⁴⁾。しかし現在は、編集者ごとにガイドラインの使用に対して若干異なる立場を取りつつも、多くの医学系雑誌で著者にガイドラインの使用を勧める傾向にあるといえる。観

観察研究のメタアナリシスに関しては、たとえば、*the Annals of Internal Medicine* 誌は、論文著者に推奨されている研究報告のガイドラインやチェックリストを使って原稿を準備するよう使用を勧めている¹⁵⁾。

BMJ 誌は EQUATOR (Enhancing the Quality and Transparency of Health Research) Network から該当する研究報告ガイドラインを用いることを推奨している¹⁶⁾。その説明として、ガイドラインを使用することによって、「著編集者や査読者、読者に、研究がどのように行われたかを理解させ、その結果が信頼できるか判断するのに十分な情報を提供したことを保証することになる」と記載されている。さらに、JAMA 誌は、投稿に際して MOOSE を含む各種執筆ガイドラインのチェックリストの提出を求めている¹⁷⁾。ガイドラインの使用は、あくまで、著者の明確な研究の報告を目的としており、研究そのものの質や妥当性を評価することを目的とはされていない。

今後、観察研究のメタアナリシスの質と完成度を高める手がかりとして、MOOSE のガイドラインがより広く認知され、活用されることが望まれる。より数多くの観察研究を統合させるためにも、一次研究である観察研究の報告の質の向上も不可欠であり、観察研究の報告に関する執筆ガイドラインである STROBE 声明の普及も強く期待したい。

文 献

- 1) 丹後俊郎. メタアナリシス入門: エビデンスの統合をめざす統計手法. 朝倉書店, 2002.
- 2) Guyatt G. Evidence-based medicine. ACP Journal Club 1991; 114: A-16.
- 3) 中山健夫. EBM を用いた診療ガイドライン: 作成・活用ガイド. 金原出版, 2004.
- 4) Shapiro S. Meta-analysis/shmeta-analysis. Am J Epidemiol 1994; 140: 771-8.
- 5) Greenland S. Can meta-analysis be salvaged? Am J Epidemiol 1994; 140: 783-7.
- 6) Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. N Engl J Med 2000; 342: 1887-92.
- 7) Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. N Engl J Med 2000; 342: 1878-86.
- 8) Shikata S, Nakayama T, Noguchi Y, et al. Comparison of effects in randomized controlled trials with observational studies in digestive surgery. Ann Surg 2006; 244: 668-76.
- 9) Pocock SJ, Elbourne DR. Randomized trials or observational tribulations? N Engl J Med 2000 22; 342: 1907-9.
- 10) Ioannidis JP, Haidich AB, Pappa M, et al. Comparison of evidence of treatment effects in randomized and nonrandomized studies. JAMA 2001; 286: 821-30.
- 11) Cziraky MJ, Willey VJ, McKenney JM, et al. Statin safety: an assessment using an administrative claims database. Am J Cardiol 2006; 97 (8A): 61C-68C.
- 12) 厚生労働省. レセプト情報・特定健診等情報データの利活用の促進に係る中間とりまとめ.
<http://www.mhlw.go.jp/stf/shingi/0000041524.html>
- 13) Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA 2000; 283: 2008-12.
- 14) Vandembroucke JP. STREGA, STROBE, STARD, SQUIRE, MOOSE, PRISMA, GNOSIS, TREND, ORION, COREQ, QUOROM, REMARK... and CONSORT: for whom does the guideline toll? J Clin Epidemiol 2009; 62: 594-6.
- 15) <http://annals.org/SS/authorsinforeviewssystematicand-metaanalyses.aspx>, Accessed May 26, 2015.
- 16) <http://journals.bmj.com/site/authors/editorial-policies.xhtml#reporting>, Accessed May 26, 2015.
- 17) <http://jama.jamanetwork.com/public/instructionsForAuthors.aspx#ManuscriptPreparationandSubmissionRequirements>, Accessed May 26, 2015.