
学内活動報告

順天堂大学医療看護学部 医療看護研究20
P.63-67(2017)

第7回医療看護学研究科FD講演会 システマティックレビューの基礎 —リスクオブバイアスとメタアナリシス—

Basis of Systematic Reviews — Risk of Bias and Meta-analysis —

大 田 えりか*
OTA Erika

1. システマティック (系統的) レビューとは？

システマティックレビューとは、ある課題に関して行われた研究を網羅的に検索し、そのバイアスを系統的に吟味し、その結果に応じて統計的に統合する手法である。Reviewには、大きくNarrative Reviewの中にSystematic ReviewとMeta-analysisが、一部重なり合っており含まれている。つまり、システマティックレビューは、全世界の論文たちを網羅的に集めて統合して、より良いエビデンスを導くために実施する方法である。

なぜ系統的レビューか？ 例えば、ある新しい薬Aと今まで使用されていた標準薬があり、どちらを推奨すべきかという問題があるとする。高血圧の治療は、低リスク患者に新薬Aは、標準薬に対して有効か？という問いに対して、検索された既存のランダム化比較試験の結果を比較する。個別の論文を評価すると、新薬Aを強く推奨する説明では、“新薬Aは、Otaの論文より脳卒中防止作用があり、Kouseiの論文より血圧低下作用もある。さらに、Toukyoの論文と突然死増加は同じである。”一方、標準薬を強く推奨する説明では、“標準薬は、Yuasaの論文などより脳卒中防止作用があり、Kitaの論文より血圧低下作用もある。さらに、Yuasaの論文より突然死増加が少ない。”

という結果となった。どちらが有効であるかの結果として、標準薬の方が新薬Aよりも有効であることが理解できる。したがって、論文単位でアウトカムを選んで参考文献を挙げるのは問題であり、論文単位でエビデンスにしないためには、論文を並べて、設定したアウトカム毎に論文の結果を串刺しのように読む必要がある。

2. 系統的レビューのプロセス

1) 疑問の定式化 (PICOs)

疑問を下記の項目に基づいて、定式化していく。

- Participants：どんな対象が
- Intervention：どんな治療、検査をするのか
- Comparison：どんな治療、検査と比べて
- Outcome：どうなるか
- Study design：研究デザイン

リサーチクエストionsとして、“妊娠中の女性への亜鉛のサプリメントの摂取は、母子の予後を改善するのか？”とした場合は、研究デザインを分かりやすくする方法として、上記の項目に準じて定式化する。

(1) P：Participants /Population

— 研究対象は？ 妊娠した女性全員？ ハイリスクの女性も入れる？ 妊娠何週から？

(2) I：Intervention/Exposure

— どのような介入について研究するのか？

— 亜鉛のサプリメントはどのくらいの量をいつ摂取するのか？

* 聖路加国際大学大学院看護学研究科・コクラン日本支部
St. Luke's International University Graduate School of Nursing
Science, Cochrane Japan

- (3) C : Comparison/Control
 - 何と比較する？
 - プラセボ？ 他のサプリメント併用の場合はどうする？
- (4) O : Outcome
 - どのような結果を見るのか？

これらを吟味して、妊婦さんへの亜鉛のサプリメント介入研究を定式化する。

- P (対象)
 - 全ての妊娠女性を対象、ただしHIVや栄養不良のハイリスクへの研究は除く
- I (介入)
 - 亜鉛サプリメント
- C (対照)
 - プラセボまたは、サプリメントなし
- O (結果)
 - 早産、SGA、低出生体重児、周産期死亡、新生児死亡、死産、出生体重
 - 妊娠中の出血、産後の出血、妊娠性高血圧、過期産、前期破水など

Ota E, Mori R, Middleton P, Tobe-Gai R, Mahomed K, Miyazaki C, Bhutta ZA. Zinc supplementation for improving pregnancy and infant outcome. Cochrane Database of Systematic Reviews 2015, Issue 2. Art. No. : CD000230. DOI : 10.1002/14651858.CD000230.pub5.

次に、優先的アウトカムとして下記について考慮する。

- 主要アウトカム (最大3項目)
 - 通常起こりうる有害作用を1つ以上含める
- 副次的アウトカム
 - 残りの主なアウトカム
 - 優先性が低い追加的なアウトカム
- 主なアウトカム (最大7項目)
 - 意思決定に必須となるアウトカム
 - 解析と要約に基づいてまとめる

アウトカムは、真のアウトカム 'True outcome' と代用のアウトカム 'Surrogate outcome' の2種類ある。前者は、患者にとって重要なアウトカムであり、その評価には長い時間や費用を要する場合も多い。例えば、死亡率などがその例である。後者は、研究では用いられることが多いものの、あくまでも真のアウトカムの代用であり、直接の効果までは説明できない。

選ぶ研究デザインについて、疑問のための最も適切

なデザインを選択することは言うまでもない。常に、研究デザインの選択に対する合理的理由を示す必要がある。コクランレビューの大部分は、ランダム化比較試験が採用されている。その理由は、グループ間の系統的相違を防止するためである。時には、比較臨床試験 (準ランダム化試験) も採用されている。また、知りたいカテゴリーと研究デザインは概ね決まっており、頻度を知りたい場合は横断研究であり、リスクを知りたい場合はコホート研究と症例対照研究であり、治療・予防を知りたい場合は、臨床試験であるランダム化比較試験である。

系統的レビューで考慮すべき4種類のランダム化比較試験の研究デザインには、対象者一人一人をランダム化する Individual RCT (ランダム化比較試験)、施設や村、学校などのクラスター単位をランダム化する Cluster RCT (クラスターランダム化比較試験)、来た順番に交互に、または生年月日など予測が可能なように分ける方法である Quasi RCT (準ランダム化比較試験)、介入群を実施した後に一定期間を置いてコントロール群になる Cross-over trial がある。

2) 網羅的検索

文献検索の基本戦略として、population/participants、design、Interventionの3要素を主として検索する。そして、データベースシステムを用いて検索するのは当たり前のことであるが、その理由として、重要な発見を見落とさない、参考にする文献が偏らない、検索の効率を高めることである。医療系の三主要データベースシステムは下記の通りである。

(1) Medline

- 米国発、PubMedを通して無料公開、包括的、米国・先進国に偏りがある。Medlineでは、文献を管理する際に文献の内容を表す適切な用語を文献に付与し、この用語により文献を検索・管理できるため、米国国立医学図書館 (NLM) が定める生命科学用語集 (シソーラス) である MeSH (Medical Subject Headings) に基づく。

(2) Cochrane Library

- 英国発、RCT・SRなど質の高い研究を重視し、英国・豪州・ブラジル・途上国のみ無料公開しており、同様に MeSH に基づく。

(3) EMBASE

- 欧州発、薬学にも強く、日本では有料の図書館が多い。

これらの他には、国内の医学中央雑誌、コメディカルのデータベースとしてCINAHL、看護系のみBritish Nursing Index、精神・心理系のPsycINFO、国際保健を主としたAfrican Medical Indexがある。これらを活用して文献を網羅的に検索し、それぞれのデータベースの特徴を理解し、検索内容によってデータベースシステムを取捨選択する。検索の方法として、キーワード毎に、検索を拡大するOR、検索を狭小化するANDのいずれも駆使して得られた用語をまとめる。研究論文の選択は、タイトルとアブストラクトを2名で独立してスクリーニングする。クライテリアに該当する研究論文はフルテキストを入手して含めるか除外するか検討する。この際に、各研究論文のフローを表したプリズムチャートと除外理由を記載する必要がある。スクリーニングに有用なソフトとして、EndNote、EPPI-Reviewer 4、Abstrckr、そしてお薦めサイトとして、COVIDENCE (<http://www.covidence.org/>)がある。

3) RCTの批判的吟味

コクランによる‘Risk of bias’アセスメントには、エビデンスに基づく下記7つのドメインがある。また、レビューの著者による判断として、Low risk of bias (低バイアスリスク)、High risk of bias (高バイアスリスク)、リスクの程度不明のUnclearがある。リスク判断をサポートするために、研究論文または他の資料からのエビデンスを引用したり、著者の説明をレビューしたりする。

[対処すべき7つのドメイン]

(1) ランダム系列の生成

対象の割り付けがランダム化されているかどうかを吟味するため、低リスクの場合は予測不可能であり、手段としては、乱数表、PCによる乱数生成器の活用、階層ランダム化、さらに低い手段としてコイントス、カードシャッフル、封筒法、サイコロ投げ、くじ引きなどがある。

高リスクの場合は予測可能であり、主に2種類の方法がある。最初に、生年月日、受診日、IDまたは記録番号などにより交互割り付けする準ランダム化がある。次に、医師または参加者の選択、検査結果など入手可能性による非ランダム化がある。

(2) 割り付けの隠蔽化

このドメインは、臨床試験の開始時に参加者の割り付け中に発生し、選択バイアスを避けるために行われる。研究への採用時に参加者自身が割り付けられる群を誰も予測できない。そして、順序の変更を防止したり、誰を採用するかに関する選択を防止するなどランダム系列の厳格な実行を保証するためのドメインである。

(3) 参加者と医療提供者のブラインド化

このドメインは、介入群と異なる治療や異なる参加者の期待、実際のアウトカムにおける変化を招くなどの施行バイアスを避けるためにブラインド化する。このブラインド化は慎重に査定し、ブラインド化が破られる可能性があるのかどうかを吟味し、ブラインド化自体がこの介入に適していない場合であってもその影響を考慮する必要がある。

ブラインド化がなされている、もしくはブラインド化が破られた可能性が低い場合、またブラインド化されていないか、あるいは不完全であるが、アウトカムがその影響を受ける可能性が低いと判断された場合は、低リスクと評価する。一方で、ブラインド化がなされていないか、もしくは不完全であり、アウトカムがその影響を受ける可能性が高い場合に高リスクと評価する。

(4) アウトカムアセスメントのブラインド化

これは参加者が受療する介入法に関する知識を有していることによって影響されるアウトカムの測定において、検出バイアスを避けるために評価されるドメインである。施行バイアスと同様に、慎重に査定されなければならない。ブラインド化が破られる可能性があるかどうかを吟味し、参加者や医療提供者がブラインド化されていない場合であっても、また参加者や医療提供者もアウトカムの評価者である可能性について評価する。

検討した結果、ブラインド化がなされており、ブラインド化が破られた可能性が低い場合、およびブラインド化がなされていないが、測定が影響を受ける可能性が低い場合は、低リスクと評価する。そして、ブラインド化がなされていない、もしくはブラインド化が破られている、および測定が影響を受ける可能性が高い場合は、高リスクと評価する。

(5) 不完全なアウトカムデータ

参加者全ての完全なアウトカムデータを利用することができない場合は、損耗バイアスを招く危険性がある。

る。具体的には、ドロップアウトがあり、フォローアップの失敗や測定の中止、その他のデータの欠損などが含まれる。また、報告に利用可能なデータの一部が含まれないといった除外が事実として存在する。

この場合に考慮すべき事項として、各グループからどの程度の数値を含むデータが失われたのか、なぜ失われたのか、そもそもデータはどのように解析されたのか、入力ミスなどを吟味して評価する。

このドメインでは、欠失データが無い、アウトカムと無関係なデータ欠失の理由が記載されている、欠失データがグループ間でバランスが取れており、その理由も類似する場合、そして欠失比率や妥当な効果サイズが臨床的に重要な効果を持つには十分ではない場合は、低リスクと評価する。なお、大部分の研究論文が判定されうる可能性が高い評価として、リスクの程度が不明なUnclearがある。一方で、欠失理由がアウトカムに関連し、欠失数またはその理由が群間で異なる場合、欠失比率や妥当な効果サイズが十分に臨床的に重要な効果を持つ場合、そして‘as-treated’解析の各群の症例数が当初の割り付けと大きく異なる場合は、高リスクと評価する。

(6) 選択的報告

このドメインでは、報告バイアスを招く危険性があることを評価する。その特徴として、狙い通りにかつ詳細に統計的に有意な結果が報告されやすいことである。比較的判定が難しいため、次の事項を吟味する。方法と結果を比較して、データの欠失や追加、サブセット、予期されていない測定項目や方法など事前に設定されているように、あるいは予想されたようには報告されていないアウトカム、さらに報告が不完全であり、メタ解析には加えられないアウトカムと判断された場合は、高リスクと評価する。

(7) その他のバイアス

ある要因がなぜバイアスの原因となるのかについては、明確な合理的理由が必要である。その理由には、サンプルサイズが小さいなどの不正確性、多様性、質に関するその他の測定指標は含まれない。可能であれば、プロトコルで重要な問題を特定する。

その他のバイアスの起源が無いと思われる研究は、低リスクと判断する。例えば、交差試験におけるキャリーオーバーやクラスター RCT における採用バイアス、非ランダム化試験などのように明らかに研究デザインに特有の問題がある場合、ベースラインデータの不均衡、非ブラインド化臨床試験におけるブロックラ

ンダム化、鑑別診断活動、その他のバイアスがあると判断する場合は、高リスクと評価する。

4) メタ解析

(1) メタ解析を行う理由

メタ解析を行う理由は、治療効果とその不確実性を定量化し、検出力と精度を高め、研究間の相違を調べ、矛盾する研究の議論を解決し、新たな仮説を生み出すためである。改めて、二群を比較する次の指標を理解する必要がある。効果測定指標には次の3つがあり、リスク比 (RR; 相対リスク)、オッズ比 (OR)、リスク差 (RD; 絶対リスク減少) である。全ての推定値は不確実であり、信頼区間と共に提示する必要がある。

[リスク比]

$$\text{リスク比} = \frac{\text{介入群のリスク}}{\text{対照群のリスク}}$$

リスク比 = 1 の場合は群間に差は無いことを示している。具体的に記述する場合は、例えばリスク比が 1.79 だった場合は「介入が頭痛のリスクを 79% 増加させた。」と表す。リスク比が 0.79 だった場合は「介入は頭痛のリスクを 21% 減少させた。」と表す。

[オッズ比]

$$\text{オッズ比} = \frac{\text{介入群のオッズ}}{\text{対照群のオッズ}}$$

オッズ比 = 1.0 の場合は群間に差が無いことを示している。例えばオッズ比が 2.06 だった場合は「介入は頭痛のオッズを 2 倍にする。」もしくは、「介入は頭痛のオッズを 106% 増加させた。」と表す。オッズ比が 0.06 だった場合は「介入は頭痛のオッズを対照群の 6% に減少させた。」もしくは「介入は頭痛のオッズを 94% 減少させた。」と表す。

[リスク差]

$$\text{リスク差} = \text{介入群のリスク} - \text{対照群のリスク}$$

リスク差 = 0 の場合は、群間に差が無いことを示している。リスク差が 0.11 だった場合は「介入は頭痛のリスクを 11 ポイント増加させた。」、「カフェイン群では頭痛を経験した人は対照群より 100 人中 11 人多かった。」と表す。リスク差が -0.11 だった場合「介入は頭痛のリスクを 11 ポイント減少させた。」、「カフェイン群で頭痛を経験した人は対照群より 100 人中 11 人少なかった。」と表す。

5) GRADE : エビデンスの質の評価

研究結果の信頼性を評価するために、アウトカム毎に複数の論文を集めてエビデンスの質を評価する。総体エビデンス (Quality of a body of evidence) の質は下記の通りである。

総体エビデンスの質	
高	high
中	moderate
低	low
非常に低	Very low

最後に、系統的レビューのはじめ方として、コクラン系統的レビューのタイトルを登録する。コクラン以外のジャーナルに投稿する場合はPROSPEROに登録する (<http://www.crd.york.ac.uk/PROSPERO/>)。