

因果関係のためのGRADEアプローチと Bradford Hill 基準



相原守夫¹

(記 平成 25 年 6 月 25 日)

RCT や観察研究(コホート研究、症例対照研究)によって曝露とアウトカムとの関連(association)があることが確かめられたらば、次の問題は、観察された関連がはたして因果関係(causal relationship)がどうかということである。本稿では、特に観察研究からのエビデンスの質を評価する上で、研究デザインや因果関係について GRADE アプローチをどのように適用させるかについて Bradford Hill 基準との関連も含めて解説する。

1. 観察研究のデザイン

観察研究の質を評価するとき、前向き研究と後ろ向き研究を区別することはあまり意味がない。なぜならば、サンプル抽出にバイアスがないこと、患者の特性やアウトカムに関する質の高い測定が実施されていること、そして追跡が完全に行われているという質基準が、研究が前向きか後ろ向きかにかかわらず重要だからである。これらの質基準は、どちらかという後ろ向き研究よりは前向き研究が満たす傾向があるが、前向き研究が質基準を満たしていない場合や、後ろ向き研究が質基準を満たしている場合もある。したがって、前向きか後ろ向きかの相違よりも、内的対照(internal control)もしくは内部比較(internal comparison)が存在するか、存在するならばその内容はどうか問題である。内部対照を伴わない研究は、次の基準を満たさない限りは、症例集積とみなされる。(1) データ収集の開始前に研究のプロトコルが確立されている、(2) 組み込み基準と除外基準が定義されている、(3) 実際の研究で除外された患者の数が報告されている、(4) 実際の研究で追跡スケジュール、予期されるアウトカムの調査およびアウトカムの定義のために使用した基準に関する説明を含む標準的な観察が実施されている、(5) 追跡からの脱落がすべて報告されている(文献1: 診療ガイドラインのためのGRADEシステム p42より)。

2. 因果関係

一方が原因で他方が結果というつながりを因果関係という。1965 年、英国の統計学者でロンドン大学教授の Sir Austin Bradford Hill は、ある疾患と環境因子の関係を決定する際、それが因果関係による相関なのか単なる相関なのかを考えるためのいくつかの特性を提案した[2]。因果関係の推定(因果推論)において、Bradford Hill が提案した9つの項目をすべて満足する必要はなく、各項目の重要度がすべて等しいというわけでもない。Bradford Hill の基準は修正されつつ現在でも広く用いられているが、因果関係を評価するためのこの基準を GRADE の枠組みで考慮することが要請されている[4]。

GRADE working group は、制定後半世紀を経た今なお Bradford Hill の基準は因果関係に対する確信性を左右する関連要因であると認識している。これらの基準の一部は、因果関係について判断を下す前の段階における効果推定値が正しいという、あるいは曝露とアウトカムの関連が信頼できるというわれわれの確信を左右するものである。しかし、因果関係の立証または要約を行い、介入に踏み切るまでの間に、予期される因果関係に影響を及ぼすいくつかの段階があることに注意されたい。

曝露とアウトカム間の因果関係を立証したからといって、必ずしも曝露を修正または除外するため

¹ Member of the GRADE working group

の全て的手段によりアウトカムの変化がもたらされるという確信的推論を導けるわけではない。単純な薬剤介入の枠組みを超えた複雑な介入の場合は特にそうである。この場合の判断においては曝露の除外が達成可能であることに対する確信性を判断する必要があり、GRADE における直接性に関する判断をする際に考慮される。

それにもかかわらず、GRADE では Bradford Hill の基準の多くを明示的または暗示的に採用してきた。しかしながら、GRADE が因果関係に関する基準をどのような形で組み込んでいるのかわかりづらいと感じる読者もいると考えられる。そこで本稿では、GRADE アプローチにおいて Bradford Hill の基準がどのように考慮されているのかを説明し、特定の基準について考慮していない理由などを、文献 3 (Schunemann HJ, et al: The GRADE approach and Bradford Hill's criteria for causation. J Epidemiol Community Health 2011;65:392-395) に記載されている内容の一部を紹介しながら説明する。

GRADE は、エビデンスの質とは総体エビデンスから得られた効果推定値(因果関係)に対する確信性であると定義する。したがって本稿ではこの確信性が上がる場合をグレードアップ、下がる場合をグレードダウンの用語を使う。

(1) 関連の強さ(Strength of the association) 。Bradford Hill は、因果関係は強い関連により裏付けられると示唆している。この基準は GRADE では、グレードアップにより直接考慮されている。GRADE システムにおいては、介入または曝露とアウトカムとの間の強い関連が、エビデンスの質のグレードアップにつながり、これはすなわち、当該介入により、そのアウトカムの発生率に変化が生じる確信が高まることを意味する。第 2 の基準である不精確さは、強い関連があっても効果に対する確信を制限するものであり、効果の大きさに確信性を損なうほどの不確実性があるならば関連に対する確信を低下させる点において、本項目に間接的に関連している。

(2) 一貫性(Consistency) 。Bradford Hill は、様々な研究からの結果が一貫していれば因果関係が存在する可能性が高まると示唆している。この基準は GRADE において直接考慮されている。GRADE アプローチは、一貫性に欠けるエビデンスが存在する場合、すなわち質が類似した複数の研究が効果推定値における説明のつかない異質性を示している際は、エビデンスの質をグレードダウンするよう示唆している。

(3) 時間的關係または研究デザインの適合性(Temporality or study design suitability) 。Bradford Hill は、曝露とアウトカムの間に時間的關係がなくてはならないと説明している。この基準は、通常、特にデザインや実施が適切でない観察研究よりも優れた研究を指すもので、GRADE においては少なくとも 3 とおりの方法で間接的に考慮されている。第一に、GRADE ではそもそもこの時間的關係を立証するものであるランダム化比較試験からのエビデンスは、この關係を立証しない研究からのエビデンスよりも質が高いものとみなす。第二に、同時対照群を組み入れた縦断的観察研究は、横断研究よりも質の高いエビデンスを提供するものと考えられる。第三に、GRADE では、観察研究のデザインを評価する際は、誤った關係を提示する原因となりうる交絡因子や共変量を批判的に考慮するよう要求している。

(4) 生物学的用量反応勾配(Biological gradient) 。Bradford Hill に記載されたように、曝露と効果の大きさ間の生物学的用量反応勾配が因果關係の確信性を高める。用量反応關係によるエビデンスの質のグレードアップという GRADE の基準は、この原則を直接適用したものである。

(5) 特異性(Specificity) 。Bradford Hill によれば、ある特定のアウトカムとある特定の曝露との間に、原因が変われば疾患のアウトカムが変わるような關係がみとめられる場合は、因果關係が存在する可能性が高い。GRADE では、曝露とアウトカムの双方が直接測定されているかの評価にお

いて、なおかつ最初の段階における疑問の定式化や集団、介入、比較対象、アウトカムの選択により、この基準を間接的に考慮している。しかし、単一の曝露または介入は必ずといってよいほど数多くのアウトカムに関係しており、その逆もまた然りである。この基準は、介入効果の評価においては重要な基準ではない²。

(6) 生物学的妥当性 (Biological plausibility)。Bradford Hill のアプローチにおいては、関連が妥当性のあるものかどうか因果関係を左右する。GRADE は、妥当性の問題を厳密に Bradford Hill で取り上げられたような形では考慮していない。これは、関連がみとめられれば研究者には必ず考えつく説明があることから、どのような関係も妥当であると説明されうるためである。しかし GRADE では、介入が代理アウトカムにどれだけ直接的に関係しているかを評価する際に、妥当性の考慮がある程度行われる。たとえば我々は、介入への反応において、患者にとって重要なアウトカムと同様の反応が繰り返しみとめられた代理アウトカムは容認することが多い。たとえば、CD4 数と HIV ウィルス量を、死亡やその他の患者にとって重要なアウトカムの代理アウトカムとして使用することは容認されており、この容認理由の 1 つとして、CD4 数と HIV ウィルス量が疾患と治療成功の決定因子であるという生物学的妥当性があげられる。また、GRADE において生物学的妥当性は観察されたサブグループ効果の信頼性を評価するための基準として捉えられている。さらに、関心のある疑問を提起し、その疑問を支持もしくは覆すエビデンスを特定することにより、生物学的妥当性の項目が間接的に考慮される。

(7) 整合性 (Coherence)。Bradford Hill によれば、観察された事象が当該疾患の自然経過により裏付けられ、なおかつその自然経過に合致するものならば、因果関係が存在する可能性が高い。GRADE はこの基準を明示的には考慮していないが、代理アウトカムの妥当性評価、ならびに適切な医療疑問の定式化においてこれらの考慮が黙示的に組み込まれている。さらに、疑問の定式化とエビデンスの評価においては短期的なアウトカムよりは直接的なアウトカム (例: 集団にとって重要な長期的なアウトカム) が重視される。

(8) 実験的エビデンス (Experimental evidence)。実験的エビデンスは因果関係の確率を高める。GRADE は厳格な実験的デザインを重視し、この基準を直接考慮している。RCT は因果関係を立証するのに理想的な実験的研究デザインを持ち、RCT におけるランダム化は関連への確信性を高める主要な実験因子である。RCT の実験的デザインや実施における欠陥は、エビデンスの質のグレードダウンにつながる。独立した対照群を持ち実験的に良くデザインされた観察研究は、デザインの劣る観察研究よりも質が高いと評価される。

(9) 類似性による推論 (Reasoning by analogy)。Bradford Hill は、因果関係は既存の類似した関連により裏付けられると示唆している。この基準は GRADE アプローチにおいて間接的に考慮されている。他の重大な複数のアウトカムに関する質の高いエビデンスが存在し、その関連が同じ方向ならば、1 つの重大なアウトカムを理由にエビデンスの全体的な質がグレードダウンされることはないかもしれない。たとえば、曝露 (例: 大気汚染) を軽減する介入が死亡と慢性呼吸器疾患に関連することが中程度 (⊕⊕⊕⊕)³の質のエビデンスに裏付けられ、脳卒中のような第 3 のアウトカムについては質の低い (⊕⊕⊕⊕) エビデンスしか存在しないが、全ての関連が同じような効果を示す場合は、たとえ質の低い 1 つのアウトカムが重大なものであっても、それのみを理由に全体的な質がグレー

² 特異性は因果関係のエビデンスとして最も弱いものであり、将来的にはリストから削除されるべきという意見もある。

³ GRADE システムによるエビデンスの質評価は、高 (⊕⊕⊕⊕)、中 (⊕⊕⊕⊕)、低 (⊕⊕⊕⊕)、非常に低 (⊕⊕⊕⊕) の 4 段階である。

ドダウンとなることはない。さらに、GRADE は直接的エビデンスが存在しない場合は非直接的エビデンスを考慮する。

以上のように、Bradford Hill の基準に関し、GRADE アプローチでは Bradford Hill が提案する考慮事項の大半が適切に組み込まれているものと考えられる。

システマティックレビューや診療ガイドラインにおける GRADE システムの使い方に関しては、文献 1,5,6、Bradford Hill 基準については文献 7 も参照していただきたい。

表 1 因果関係の Bradford Hill 基準と GRADE [table 1 ref.3]

Bradford Hill criteria		Consideration in GRADE ⁴
1	関連の強さ (Strength of the association)	Strength of association and imprecision in effect estimate
2	一貫性 (Consistency)	Consistency across studies, ie, across different situations (different researchers)
3	時間的關係 (Temporality)	Study design, specific study limitations; RCTs fulfil this criterion better than observational studies, properly designed and conducted observational studies
4	生物学的用量反応勾配 (Biological gradient)	Dose-response gradient
5	特異性 (Specificity)	Indirectness
6	生物学的妥当性 (Biological plausibility)	Indirectness
7	整合性 (Coherence)	Indirectness
8	実験 (Experiment)	Study design, randomisation, properly designed and conducted observational studies
9	類似性 (Analogy)	Existing association for critical outcomes will lead to not downgrading the quality, indirectness

⁴ GRADE システムにおいては、エビデンスの質評価のためのグレードダウンの 5 要因 (limitations, inconsistency, indirectness, imprecision, publication bias) とグレードアップの 3 要因 (large effect, dose response, confounding) がある。

参考文献：

1. 相原守夫、三原華子、村山隆之、相原智之、福田真作：診療ガイドラインのためのGRADEシステム、凸版メディア、2010年
2. Bradford Hill A: The Environment and Disease: Association or Causation? Proc R Soc Med. 1965 May; 58(5): 295–300.
3. Schunemann HJ, Hill S, Guyatt G, Akl EA, Ahmed F: The GRADE approach and Bradford Hill's criteria for causation. J Epidemiol Community Health 2011; 65:392-395. doi:10.1136/jech.2010
4. Durrheim DN, Reingold A. Modifying the GRADE framework could benefit public health. J Epidemiol Community Health 2010; 64:387.
5. 相原守夫：GRADE システムの使いかた-エビデンスから推奨へ。
Available from http://www.grade-jpn.com/how_to_use_grade_aihara_20120426.ppt
6. 相原守夫：GRADE ハンドブック（日本語翻訳版）
http://homepage3.nifty.com/aihara/grade_handbook3.2.html
7. 南郷栄秀：はじめてケースコントロールシート ver.3-0 http://spell.umin.jp/BTS_CCS3.0.pdf