

Pawson, R., Greenhalgh, T., Harvey, G. and Walshe, K., 2005, "Realist Review -A New Method of Systematic Review Designed for Complex Policy Interventions," *Journal of Health Services Research & Policy*, 10(Suppl 1): 21-34.

R. パーソンほか, 2005, 「リアリストレビュー：複雑な政策介入のためにデザインされた新しい系統的レビューの方法」

レジュメ作成者による紹介

「何が有効か」について、しばしば最も良質なエビデンスを提供するとされる伝統的な系統的レビューの方法に対し、批判的実在論に根ざした新しいレビューの方法を提示した論考。複数の段階やステークホルダーを伴う複雑な介入の有効性をレビューする際に、この方法がどのように適用できるかを解説している。

1 導入 (pp. 21-22) ¹

- リアリスト²レビュー (realist review) とは、因果関係の判断よりもその説明に重きをおいた、研究を統合するための比較的新しい戦略である。ここでは、特定の文脈や状況において、複雑なプログラムがどのように機能するのか（あるいはなぜ失敗するのか）というメカニズムを解明しようとする。
- ここでのリアリズムとは、哲学分野の議論³に根ざし、様々な分野で応用されている方法論的立場である。しかし、ヘルスケアにおけるエビデンス統合のアプローチとしての応用はほとんど進んでいない。
 - 臨床試験と比べて、ヘルスケアのマネジメントや政策介入に関する文献は、認識論的に複雑で、方法論的にも多様であるため、リアリストレビューを用いるのに適していると著者たちは主張する。
- 社会的介入において何が有効かを理解する試みには因果関係を明らかにすることが含まれるが、リアリストの研究は因果性への独特の理解にその特徴がある。
 - 因果性の継起的 (successionist) モデルでは、原因 X の後に続いて結果 Y が生じたときに、因果関係が証明されたと主張される。これは、臨床試験の基礎にある考え。

¹ セクション冒頭の番号はレジュメ作成者が割り振った。なお、議論の構造が明確になるよう、一部原文の階層構造から変更した箇所がある。

² ここでのリアリスト (realist) とは、下の脚注 2 で挙げている文献からもわかるように、基本的には哲学における批判的実在論に立脚した立場を指していると思われる。

³ Bhaskar, R., 1978, *A Realist Theory of Science*, 2nd edn, Brighton: Harvester Press. Harré, R., 1978, *Social Being*, Oxford: Blackwell. Putnam, H., 1990, *Realism with a Human Face*, Cambridge: Harvard University Press. Collier, A., 1994, *Critical Realism*, London: Verso.

- 因果性の生成的（generative）モデルでは、X と Y という 2 つの出来事間の因果的なアウトカム O を推論するために、X と Y をつなぐ基底的なメカニズム M と、両者の関係が現れる文脈 C を理解する必要があると主張される。これが、リアリストの研究の基礎にある考え。
 - リアリズムの立場をとる場合、政策やプログラムの評価に関する基本的な問いは、「何が有効か」ではなく「このプログラムの何が、どのような状況で、誰にとって有効なのか」というもの。
- 本稿では、上記の説明手法を研究の統合でどのように用いることができるかを示す。
 - リアリストのアプローチは、量的方法と質的方法のどちらか一方を特に好むということはない。むしろ、介入のプロセスとインパクトの両方を調査できるよう、量的・質的を組み合わせた多様な方法を用いることにメリットを見出す。

2 介入の性質（pp. 22-23）

- 「介入」は包括的で便利な用語だが、方法論的に別々の戦略を一緒くたにしてしまう。
 - 臨床的な「治療」は、ヘルスケアの「プログラム」と同じではない。後者は、ヘルスケアの「サービス提供」と混同されてはならず、ヘルスケア「政策」とも別物。
- 研究をレビューするときに重要なのは、レビューの方法を主題に適合させること。複雑なサービス介入には、7つの決定的な特徴がある。
 - 1) 複雑なサービス介入は理論である。すなわち、介入はつねに「もしこのような方法でプログラムを提供すれば、あるいはこのようにサービスを管理すれば、いくらか改善されたアウトカムをもたらすだろう」という仮説に基づく。
 - 2) 介入は能動的（active）である。すなわち、臨床医や教育者といった諸個人の能動的なインプットによって効果が実現する。ランダム化実験などでは、人間の意思の働きは汚染するもの（contaminant）＝介入効果を推定するために取り除かれるべきものとみなされる。対して、能動的なプログラムはステークホルダーの推論（reasoning）を介してはじめて機能し、そうした推論についての知識もアウトカムを理解するうえで不可欠。
 - 3) 介入の理論には長い道程がある。医療のパフォーマンスに関する情報公開政策を例にとると、このような介入が効果をもつためには、いくつかの段階や複数のステークホルダーが必要（後述の 4-3 の例を参照）。ただし、一連の出来事の基底にある諸理論はすべて誤りやすい＝意図しない結果を引き起こしうる。
 - 4) 複雑なサービス介入の実施は段階を直線的に進むではなく、それぞれの関係者の影響力によって方向づけられ、場合によっては前の段階に後退することもある。
 - 5) 介入は多数の社会システムに埋め込まれた脆弱なものである。文脈の影響により、あるプログラムがあらゆる状況で等しく有効であることは、きわめて稀である。リアリストの探究における 1 つの重要な要求は、介入をつくりあげる社会的現実の様々なレイヤーに留意すること。

- 6) 介入には漏れがあり、借用されやすい。介入を具体化する際、実践者は同僚と話し合い、アイデアを交換する。そこでは解決策が比較検討され、変更される。同じ介入が改良や再発明、ローカルな状況への適応によって変異すると予期すべき。
- 7) 介入はそれ自体がフィードバックするオープンなシステムである。介入が実施されるにつれて、介入それ自体が、当初その介入を機能させていた条件を変化させていく。この進化と適応は、長期的に意図しない結果をもたらさう。

3 リアリストレビュー：理論的・実践的な限界（pp. 23-24）

- 上記のような特徴をもつ複雑なサービス介入の有効性をレビューする際、レビューワーには3つの重要な理論的制約がかかる。
 - a) カバーできる領域の制約。介入には多数の段階があり、それぞれに関連する理論があり、それをとりまく状況は無限に変化する。レビューワーは、特定の状況における特定のプロセスと理論の調査に優先順位をおく必要がある。
 - b) 引き出せる情報の性質と質の制約。経験的な研究から、個人間の関係や権力闘争、文脈的条件に関わるインフォーマルな情報の入手は困難。実際的な観点からは、リアリストのレビューは多様な一次資料をもとに幅広い情報を利用することが必要。
 - c) 提言として期待できるものの制約。リアリストレビューは、一般化可能な事実よりも啓発（illumination）を、標準化よりも文脈的な調整をもたらす。

4 リアリストによる系統的レビューのテンプレート（p. 24）

- リアリストレビューのステップは以下のように集約できる（4-1～4-8）。
 - リアリストレビューでは、伝統的な系統的レビューのステップを出発点としつつ、より多くのサブステージが考慮される。
 - 実際には、これらのステップは重複しており、また反復される。

4-1 レビューの問いを特定する（pp. 24-25）

- すべてのレビューは概念を磨くことから始まる。研究の統合において追求され（てい）る問いの正確な定義と精緻化が、そこでは試みられる。
 - 薬物治療の系統的レビューでは、レビューワーが対象集団や、薬剤の投与量・投与期間・投与方法などのほか、アウトカム指標を定義。
- リアリストレビューは、同じく問いを磨くことから開始する。しかし、研究対象が複雑な介入であり、因果関係の判断よりも説明に重きをおくという違いがあることから、その仕事は通常の系統的レビューと大きく異なる。
 - リアリストレビューにおける問いの特定は、時間のかかる継続的な作業となり、しばしば作業全体の中間点あたりまで続く。

- かつて著者の 1 人は、この段階を複雑なエビデンスの統合における「沼地 (swamp)」と呼び、その不確実性と反復性を認めることがレビューの成功にとって不可欠だと主張⁴。
- レビュワーと [レビューを依頼する] 委員の間で概念を明確化するにあたって重要なのは、この種の介入の何が、誰にとって、いかなる状況で、どのような点で、なぜ有効なのかという、レビューがもたらす説明の原理に両者が同意しておくこと。
 - レビュワーと委員の両者は、これらの事項のいくつかについて交渉する「事前レビュー」の段階で協力し、定期的にレビューの焦点を見直すことが必要。

4-2 レビューの目的を精緻化する (p. 25)

- 説明に重きをおいたレビューのテーマにはいくつかのテーマがあるが、少なくとも 4 つの異なるテーマがありうる。
 - i) プログラム理論の完全さのレビュー：プログラムをリアルタイムで追跡し、一連の理論の流れやその障害物を探ることが行われる。
 - ii) 競合するプログラム理論を裁定するためのレビュー：リアリストレビューは有効性についての競合する理論を裁定するためのエビデンスを明らかにできる。
 - iii) 同一の理論を異なる状況で比較するレビュー：この戦略は、プログラムが一定の状況下で一定の参加者にしか機能しないと想定するリアリストレビューの理論的根拠。リアリストレビューは「同じ」介入による成功と失敗のパターンを識別することを試みるうえで有益。
 - iv) 公式の期待と実際の活動を照らし合わせるレビュー：一般に政策形成者と実践者は、介入に着手するうえでの最善の方法をめぐって対立する。これは、リアリストレビューを通じて経験的に裁定できるような対立する理論の発生源となる。
 - どのアプローチに基づいてレビューするかを一定の段階で明確にすることが不可欠だが、最終的な決定はレビューが本格化してからでないといけないかもしれない。

4-3 検討すべき重要な理論を明確にする (pp. 25-28)

- レビュワーは、多くの情報源からアイデアを漁って、重要な介入の理論の長いリストを作成し、そこから最終的な短いリストを作成しなければならない。
 - はじめに、委員や政策形成者、その他のステークホルダーと議論し、介入にまつわる公式の推測や専門家による問題の捉え方に慣れ親しむことが重要。
 - ある時点から、なぜ介入が有効に機能するのかについての理論や直感、期待、根拠を探る目的で文献に当たらなければならない。

⁴ Greenhalgh, T., 2004, "Meta-narrative Mapping: A New Approach to the Synthesis of Complex Evidence," Hurwitz, B., Greenhalgh, T. and Skultans, V. eds., *Narrative Research in Health and Illness*, London: BMJ Publications.

- 収集すべきデータは、介入の効能（efficacy）に関するものではなく、その介入がどのように機能するはずであったか（そしてなぜうまくいかなかったか）についての理論や説明がおよぶ範囲に関するもの。
- 例：医療のパフォーマンスに関する情報公開政策⁵。
 - 情報公開はいくつかの活動（すなわち理論）からなる。
 - ①分類：ヘルスケアの特定の側面における質を測定し、パフォーマンスの順位づけをする。何をどのように測定して分類するかは決定などは論争の的となる。
 - ②公開：どのような情報を、どんな手段で、誰に公開するかといった複数の選択が含まれる。
 - ③制裁：より広いコミュニティに属する人々が開示された情報に基づいて行動すると予期される。コミュニティの中のどのメンバーを情報の受け手と見なすか、どのように制裁が下されると想定するかについて複数の選択の余地がある（サービスの規制強化に利用される、消費者の選択を刺激するなど）。
 - ④反応：情報公開の対象となる側に関するもの。基本的に、パフォーマンスが高いとされた者はその地位を守ろうとし、低いとされた者は回復を図ろうとすると期待される。ただし、後者の反応は③の制裁の内容に依存する。
 - 上記①～④は専門家によるプログラムの構想に当たるが、介入が計画通りに進むことは滅多にない。したがって、計画の成否について対立する推測（⑤～⑦）に直面する可能性が高い。
 - ⑤抵抗：公開に対する抵抗として、医師や病院など、情報公開の対象となる側から、その適用や権威に対して異議申し立てがなされる可能性がある。
 - ⑥競合する枠組み：上記のようなプログラムの枠組みが、たとえば「メディアの枠組み」になじむかはわからず、パフォーマンスが低い者の復帰を促すことを意図したメッセージとは異なるストーリーが展開されるかもしれない。
 - ⑦尺度のごまかし：他の活動を犠牲にして測定される活動に注力したりして、測定の裏をかこうとする可能性がある。
- このようにして理論をマッピングすることで得られるアイディアは多種多様。多数の理論のどの部分をどの組み合わせで短いリストにまとめるかを決めなければならない。
 - その際、包括的なレビューは不可能であり、どのプログラム理論を調査するか優先順位をつけて合意をしなければならないというシンプルな原則が重要。
 - 探求すべき仮説を表す唯一のやり方があるわけではない。

⁵ Marshall, M., Shekelle, E. P., Brook, R. and Leatherman, S., 2000, *Dying to Know: Public Release of Information about Quality of Health Care*, London: Nuffield Trust.

4-4 レリヴァントなエビデンスを探す (pp. 28-29)

- リアリストレビューにおけるエビデンスの探索は 4 つの構成要素からなると考えるのが有益。ただし、実際にはこのように整然かつ直線的に行われることはない。
 - 1. 文献の感触をつかむための背景的な探索。
 - 2. プログラム理論を特定するための漸進的な焦点化。
 - 3. プログラム理論の一部をテストするための経験的エビデンスの探索。
 - 4. 研究の統合がほぼ完了した時点での最終的な探索。
- リアリストレビューにおけるエビデンスの探索は、伝統的な系統的レビューにおけるそれとは、2 つの重要な側面で異なる。
 - 定義可能で発見可能なレリヴァントな文書には限りがあると考えない。どのレビューでも、潜在的にレリヴァントな情報源は実際にカバーできないほど多く存在する。そのため、一種の目的に応じたサンプリング（purposive sampling）戦略が必要。
 - 厳密さを理由に、ごく一部を除いたすべてのレリヴァントな研究を考慮から外すことは、結果の妥当性や一般化可能性をむしろ低減させると考える。なぜなら、それぞれの研究は、エビデンスの統合全体を構成する要素の異なる部分に貢献するから。
- リアリストレビューでは目的に応じたサンプリング、すなわち特定の問いに答えたり特定の理論をテストするために意図的に素材を取り出す。
 - 理論的なニーズが充足されたとき、あるいは問いの答えが出たときに、十分なエビデンスが集まったと判断する。これは、質的研究における理論的飽和（theoretical saturation）の概念に相当する⁶。
 - 目的に応じた探索のアプローチは、確率的サンプリングで達成されるようなきちんと定義された抽出枠をもたない。また、目的に応じたサンプリングは、理論的な理解が発展するにつれて繰り返される必要があるという点で反復的である。
- リアリストレビューは探索において索引見出しやキーワード検索、検索エンジンなどを他の系統的レビューと同様に用いる。ただし、いくつかの重要な点が異なる。
 - リアリストレビューは介入の内側の隠れた仕組みに目を向けるため、学術論文だけでなく、出版されていない文献や資料を利用する可能性が高い。
 - リアリストレビューは特定のトピックではなく介入の基底にあるメカニズムを分析の主要な単位とするため、異なる分野の文献など、より幅広い経験的研究がレリヴァントになると考えられる。したがって、探索するデータベースをきつく制限することは不適切である。

⁶ Glaser, B and Strauss, B., 1967, *Discovery of Grounded Theory*, Chicago: Aldine. 理論的飽和の概念については、社会政策/社会福祉②の第 7~9 回のレジュメも参照。

4-5 エビデンスの質を評価する（pp. 29-30）

- リアリストレビューは、研究の質を判断する方法において伝統的な系統的レビューと異なる立場をとる。
 - 生物医学的な介入の系統的レビューでは、ランダム化比較試験（randomized controlled trials: RCTs）を最上位におき、それ以外を下位におくエビデンスのヒエラルキーに依拠する。他方リアリストレビューは、より豊かな描写を行うために多様な方法を必要とするため（後述）、ヒエラルキーに基づくアプローチをとらない。
 - 複雑なサービス介入をテストするうえで RCTs が抱える問題は、そうした介入がつねに他のプログラムの影響のもとで実施されるため、介入群と対照群のクリーンな比較が困難な点にある。こうした介入について RCTs を実施することは可能であるものの、RCTs のデザインでは重要な説明要因が明確に排除されるため、得られる情報は貧弱なものになる。
- リアリストレビューは、高度に多様な主題に合わせて方法を取捨選択することで、複雑な現実を探求しようとする。ここでも、質の評価は次のように段階的に行われると考えるのが有益。
 - レリヴァンスの評価：リアリストレビューにおけるレリヴァンスとは、その研究がある特定のトピックをカバーしているかではなく、目下テストされている理論に取り組んでいるかどうかによって評価される。
 - 厳密さの評価：元の研究が導出した特定の推論が、特定の介入理論のテストにおいて方法論的に信頼できる貢献をするに足る重みをもつかどうかを評価。
- 上の 2 つは、研究を取捨選択するための絶対的な規準ではなく、ある特定の研究統合の目的に対する適合性を示すもの。
 - 例：ヘルスケアのパフォーマンスデータの公開についてのエビデンスを探す場合。
 - そうしたデータがサービス利用などに関する患者の決定にどの程度関わるかを検討したい場合、これについての研究は様々な形がありうる。
 - 優れた研究統合では、それぞれの情報源の相対的な貢献を判断しており、そこではいくつかのエビデンスが棄却される。重要なのは、優れた研究統合においては、そこでの推論が明確にされていること。
 - 研究を統合することとは、それぞれの研究の異なる貢献が意味をなすこと。たとえば、A に慎重であるべきという分析の根拠が、B から学んだことや C が示唆することによって説明されるかもしれない。このような推論の連鎖は、研究の価値は事前に行う適格さの判断によって確立されるのではなく、それらが統合されることで確立されることを示唆する。

4-6 データを抽出する（pp. 30-31）

- 伝統的な系統的レビューにおけるデータの抽出は、各論文から同じ情報のアイテムを引き出すために設計されたフォームを用いて行われる。リアリストレビューでも、資

料の選別などにフォームを用いるが、それは単一の標準化された質問リストのような形はとらない。

- これは、リアリストレビューが取り組む仮説が多面的であることや、考慮に入れるエビデンスの情報源が多様であることによる。
- リアリストレビューでは、データを抽出するというよりも、メモをとったり注釈をつけたりすることで情報を取り入れる。
 - たとえば理論を追跡する作業をする場合は、ある介入がどのように機能するはずであるかに関するアイデアを文書から探し出そうとする。そうしたアイデアにマーカーをしたり、記録したり、大まかなラベルを付与したりする。
- くわえてリアリストレビューにおけるエビデンスの読解の仕方には 2 つの特徴がある。
 - 伝統的な系統的レビューが参照した研究のリストを添付し、どれが統計分析に使われたのかを示すように、レビューが展開するなかでどの研究が使われ、どの研究が除外されたかについての解釈の痕跡（interpretive trail）を残しておくことが必要。
 - リアリストレビューのステップは直線的でなく、既存の研究に繰り返し立ち返るため、最後の段階までデータの抽出が行われる。

4-7 エビデンスを統合する (p. 31)

- リアリストレビューでは、エビデンス統合の目的は理論の精緻化であると考えられる。したがって、統合によって達成されるのは、介入がいかに機能するかについての理解を微調整することである。統合とは、説明を進歩させることを指す。
 - レビューは、「この種の介入の何が、誰にとって、いかなる状況で、どのような点で、なぜ有効なのか」に対する説明の一部もしくは全体の進歩に取り組む。

4-8 提言をまとめ、知見を発信する (pp. 31-32)

- 一般に系統的レビューは提言とともに締め括られる。最近では、提言をまとめる段階でレビュー者と [レビューを委任する] 委員が相互に対話することが推奨されている。リアリストレビューは、このような両者の対話を方法論的に要請する。
 - というのも、先述のように、レビューにおける問いの特定や理論の明確化には、実践者や政策形成者からのインプットなしには有意義に行えないと考えるから。
- レビューを普及させ、介入を実施する段階では、他の系統的レビューと同様に、現場の実践者がレビューの知見に注意しながらそれを実施することが目指される。ただし、系統的レビューの場合、それによる変化は提言が示す方向に向けた単純な行動変化として測れるかもしれない（例：医師が症状 Y に対して治療法 X を処方しているか）。他方、リアリストレビューの知見に基づく実施は、多くのアクターや複数のプロセスを含む複雑なものであり、すべての人が B をすることをやめて A をしているかといったことが問題になるのではない。

- リアリストレビューが有効に普及し実施されるとすれば、ある状況ではプログラムにおけるわずかな重点の変化が、別の状況ではプログラムの拡大が、さらに別の状況ではプログラムの完全な放棄がなされると期待できるかもしれない。
 - というのも、何が、誰にとって、どのように、いかなる状況で有効に機能するかについての知見と、プログラムの各要素がどのように合致しているのかについて、一定の情報に基づく判断がなされると考えられるから。

5 リアリストアプローチの強みと限界 (p. 32)

- リアリストレビューの強みは、哲学や社会科学に強固なルーツをもち、多元的で柔軟であることを本質とする探求の論理であるという点にある。
 - リアリストレビューでは、ステークホルダーの関与が体系化されており、また問題に対して画一的に対応するという誤りを回避するような原理を提供する。
 - くわえて、プログラム理論を分析単位とすることで、政策分野や学問分野、組織の境界をまたいだ学習を最大限に活用することが可能。
- 他方、リアリストレビューには、その適用を制約する重要な欠点がある。
 - たとえば、リアリストレビューは定型的で手続きを重視するアプローチには適用できない。リアリストレビューは、伝統的な系統的レビューのように標準化可能、再現可能なものではない。
 - リアリストレビューによる提言は、せいぜい暫定的なものにとどまる。すべての結論は文脈依存的なので、一般化可能な効果の大きさを示すことは叶わないだろう。これが欠点であるか美德であるかは見方次第である。

6 政策形成に対するリアリストレビューの貢献 (pp. 32-33)

- リアリストレビューは、政策においてどのような機能を果たしうるだろうか。
 - リアリストレビューは、Weiss⁷のいう理論ベースの評価 (theory-based evaluation) の一派。理論ベースの評価は、技術的あるいは党派的なサポートよりも、啓蒙 (enlightenment) をその役目とする。そうした啓蒙は、研究による政策への影響がデータではなくアイディアの媒介によって生じるという点にポジティブな可能性があるとされる。
 - この点でリアリストレビューには大きな利点がある。政策形成者は、多数の媒介変数 (mediator) や調整変数 (moderator) の統計的有意を示す提言に苦勞するかもしれない。一方、政策形成者は、あるプログラムがある文脈において他の文脈よりも有効に機能するのはなぜかについての説明を解釈し利用できる可能性が高い。

⁷ Weiss, C. H., 1986, "The Circuitry of Enlightenment: Diffusion of Social Science Research to Policymakers," *Knowledge*, 8(2): 274-281. なお、ここでは原文の文献注を修正した。