

Munro, E., 2014, "Evidence-Based Policy," N. Cartwright and E. Montuschi eds., *Philosophy of Social Science: A New Introduction*, Oxford: Oxford University Press, 48-67.

E. マンロー, 2014, 「エビデンスに基づく政策」

レジュメ作成者による紹介

社会政策の研究者が、社会科学の哲学の教科書に執筆した「エビデンスに基づく政策 (evidence-based policy : EBP)」についての概説。EBP におけるエビデンスの概念やそれを提供するための方法、EBP に対する主要な批判などがまとめられている。

1 導入 (pp. 48-49)

- エビデンスに基づく政策 (evidence-based policy : EBP) は、どのような政策を実施すべきかについての決定を改善する方法として、近年推奨されている新しいアプローチ。
 - 従来から政策決定者は、何らかの決定を行う際には、科学的研究に基づくものを含めたエビデンスに依拠してきた。
 - EBP の新しさは、たんに政策決定において科学的エビデンスに依拠することを奨励するだけでなく、イデオロギー的に推し進められる政治をより合理的な意思決定へ置き換えようとする点にある。
 - evidence-based な政策という語は、研究が政策を決定できるという決定論的な印象を与えるために、evidence-informed や evidence-influenced といった語も提起されている。
- EBP の主な焦点は、「X を行えば、目標 Y を達成できるだろう」という類の因果的主張の強固さ (strength) をテストすること。
 - これが、「何が有効か (what works)」というタイトルがついた文献の量産をもたらしている。
- 「何が有効か」というフレーズは多くの論点を含んでおり、EBP というトピックは多くの論争や研究をもたらしてきた。
- この章では、EBP についてどのような主張がなされてきたのかを説明し、その最も論争的な側面のいくつかについて検討する。

2 EBP とは何か（pp. 49-51）

- EBP には多くのバリエーションがあるが、そのほとんどは、[得られた知見が局所的な意義しかもたないと考える相対主義的な見方ではなく——レジュメ作成者注] 社会的な世界についての客観的で普遍的な知識を発展させることができると考える実証主義的な認識論を具体化したもの。
 - 客観的で普遍的な知識は、政治家やその他の政策決定者に問題解決のやり方を知らせることができると思定される。
- EBP の語は、1990 年代に生じた政治的関心とともに、近年広く使用されるように。
 - たとえばイギリスでは、労働党政府が 1999 年に出した「政府の近代化」のアジェンダの重要な側面として EBP が位置づけられた。
 - 国連では、ミレニアム開発目標に関する議論のなかで、EBP の推進が推奨された。
- EBP は 1980 年代から注目されたエビデンスに基づく医療（evidence-based medicine：EBM）の影響を受けてきた。
- EBP への関心の増大をもたらしたのは何か。
 - ① 専門家がもつ専門知識への懐疑の高まり。かつては医師、警察官、教師などは良い判断をすると信用されていたため、その仕事のあり方はそれほど問題とされなかった。EBP の支持者にとって、いまや専門的な判断は非常に誤りやすいものとみなされ、専門家の高い地位の正当性が疑われている。
 - ② 情報技術の発展によって、政策に関するより多くのデータが利用可能になってきたこと。
 - ③ 最も大きな要因は、EBP の魅力。EBP が支持される際には、確かなエビデンスの必要性が示され、エビデンスは客観的かつ個人的なバイアスや利害の影響を受けないと主張される。ゆえに、エビデンスに基づく政策は「本当に」役に立つだろうとされる。

3 「最良」のエビデンスとは何か（pp. 51-53）

- EBP および EBM において重要なのは「最新で最良のエビデンス」。
- そのため、EBP には、エビデンスや既存研究を判定・評価する明確な規準を開発し、政策決定者がみずからの目的にとって最良のエビデンスを決めるための手引きを提供することが含まれることが多い。
 - 研究成果の質についてレビューや批判を行い、その判断を下すための機関も設立されてきた。

- EBP におけるエビデンスとは、たんにある言明の固定的な性質を指すのではなく、何かについての、また誰かにとって何らかの主張を受け入れるための根拠として有用であるような言明の性質のことを指す。
 - 言い換えれば、研究の結果は、何者かがそれを自身の主張しようとする言い分を下支えするものと判断した際に、はじめてエビデンスとなる。
- EBP において最も重要なタイプのエビデンスは、因果関係についての主張を支持するエビデンス。
 - これは典型的には、経験的研究の知見のみを指し、その他のタイプの情報は信頼性が低すぎるとみなされる。
- このため、正しく実行されたランダム化比較試験（RCT）の結論はエビデンスとみなされる一方、たとえば長年の教育経験に基づく教師の観察は、エビデンスとみなせるほど頑健ではないとされる。
 - 後者の情報は、教師の個人的な信念などによるバイアスの可能性があるため、信頼性が低いとみなされる。
- 因果的主張を支持するエビデンスの頑健さを正確に測ることへの関心から、研究方法の違いに基づいたエビデンスのヒエラルキーが構築されてきた¹。
 - これは、ある研究デザインは他の研究デザインよりも、因果的主張の強固なエビデンスの提供において優れているという想定に基づく。
 - 典型例は、アメリカ連邦政府による医学研究のエビデンス・ヒエラルキー（表1）。
- 一般に、RCT の結果および RCT の系統的レビュー（systematic review）が、因果的主張のエビデンスとして理想的だとされ、ほとんどのエビデンス・ヒエラルキーの最上位にある。

表 1

| |
|---|
| I：適切に実行されたランダム化比較試験（RCT）；良く実行された同質的な RCT の系統的レビューまたはメタアナリシス |
| II-1：良くデザインされた、ランダム化していない比較試験 |
| II-2：良くデザインされた、コーホート研究または症例対照研究 |
| II-3：複数の介入前後の時系列デザイン；目覚ましい成果が得られた非対照実験 |
| III：臨床経験に基づく権威の意見；記述的研究または症例報告；専門家会議の報告書 |

¹ エビデンスのヒエラルキーおよびそれへの批判については、社会政策／社会福祉①の第 7 回のレビューも参照。

4 ランダム化比較試験（pp. 53-57）

- 政策決定者は、みずからの政策が実施する必要があるほど有効なものなのか、政策が成功したのかどうかを知りたいが、そこには交絡因子＝観察された結果の産出において因果的に有意であるが当の政策の要素ではない因子が存在する。
 - 例：将来的な学生数を抑制しない見込みで行われた学費の引き上げの後、入学希望者数の下落を観察した。これは政策が失敗したことのエビデンスにならない。たとえば、入学希望者の年齢層の出生率が他の年と異なるかもしれない。
- RCT は、交絡の発生をコントロールし、因果的主張を行う際のバイアスや不正確さのリスクを最小化するように設計される。
 - 中心的なアイデアは、介入を受けるか否かを除くすべての側面で似通った 2 つのグループをつくり、結果に現れたあらゆる差異がその介入の有無に帰属できるようにすること。
- RCT の特徴の 1 つは、研究参加者が介入を受けるグループ（＝実験群）と受けないグループ（＝統制群）のいずれかに無作為に割り当てられること。
 - これにより、結果に影響を与えるあらゆる交絡因子が両グループに等しく分配される確率が高まる。
- もう 1 つの特徴は、研究の手続きをコントロールして、すべてのグループのあらゆる研究参加者が、それぞれのグループ特有の因子（すなわち介入のタイプ）があることを除いて、同じように取り扱われるようにすること。
- 上記 2 つは、介入によるものではない外部からのあらゆる変化が、実験的な手法で統制もしくは無作為化されていることを保証しようとするもの。
 - これにより、研究の結果を介入およびコントロールした条件の差異に因果的に帰属することが可能。

4-1 RCT の実例

- Krueger（1999）²では、クラス規模が生徒のテストの点数に与える影響を探るために RCT を実施。
 - 2 つの実験群を設計：少人数クラス（教師 1 人につき生徒 13～17 名）、補助つき標準クラス（生徒 22～25 名でフルタイムの補助教員がつく）＋統制群である標準クラス（生徒 22～25 名）

² Krueger, A. B., 1999, “Experimental Estimates of Education Production Functions,” *The Quarterly Journal of Economics*, 114(2): 497-532.

- 合計 11,600 人の生徒およびその教師を、それぞれのグループに無作為に割り当て。生徒が入学前の幼稚園児の時点から小学校 3 年時までクラス分けを維持。
- RCT を実行する段階では、上記の研究デザインからの逸脱がいくつか発生。
 - ①標準クラスにしばしば補助教員がついていた。②小学 1 年時のはじめに標準クラスの生徒を補助つきクラスと補助なしクラスに無作為に割り当てしなおした。③約 10% の生徒が学年の変わり目に少人数クラスと標準クラスの間を移った。④実際のクラス規模が少人数クラスとして想定したものよりも変動した。⑤プロジェクトからの脱退が高い割合で生じた。
- 研究の結果、少人数クラスに参加した最初の年には、テストの平均成績が 4% ポイント上昇し、続く年には、少人数クラスの生徒のテストの点数の優位が毎年約 1% ポイントずつ上昇していたと結論。
 - いくつかの変数を統制したところ、教師の教育は「系統的な効果なし」、フルタイムの補助教員がつくことは 1 年次においてのみ成績に強い正の効果が認められた。

5 系統的レビュー（pp. 57-58）

- 系統的レビューは、エビデンスのヒエラルキーの頂点に位置する。
 - なお、メタアナリシス³とは、多数の研究の知見を統計的手法を用いて統合するもので、母集団のサイズを大きくすることで結果の信頼性を高めることを目的とする。
- 系統的レビューはナラティブレビューと対比される。
 - ヘルスケアの研究でナラティブレビューは長年用いられてきたが、文献の調べ方やレビューに含める文献の決め方について明確なルールに則っていないという意味で、しばしば系統的でないといみなされる。
 - 系統的レビューは、事前に定められた厳格かつ明確な方法論を用いて、レビューにおける恣意性やバイアスを回避することを目的とする。
 - 系統的レビューをヘルスケア専門職のあいだに広めたコクラン共同計画によれば、系統的レビューとは、事前に計画した方法を用いて既存研究を収集したうえで、そのバイアスなどを小さくするようなやり方で既存研究を評価・統合する方法論として紹介されている。

³ 系統的レビューとメタアナリシスは、ほぼ同じものを指す語として用いられることがあり、ここでとりあげている文献でも互換的に用いられている節がある。ただし、厳密には両者の意味は異なる。簡単にいえば、系統的レビューとは（後述のコクラン共同計画の説明にあるように）既存のオリジナルな研究を収集・評価・統合する一連のプロセスに関する方法論である。他方、メタアナリシスとは、レビューする複数の研究を統合する 1 つのやり方であり、統計的な手法を用いて定量的に知見を統合するやり方を指す。

- ただし、既存研究の検索・評価について系統立てられた方法論があるにもかかわらず、レビューのプロセスには依然として無数の判断が伴うと指摘する者もいる。

6 EBP への批判 1：客観的であるべきという主張について（pp. 58-60）

- EBP は、Montuschi（2014）⁴が社会科学一般において支配的だと述べるアプローチを採用する。すなわち、自然科学をロールモデルとみなし、「EBP は自然科学と同様に客観的たりうるか」という形で客観性の問題を設定する。
 - そこでの 3 つの論点：①研究対象となる社会的事実や事物は、山や星のようなものとして「実際に」存在するか。②価値を除外できる（すべき）か。③量的方法は正しい結果を確保するための唯一の方法か。
- ①：EBP が焦点を当てる社会的事実・事物は、自然現象とは異なるタイプのもの。
 - 「貨幣」「家族」「教育」などの語は人間の合意によって意味を獲得し、人間なしに存在することはないだろう。
 - 社会調査で用いられる概念が特定の集団によって構成・選択されていると主張する立場からは、権力のある集団が支配的な影響力をもち、より力の弱い集団が問題を理解する経路を阻害するという異議申し立てがなされる。
 - この異議申し立ては、権力をもつ政策決定者が、最も脆弱で政治的代表性が弱い人々に影響をもたらす決定をする社会政策においてとりわけ重要。
 - これは、研究の「客観性」が、社会でより権力をもつ人々の価値や選好が影響を与えている度合いを覆い隠すことを含意する。
- ②：EBP の支持者は（とくに研究の実行プロセスにおいて）価値を除外することを望む場合があるが、価値は広範に影響を及ぼしているという批判がある。
 - 政治学の研究は、ピュアなデータが脱政治化された領域で用いられるという理想的な見通しに警告を発している。
 - 使用する言語の影響もある。上述のように、社会的に構成された概念は、それを構成した社会集団の価値を反映する。
 - 例：イギリスの政策で用いられる「問題ある家族（troubled family）」という語。「問題ある」というラベルを付与される家族の多くはそうした記述を拒絶する。
 - どのような研究課題が（政策やプログラムの）評価研究として資金提供に値するかにも価値が影響する。

⁴ Montuschi, E., 2014, "Scientific Objectivity," N. Cartwright and E. Montuschi eds., *Philosophy of Social Science: A New Introduction*, Oxford: Oxford University Press, 48-67.

- ③：量的データは、人間の理解や経験の個々のニュアンスを把握するにはあまりにも大雑把であるため、量的方法に限定した研究は、政策実施の重要な次元を見過ごす可能性がある。
 - 例：失業対策の研究の 1 つとして、失業に伴い福祉手当を請求する人の数を測定するやり方がありうる。しかし、そうした人々の経験は多様である。ある人のニーズにとって十分な給付も、健康問題を併せもつ人にとっては著しく不十分かもしれない。政策が目的を達成しているかどうか判断するには、この種の個人的なデータの収集も必要。

7 EBP への批判 2：普遍的な知識を産出すべきという主張について (pp. 60-64)

- EBP の基本的な考え方によれば、RCT で示された因果関係に関する知見は、その母集団が RCT を実行した環境における母集団と「十分に類似」していれば、他者の意思決定において利用可能であるというもの。
 - 問題は、「十分に類似」しているとはどういうことかという点。

7-1 外的妥当性

- この問題においては、内的妥当性と外的妥当性の概念が重要。
 - ある研究が、その研究結果について強固な根拠を提供している場合、内的妥当性がある。RCT は、研究する母集団における因果関係について強い根拠を提供する。
 - RCT と「同じ介入」が「同じ結果」を別の母集団でももたらす場合、その研究は外的妥当性がある。しかし、新たな母集団が元の母集団と十分に類似しているかどうかを判断するための手引きはほとんどない。
- まず、何を「同じ」介入とみなすかが問題。
 - RCT が「何」が有効であるかのエビデンスを提供するには、試みられた介入の正確で完全な記述とその再生産が求められる（これを「モデルのフィデリティ fidelity to model」という）。
 - これには、介入のどの側面が因果的に重要かについての理論的な前提が必要。
 - 例：対象者と同じジェンダーの者によって介入が行われることは重要か。
- ある介入が実施される際には、無数の補助的な要因がそれを支えている。RCT の知見を別の文脈で利用する場合には、これら補助的な要因についても知ることが必要。
 - 例：こどもの栄養状態を改善するために母親たちを教育するという政策。ある国での RCT ではその有効性が確認されたが、別の国で実施すると、その国では母親が食材の購入やこどもの食事を決定する役割を担っていなかったため、期待された効果がなかった。

- Cartwright and Hardie (2012)⁵が言うように、原因として強調されるもの（この場合は政策）が、単独でその結果をもたらすほどの寄与をすることは滅多になく、「チームの助け」（上記の「補助的な要因」、または例における母親の役割）が必要。
- Cartwright and Hardie (2012) は、RCT の結果が他の文脈でも有効かどうかを予測する作業は、多くの EBP 支持者が考えるよりもはるかに複雑だと主張。
 - 研究した母集団で働いているのと同じ因果的な仕組みが、介入する母集団でも働いていると考えるためには根拠が必要。さもなければ、研究の結果は、介入する母集団とはただただレリヴァントでないものになると主張。

7-2 鍵概念の社会的構成

- 社会的世界の鍵概念は社会的に構成されており、社会調査における鍵概念も、社会集団のあいだで（また同一の社会集団内でも時間の隔たりによって）異なる意味を付与される。
- これは、ある研究結果を普遍的なレリヴァンスをもつものとして単純に扱うことはできないことを含意する。
 - この点は、多くの異なる RCT をひとまとめにする系統的レビューに対して深刻な影響を与える。

7-3 社会的世界における因果性への誤った見方

- 標準的な EBP は、社会的世界における因果プロセスについて、過度に単純な見方をしているという批判も。
 - 標準的な EBP では、観察された因果関係からの一般化を許容するような一定の規則性が想定されている。
- 批判者からは、社会的世界はオープン・システムであり、そこでの因果プロセスは非線形であるという主張がなされる。
 - オープン・システムでは外部からのインプットがあり、これが組織内部における変化を評価するにあたって問題を発生させる。というのも、この場合、組織に影響を与えうるすべての要因をコントロールすることはできないから。
 - この点は、プロセスを予測・統制する能力についての重要な含意をもつ。

⁵ Cartwright, N. and Hardie, J., 2012, *Evidence-Based Policy: A Practical Guide to Doing it Better*, Oxford, UK: Oxford University Press.

- 制度的な力も影響力が大きい。同じ「政策」は、異なる文脈で実施されることで異なる形をとるかもしれない。
 - 例：学校で新たな試験を導入する政策は、各学校で実施されるなかで変更されていき、さまざまな効果をもたらすかもしれない。
- EBP における因果性を見方を批判する立場も、いくつかのありうる説明に裁定を下すうえで経験的研究を行うことの価値は認めている。
 - ただし、その結果は、しばしば主張されるよりも限定的なものであると指摘される。

8 EBP への批判 3：RCT を好む傾向は研究の問いを限定する（p. 64）

- エビデンスの源として RCT が注目されることは、RCT が答える問いばかりが注目され、その他の問いの価値が損なわれるのではないかと危惧する向きがある。
 - Solesbury (2001)⁶は、EBP の必要性を「何が有効か」という観点だけで解釈してしまうと、「何が進行しているのか」「何が問題なのか」「それは……より良い／悪いか」といった、政策の展開において重要な他の問題を度外視してしまうと述べる。

9 政策策定における研究エビデンスの限定的役割（pp. 64-65）

- EBP それ自体への批判ではないが、エビデンスが実際の政策形成にもたらす貢献はささやかなものであるという点には注意が必要。
 - 研究のエビデンスが政策を決定づけることはできず、多くの他の要因が考慮されなければならない。
- Weiss (1977)⁷は、政策決定に影響をおよぼす要因として、とくに4つの「I」（＝情報、利害、イデオロギー、制度）を挙げる。

10 結論（pp. 65-66）

- EBP を批判する立場のなかには、社会調査からの一般化の可能性を完全に否定する向きもある。
- しかし、批判者の大多数は、経験的な研究が政策決定者にとって有益なリソースになりうると依然として考えている。しかし、彼らが主張するのは、その有益さは、EBP が想定するよりも複雑で限定的なものであるということである。

⁶ Solesbury, W., 2001, *Evidence-Based Policy: Whence it Came and Where it's Going*, London: ESRC UK Centre for Evidence Based Policy and Practice.

⁷ Weiss, C. H., 1977, "Research for Policy's Sake: The Enlightenment Function of Social Research," *Policy Analysis*, 3(4): 531-545