

Hammersley, M., 2005, "Is the Evidence-Based Practice Movement Doing More Good than Harm? Reflections on Iain Chalmers' Case for Research-Based Policy Making and Practice," *Evidence & Policy*, 1(1): 85–100.

M. ハマーズリー, 2005, 「エビデンスに基づく実践を目指す運動は害よりも利益をもたらすか？」

#### レジュメ作成者による紹介

質的研究の方法論に関する研究で知られる著者が、エビデンスに基づく政策形成や実践を擁護する論者を批判した論文。実践者の意見と研究のエビデンスを明確に区別することで、研究のエビデンスが政策形成や実践にもたらす利益が過大評価されていると主張している。

#### 導入<sup>1</sup> (pp. 86-88)

- 政策形成や実践が「エビデンスに基づく」ものであるべきという考えが広く受け入れられるようになってきた。エビデンスに基づく実践を目指す運動では、政策形成や実践が、系統的レビュー（systematic review）の形式で提示される、研究に基づいたエビデンス（research evidence）に基づくべきだと主張されてきた。
  - 系統的レビューとは、一定の方法論的厳密さの基準を満たした関連する研究の知見を統合したもの<sup>2</sup>。
- ここで用いられる「厳密さ」は、広い意味での実証主義（positivist）の特徴をもつ研究方法論の概念を前提としている。結果的に、エビデンスに基づく実践に向けた運動は、多くの社会学者にとって重要な以下のいくつかの論点を提起する。
  - 政策形成や専門的な実践に関する他の情報源と比較して、研究に基づくエビデンスが果たす適切な役割とは何か。
  - 意思決定において判断（judgement）が果たす役割に対して、エビデンスが果たす役割とは何か。
  - 文献レビューにおいて研究に基づくエビデンスのレリヴァンスはどのように定義すべきか、また文献の検索は網羅的である必要があるか。
  - 方法論的な健全さ（soundness）をどのように決定すべきで、それはどれほど重要なことなのか。
  - 異なる研究の知見を「統合」する際の価値あるやり方とはどのようなものか。
  - 研究が政策形成や実践になしうる貢献とはどのようなものか。

<sup>1</sup> この1つ目の見出しはレジュメ作成者が割り振った。

<sup>2</sup> 系統的レビューについては、社会政策／社会福祉②の第10、11回なども参照。

- エビデンスに基づく実践を目指す運動は保健医療の領域に端を発するが、その影響は他の領域にも広まっている。ここでは、医療の領域と同様に、運動に対する抵抗も見られる。本稿では、Chalmers (2003)<sup>3</sup>をとりあげて、エビデンスに基づく実践を支えるものとしての研究を擁護する立場の議論を検討する。
  - Chalmers は、医療の領域においてエビデンスに基づく実践を擁護する最も影響力ある者の1人で、他の領域における運動の展開も支持している。
- Chalmers は、エビデンスに基づく実践に向けた運動の主要なテーマを次のように要約する：専門的な実践者は他人の人生に介入する際、ときに利益よりも害をもたらすことがあるため、その政策と実践は厳密で透明性の高い、最新の評価に基づいたものであるべき。
  - このような要求は、多くの臨床的判断が最新の研究エビデンスに基づいていなかったという認識や、臨床医の標準的な治療が患者の改善をもたらさず、ときに悪化させていたという発見によって活気づけられてきた。
  - こうした議論が他の実践へと拡大していく中でも、厳密さを確保するうえで手続きの「透明性」が重要であると強調される点は変わらなかった。
- 著者は Chalmers の主張のうち、以下のような点には広く同意が得られるべきだとする。
  - あらゆる領域の実践者は、ときに利益よりも害をもたらす。
  - 政策形成や実践に関する情報提供において、研究は重要な役割を果たす。政策が実際にどのように実施されたのか、その文脈に応じてどのような変化が生じたのか、政策や実践がもたらす効果はどのようなものであるかといった点についての知識を、研究によって提供することができる。
  - 個々の研究の知見をそのまま信頼性の高い情報として提供するのではなく、利用可能な文献 (literature) のレビューを通じて一般の聴衆に提供すべき。個々の研究の知見はせいぜい健全な知識に向けた一步に過ぎず、文献レビューを行うことで研究を政策形成や実践の世界で活用することが可能となる。
- しかし、著者は Chalmers の議論の主要部分について、以下のような疑念があるとする。なお、Chalmers の議論の主要部分の多くは、エビデンスに基づく政策や実践を擁護する論者の常套句 (stock-in-trade) である。
  - 実践者がときに利益よりも害をもたらすのを防ぐことは可能か？
  - 実践者が利益をもたらすか、それとも害をもたらすかを、研究のみで確実に決定することは可能か？
  - 政策や実践を研究によって評価することは、つねに害よりも利益をもたらすか？

<sup>3</sup> Chalmers, I., 2003, "Trying to Do More Good than Harm in Policy and Practice: The Role of Rigorous, Transparent, Up-to-Date Evaluations," *Annals of the American Academy of Political and Social Science*, 589(1): 22-40.

- これらの疑念は、Chalmers の議論にみられる、実践者の意見（practitioner opinion）と科学的な研究のエビデンスとの明確な区別から生じている。
  - Chalmers によれば、臨床医はしばしば自身が用いる治療法の価値を確信しているが、いくつかの調査は、治療法の適用の仕方は臨床医ごとにバラバラであることを示している。研究に基づくエビデンスは何が有効で何が有効でないかを立証することで、そうした意見の衝突を「裁定（adjudicate）」できると Chalmers は主張。
  - これに対して著者は、意見とエビデンスとの区別は正当化できるものではなく、研究のエビデンスを「裁定者」の役割を果たすものとして扱うべきではないと考えるいくつかの理由を論じる。

### 研究が果たす実際の役割（pp. 88-89）

- Chalmers は実践が研究のエビデンスに基づくべきだという信念をもっているが、この考えはミスリーディング。なぜなら、そうした実践は必然的に（研究だけでなく）様々な情報源からの情報を組み合わせた判断の問題となるから。
  - つねに正しいエビデンスというのは存在しないので、政策形成者や実践者は、研究報告書（さらには系統的レビュー）の主張を批判的に評価することが求められる。その際、彼（女）らは自らの経験や背景知識を当てにせざるを得ない。
  - 集団についての研究エビデンスをどのようにして個別ケースに適用するか、研究エビデンスと他の情報源からの情報が矛盾する場合に両者をどのように比較検討するかという問題もある。重要なのは、研究のエビデンスが、最良の実践についての競合する考え方を裁定する機能を果たすことはできないという点。なぜなら、実践者が優れた実践に従事するうえで必要なすべての情報を研究がもたらすことはできず、また研究の知見はつねに解釈され、ときに誤解される可能性があるから。
  - さらに、実践的判断が依拠する様々な情報源はしばしば通約可能（commensurable）でなく、同じ尺度でその重要性をはかることができない。
- 実践者の判断が果たす役割や、研究以外の情報源の貢献は、実践のタイプの違いに応じて大きく異なる可能性がある。
  - 例えば医療では、教育やソーシャルワークの実践よりも研究のエビデンスが大きな役割を果たす可能性がある。
  - 実践者が取り扱う問題の性質や、実践者が活動する環境の性質によって、ある領域において実践的な判断が果たす役割は他の領域よりも重大になりうる。
- Chalmers は実践者の偏見や独りよがりの側面に目を向ける。そうした危険性はつねにあるが、これはたんに悪しき実践が行われているということにすぎない。
  - 実践者が個人的な経験や過去に教わったことに依拠するのは必ずしも無批判であることを意味しないし、その判断が必ずしも独善的であるわけでもない。

- もしそうした悪しき実践が一般的だとしても、実践者に対して最新の研究エビデンスに精通し、それに基づいて決定を行うよう求めることが問題を改善するとは考えられない。そのエビデンスと、エビデンスが他の情報との関係で実践上の決定にもつ含意を賢明に評価する場合のみ、健全な判断を行うことができる。

### 研究に基づくエビデンスの性質（pp. 89-93）

- Chalmers は、研究によって、実践者の経験に基づく場合よりも妥当性の高い実践的助言が生み出されるとする。しかし、この考えには以下のような理由から疑念がある。
  - あらゆる研究が誤りを免れない。他のあらゆる実践と同様に、研究もまた判断や解釈に依存している。
  - 研究者の信念といった外的な要因のバイアスを防ぐためにどんな試みをしたとしても、歪みがないことを保証することはできない。
  - 人間の社会関係を研究する場合に独特の、妥当性確保の難しさがある。
  - 以下、研究のエビデンスにまつわる様々な妥当性確保の困難について、ランダム化比較試験や系統的レビューとの関連で見えていく。
- Chalmers は、「何が有効か」について確実なエビデンスを生み出すランダム化比較試験の性能を強調する。しかし、ランダム化比較試験の実行とその結果の解釈は、Chalmers が示唆するほど簡単ではない。事情は分野ごとに異なる。
  - 治験の場合、特に外的妥当性の確保に困難がある。なぜならサンプルはより広範な母集団を代表しておらず、したがって何が、誰にとって有効かといった点がわかるとは限らない<sup>4</sup>。
  - 治験の場合、通常は処置が標準化可能で、場合によってはアウトカムをほとんど誤差なく測定できる。一方、他の領域では同じように処置を標準化することができず、アウトカムについて信頼性の高い測定ができない（例：ある教育手法が子どもの学習にもたらす影響）。これは、ランダム化比較試験を実施するうえでの困難というより、研究対象の性質に起因する問題である（例：教師が有効に仕事を行うためには、子どもたちの特徴といった状況に適応することが必要なため、教師の振る舞いを容易には標準化できない）。
- Chalmers はレビューの作業に科学的手法を適用したものとして系統的レビューを提示。Chalmers はこの手法が手続き上の客観性に基づいていると解釈する。しかしこれは、明確なルールによって研究実践が「透明化」されたり、その結果の妥当性が保証され

<sup>4</sup> 次の文献などが参照されている。Dehue, T., 2002, "A Dutch Treat: Randomized Controlled Experimentation and the Case of Heroin-Maintenance in the Netherlands," *History of the Human Sciences*, 15(2): 75-98.

たりすることはないという、我々が自然科学の性質について学んできたことや、社会科学において発見されてきたことの多くを無視している<sup>5</sup>。

- ある程度まではガイドラインに従うことで科学的研究の質は改善するかもしれないが、それを超えると方法論的なルールが適切な判断なしに適用されることで、むしろ害をもたらす可能性がある。
- 系統的レビューの手法に関する文献はレビューのプロセスについて単一の考え方、すなわち政策や実践の機能についての結論の確かさを最大化するために、多数の研究のデータや知見をプールすることを促す傾向にある。しかし、研究の知見を統合するうえで有用な方法はこれだけではなく、またこの種の統合がレビューの唯一の（あるいは最も重要な）目的ではない。
  - 従来のレビューでは、同じ現象の異なる側面についての知識を提供する補完的な研究をまとめ上げることも等しく重要。
  - また、レビューの目的がある領域で確立された知識とされているものを示すことなのであれば、関連する研究を網羅的に検索する必要はない。たとえ網羅的な検索が必要な場合でも、そこにはコストがかかり、ある特定の研究を読みそれを評価するのに割ける資源が少なくなることを考慮しなければならない。
- 系統的レビューにおいて関連する研究の妥当性を評価する際にしばしば取られるアプローチにもいくつかの批判が向けられる。
  - ランダム化比較試験が最良のエビデンスを提供するとし、質的な事例研究で得られるエビデンスが最も劣位に置かれる傾向にある。しかし、これらの方法から得られるデータはそれぞれに異なった妥当性確保の困難さを孕む。
  - 研究を評価する際には標準化された規準を用いるべきだとされがちだが、その研究がレビューにおいて果たす機能に応じて異なる規準を用いることが必要なときもある（例：レビューの目的が健全な知識を特定することではなく、将来的に有望な研究の方向性を示すことでもある場合）。
  - 知見の妥当性は、そこで用いられたリサーチデザイン、具体的には処置以外の要因が統制されているかどうかに基づき評価できると考えられることが多い。しかし、知見の妥当性を評価するうえで考慮すべき重要な点はこれだけではなく。例えば、すべての研究が理論をテストすることに関心があるわけではない。記述的な研究においては、対照比較を行う必要はないかもしれない。ここでの指摘は、統制が重要ではないということではなく、それがつねに知見の妥当性を最大化するわけではなく、妥当な知見を生み出すうえでつねに必要なわけでもないということ。

---

<sup>5</sup> この議論をさらに発展させたものとして次の文献が参照されている。Hammersley, M., 2002, "Systematic or Unsystematic: Is that the Question? Some Reflections on the Science, Art, and Politics of Reviewing Research Evidence," Paper presented to the Public Health Evidence Steering Group of the Health Development Agency.

- 以上のように、専門的な実践の性質と研究の性質の両方を踏まえると、Chalmers や他の論者のように実践者の意見と研究のエビデンスの間にはっきりとした区別を設けることに疑問を呈することが重要。
  - エビデンスに基づく実践の擁護者は、知識と判断の明確な区別を主張する点で、ある種のポストモダニストとよく似た立場に立つ。一方は、主観的判断を矯正する手段として科学的知識の役割を過剰に強調し、他方は、科学をイデオロギー的で抑圧的なものとして拒絶し、恣意的な判断とみなされるものに依拠するほうを好む。
  - 両者はともに知識と判断の明確な区別を主張するが、より賢明なのは、ときに通約不可能で、実践的な判断が必ず伴う様々なエビデンスを認識するアプローチである。

#### 批判に対する Chalmers の応答 (pp. 93-94)

- Chalmers は、エビデンスに基づく実践を擁護する他の人々と同様、ランダム化比較試験や系統的レビューの特権化を批判する人々の指摘に真剣に向き合っていない。
  - エビデンスに基づく実践を擁護する人々は、批判が「政治的」であるとか「イデオロギー的」であるとか「ポレミック」であるとしてそれらをはねつける。
  - Chalmers は、ランダム化比較試験に限界があると示すことは、ランダム化を拒否し、自分自身が全知全能であると信じていることを意味するという懸念をもっているようである。批判を向ける人々を「エビデンスを無視」し「現実に向き合っていない」として非難する。
- Chalmers は、著者の Hammersley への応答の中で、著者が書いたことを歪曲している。
  - Chalmers は、著者である Hammersley が「バイアスは最小化されねばならないという考えを否定」していると述べたうえで、「レビューにおけるバイアスについての懸念が傲慢にも欠如している」と明言する。
  - 著者の主張は、系統的レビューには実証主義的な前提、すなわち主観性はバイアスの発生源であるため、レビュープロセスの手続き化（proceduralising）を通じて最小化されねばならないという前提があるというものだった。著者は、この前提が誤りであり、主観性がつねにバイアスの発生源となるわけではないこと、手続き化された研究がつねに妥当な結果を保証するわけではないことを主張した。
  - 著者は、妥当な結論を得る可能性を高めるためにバイアスが最小化されねばならないという考えを否定していない。むしろ、バイアスの最小化は科学的探究の中核にあり、レビュープロセスに必要不可欠な要素である。著者が否定したのは、主観性の最小化が必要もしくは望ましいという考え方。

## 結論（pp. 94-95）

- Chalmers やエビデンスに基づく実践を擁護する論者によって提出された議論には、いくつか重要なものもある。しかし、その結論は、実践者の意見と研究のエビデンスの間の明確な区別に基づいているために、かなりミスリーディングなものとなっている。
  - このため、政策形成や実践を導くうえで研究が果たしうる役割について行き過ぎた主張がなされている。
- ランダム化比較試験や系統的レビューの実施には付随的なコストが発生する可能性があり、それ自体が害をもたらす可能性もある。この点を Chalmers は無視している。
  - 政策形成者や実践者が研究の知見を過剰に重視するよう促された場合、価値あるはずの政策や処置が用いられなかったり、政策や実践が実際よりも信頼のおけるものとして扱われたりする可能性がある。
  - ここでの著者の主張は、研究は誤る可能性があるため、その価値が損なわれるということではない。そうではなく、研究がもたらす利益についての Chalmers の一方的な説明に対してバランスをとる必要があるということ。
- また Chalmers は、ある種の研究が政策や実践の問題を直接的に解決できるかのように述べており、価値判断を必然的に伴う実践的な判断に対する技術的な解決策があるかのように述べている点も問題である。
  - 医療のいくつかの分野では、価値や目標の競合が生じないために、この点はそれほど問題にならないかもしれない。
  - しかし、例えば文章読解の様々な指導法を評価する場合、それらの方法が目指すものが同じとは限らず、そうした差異は教育的価値などの違いによって生じる側面がある。こうした状況において、「何が有効か」を表すものとして研究をみなすことは、異なる教育目標に与えられるべき相対的な重みについて適切に議論する可能性を閉ざすイデオロギー装置として機能しかねない。