



RIETI Policy Discussion Paper Series 26-P-004

EBPM（エビデンスに基づく政策形成）入門

序論 EBPMとは何か、なぜ必要か？

第1話 本来のEBPMの概要と経緯

関沢 洋一
経済産業研究所



Research Institute of Economy, Trade & Industry, IAA

独立行政法人経済産業研究所

<https://www.rieti.go.jp/jp/>

EBPM(エビデンスに基づく政策形成)入門¹

序論 EBPM とは何か、なぜ必要か？

第1話 本来のEBPMの概要と経緯

関沢洋一（独立行政法人経済産業研究所）

要 旨

- ・ EBPM は、個々の政策介入に効果があるかどうかを可能な限り厳密に検証して、その検証の結果として得られる証拠（エビデンス）を政策に反映させるものであり、エビデンスを用いた政策の品質管理である。
- ・ 効果があるという証拠を得るためには因果関係の検証が必要になる。
- ・ EBPM という言葉は 1990 年代のイギリスで登場したが、実際には 1960 年代からアメリカではプログラム評価として多くの社会実験が行われ、EBPM に近いことが行われてきた。
- ・ 日本では 2016 年頃から EBPM を推進し始めたが、効果の検証よりもデータ整備やロジックモデルの作成が重視され、本来の EBPM から乖離した。
- ・ 本稿が提示する EBPM2.0 は、新・三本の矢から構成されており、それらは①因果推論に基づく政策効果の厳密な検証と検証結果の政策への反映（狭義の EBPM）、②セオリー評価に基づく政策の確からしさの向上（準 EBPM）、③業務情報へのアクセスの拡大からなる。

キーワード：政策介入、効果、因果関係、エビデンスに基づく医療（EBM）、「何に効果があるか（what works）」、ランダム化比較試験（RCT）

JEL classification: H11 H43

RIETI ポリシー・ディスカッション・ペーパーは、RIETI の研究に関連して作成され、政策をめぐる議論にタイムリーに貢献することを目的としています。論文に述べられている見解は執筆者個人の責任で発表するものであり、所属する組織及び（独）経済産業研究所としての見解を示すものではありません。

¹ 本稿の原案は、経済産業研究所（RIETI）のポリシー・ディスカッション・ペーパー検討会で発表を行ったものです。検討会参加者からの有益なコメントに感謝申し上げます。

序論 EBPM とは何か、なぜ必要か？

お役所（日本政府や地方公共団体）は政策という名の下に様々な取り組みを行っています。それらはお金の提供（給付金や補助金など）、課税やその減免、公的融資や債務保証、法的規制、行政指導、啓発活動や情報提供、などなど多岐に渡っています（以下では「政策介入」と呼びます）。

現実に行われているほとんどの政策介入に効果がある、つまり、それぞれの政策介入はそれが目指す目的の達成に貢献していると、多くの人々は漠然と思っているかもしれません。また、政策の立案や執行に当たる人々は個々の政策介入に効果があることを当たり前の前提として行動してきました（無謬性）。

ところが、実際には政策介入には効果がなかったり乏しかったりすることが頻繁にあります。「地獄への道は善意で舗装されている（The road to hell is paved with good intentions）」ということわざがありますが、関係者が社会を良くしようという真面目な意図で政策介入を立案して実施している場合でも、実際には良くならない場合が多いのです。

効果のない政策が 8 割？

日本では政策介入の効果を実際に検証した例は少ないのですが、他の国々、特にアメリカでは多くの効果検証がプログラム評価として行われてきました（プログラムは政策介入とほぼ同義）。

様々なプログラム評価の結果を踏まえて、政策介入のうち約 80% は効果がないことを意味する「80%ルール」を提案した論文があります²。一見極端に見えますが、更に極端な指摘として、プログラム評価の大家であるロッシは、全てのプログラムの効果の期待値は厳密な検証を行うほどゼロに近づくと唱えました³。

また、社会学者のマーティンソンは、犯罪履歴のある人々や犯罪に手を染めるリスクの高い人々に対する矯正的活動について「何をやっても効果がない（Nothing Works）」と指摘し、多くの人々から支持されてきました⁴。

ロッシやマーティンソンは自身の指摘が行きすぎていることを後に認めています⁵、多くの政策介入には効果がないことは否定しがたいように思えます。

² White (2019)

³ Rossi (2003)

⁴ Martinson (1974)。Alsan et al. (2025)の最近の研究では効果があるものが存在することを指摘しています。

⁵ Martinson (1978); Rossi (2003)

以上の政策介入との対比として医薬品について考察します。医薬品においては効果がない場合（たとえば薬を飲んでも効かない場合）は一般的な政策介入に比べれば少ないと思います。なぜそう言えるかという、医薬品はランダム化比較試験（RCT, Randomized Controlled Trial）という実験によって効果と安全性を示せなければ市場に投入することへの承認を政府から得られないからです。ある研究によると、3段階ある実験を経て承認に至る医薬品の割合はわずか10%なのだそうです⁶。市場に投入される医薬品について事前に厳密な品質チェックが行われているために、市場に投入される段階では相当信頼できるものとなっています。

医薬品と異なり、政策介入においてはRCTのような実験によって事前に効果のチェックが行われることはめったにありません。つまり政策介入のほとんどは品質チェックが行われないうちに実社会において実行に移されます。せめて、実社会で実行に移された政策介入について実際に効果があったかどうかの検証が事後的に行われればいいのですが、そのようなこともめったにありません。

この点はアメリカと比較した場合の日本においては特に顕著です。アメリカでは社会実験としてRCTによる政策介入の効果検証が頻繁に行われ、多くの政策介入に効果があるのかが明らかにされてきました。日本ではこのような実験はほとんど行われてきませんでした。

この結果として、日本では、大部分の政策介入は効果がなく無駄がはびこっていることが想像されるものの、実際にどの政策介入に効果があってどれに効果がないかがわからなくなりました。

EBPM とは何か？

EBPM（エビデンスに基づく政策形成, Evidence-Based Policy Making）はこのようなやり方を改めて、個々の政策介入に効果があるかどうかを可能な限り厳密に検証して、その検証の結果として得られる証拠（エビデンス）を政策に反映させようとするものです。EBPM はいわばエビデンスを用いた政策の品質管理です。

とはいえ、全ての政策介入を実行に移す前提条件としてRCTのような実験の実施を求めるといふ医薬品みたいなことは現実にはできません。それでも、エビデンスを得ることを容易にする2つの動きが生じています。

第1に、因果推論と呼ばれる分野の研究の進展です。政策介入の効果の有無を明らかにすることは因果関係を特定することに近く、そのためのほぼ唯一の手段がRCTでした。ところが、この数十年の間に、RCTのような実験に頼らなく

⁶ Mullard (2016)

ても因果関係の検証を行える手法の開発が学際的に進みました。この結果、政策介入の効果を検証できる場合が着実に増えています。

第2に、分析に利用可能なデータが増大しました。その背景としては ICT（情報通信技術）の進歩やそれに伴うビッグデータの登場に加えて、各国政府が保有する業務情報（日常業務を遂行する過程で役所に集まってくる情報）の研究目的での利用が拡大していることがあります。

以上の結果、RCTのような実験に頼らなくても、行政機関が既に保有している業務情報など既存のデータを活用して政策介入の効果についての因果関係を検証できる場合が増えてきており、政策立案に貢献できそうな信頼できるエビデンスを得られる機会は着実に増加しています。このため、本格的な EBPM を日本でも推進することを可能とする環境が整いつつあります。

本入門の目指すもの

これまでに多くの政治家や行政官が様々な政策介入を実施に移し、自らの業績として誇示してきました。しかし、政策介入は実現するだけでは意味がなく、効果があつて初めて意味があります。効果検証がないままに効果があるかないかわからない政策介入（その多くは効果がないと予想されるもの）を漫然と続けることはリソースの無駄遣いにつながります。

特に、日本の場合には、少子高齢化に伴う人手不足により、利用可能なリソースの制約が日毎に高まっています。政府が講じる諸政策も例外ではなく、本当に効果がある政策介入を見極めた上でそこへの選択と集中が必要になります。EBPM は政策の品質管理の手段としてこのような選択と集中を促すポテンシャルを有しています。

この EBPM 入門では EBPM を巡る重要なポイントについて数回に分けて述べることにします。

引用文献

Alsan, M., Barnett, A., Hull, P., & Yang, C. S. (2025). "“Something Works” in U.S. Jails: Misconduct and Recidivism Effects of the IGNITE Program," *The Quarterly Journal of Economics*, 140(2), 1367-1415.

Martinson, R. (1974). "What works? Questions and answers about prison reform," *Public Interest*, 35, 22-54.

Martinson, R. (1978). "New findings, new views: A note of caution regarding sentencing reform," *Hofstra Law Review*, 7, 243.

Mullard, A. (2016). "Parsing clinical success rates," *Nature Reviews Drug*

Discovery, 15(7), 447-447.

Rossi, P. H. (2003). "*The iron law of evaluation reconsidered.*" Paper presented at the Remarks presented at the 2003 APPAM research conference, Washington, DC.

White, H. (2019). "The twenty-first century experimenting society: the four waves of the evidence revolution," *Palgrave Communications*, 5(1), 47.

第1話 本来の EBPM の概要と経緯

(要約)

- ・ EBPM は、個々の政策介入に効果があるかどうかを可能な限り厳密に検証して、その検証の結果として得られる証拠（エビデンス）を政策に反映させるものである。
- ・ 効果があるという証拠を得るためには因果関係の検証が必要になる。
- ・ EBPM という言葉は 1990 年代のイギリスで登場したが、実際には 1960 年代からアメリカではプログラム評価として多くの社会実験が行われ、EBPM に近いことが行われてきた。
- ・ 日本では 2016 年頃から EBPM を推進し始めたが、効果の検証よりもデータ整備やロジックモデルの作成が重視され、本来の EBPM から乖離した。
- ・ 本稿が提示する EBPM2.0 は、新・三本の矢から構成されており、それらは①因果推論に基づく政策効果の厳密な検証と検証結果の政策への反映（狭義の EBPM）、②セオリー評価に基づく政策の確からしさの向上（準 EBPM）、③業務情報へのアクセスの拡大からなる。

政策介入の効果とは何か？

EBPM は、個々の政策介入に効果があるかどうかを可能な限り厳密に検証して、その検証の結果として得られる証拠（エビデンス）を政策に反映させるものです。

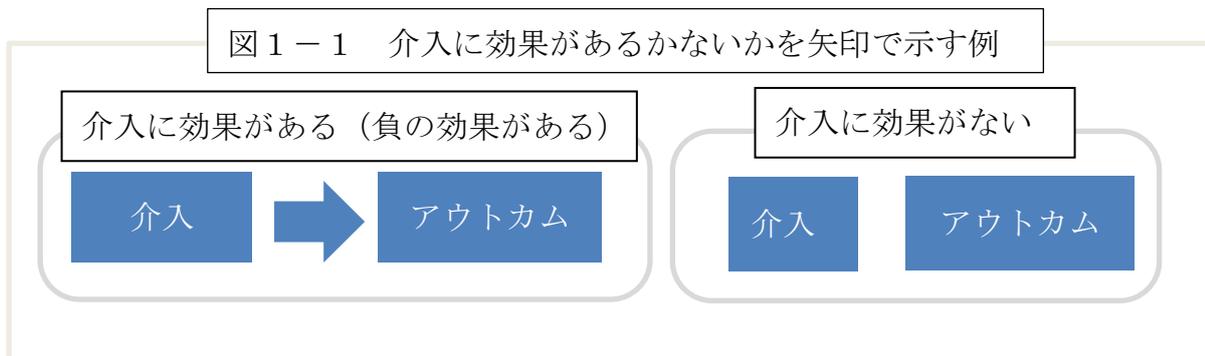
「政策」や「政策介入」の定義は 2001 年に成立した政策評価法（行政機関が行う政策の評価に関する法律）が参考になります。同法では政策の定義として「行政機関が、その任務又は所掌事務の範囲内において、一定の行政目的を実現するために企画及び立案をする行政上の一連の行為についての方針、方策その他これらに類するもの」としています。また、同法では政策効果として「当該政策に基づき実施し、又は実施しようとしている行政上の一連の行為が国民生活及び社会経済に及ぼし、又は及ぼすことが見込まれる影響」と定義しています。本稿での政策介入はこの定義中の「当該政策に基づき実施し、又は実施しようとしている行政上の一連の行為」とほぼ同義で、個々の政策に基づいて実施される様々な取り組みを指します。

ここで重要なことは政策介入の効果を見極めることですが、「効果」とは何でしょうか。キーワードは因果関係です。

本稿では、特定の目的をもって行われる何らかの介入がその目的に照らして望ましい結果を生じさせる場合に「効果がある」という表現を用い、悪い結果を生じさせる場合には「負の効果がある」という表現を用い、介入が何らの結果も

もたらさない場合は「効果がない」という表現を用います。

何らかの介入がある場合、その目的の達成度を明らかにする何らかの数値が存在しており、これはアウトカムと呼ばれます。介入がアウトカムに対してプラス（マイナス）の影響を及ぼしているという因果関係があれば、その介入に効果（負の効果）があることとなります。図1-1の左側はこれを指します。矢印は因果関係を示します⁷。これに対して、図1-1の右側は介入とアウトカムの間に関係がありません（矢印がない）。つまり、その介入はその目的に照らして役に立っておらず、効果がないこととなります。



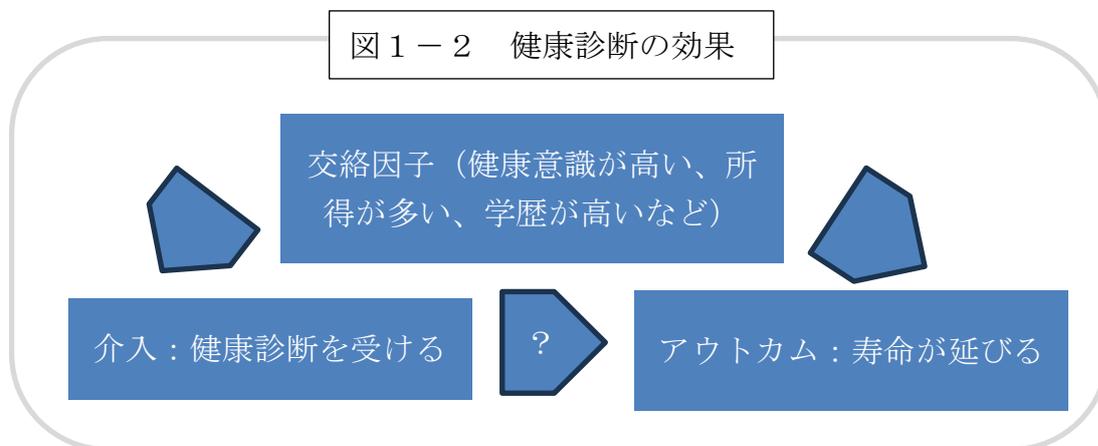
私たちが知りたいのは図1-1において、介入とアウトカムの間に因果関係があるか（矢印で結びついているか）どうかです。言い換えると、その介入がなかった場合に比べて介入があったことによってアウトカムについてどのような違いが生じたのかが知りたいことです（反実仮想 [counterfactual] と呼ばれます）。歴史に **IF**（もしも）は禁物と言われますが、反実仮想で探求するのは **IF** です。しかし、タイムマシンがない以上、その介入が行われた場合と行われなかった場合を直接比較することはできません。

これに対する対応策は、介入を受けた人々と受けなかった人々で比べることです。介入を受けた人々を介入群、受けなかった人々を対照群と呼びます。

ただ、実際には介入群と対照群の単純比較ではうまくいかないことがほとんどです。たとえば、健康診断を受けることによって寿命が延びるかどうかを検証したいとします。健康診断を受ける人々と受けない人々を単純に比べると、2つの集団の間に元々の違いがあります。たとえば、健康診断を受ける人々は受けない人々に比べて、健康意識が高い、所得が多い、学歴が高いなどの違いがあります。こうした要因は寿命の延びにつながるものが想定され、健康診断を受ける人々が受けない人々より長生きしても、健康診断を受けたから寿命が延びるという因果関係は存在しない可能性があります。図1-2はこのような

⁷ 図1-1は因果ダイアグラムと呼ばれるもので、矢印は介入がアウトカムに及ぼす効果があることを示します(Pearl & Mackenzie, 2018)。

状況を示します。介入とアウトカムの双方に影響を及ぼす要因は交絡因子と呼ばれます。



単純比較ではうまくいかない場合、健康診断を受けることによって寿命が延びるかどうかについての因果関係を明らかにするためにはどうしたらいいでしょうか。より一般的に、介入がアウトカムの改善につながっている（介入に効果がある）かどうかを知るためにはどうしたらいいでしょうか。

この問いに対処することは簡単ではなく、数少ない方法として古くから行われてきたものは、ランダム化比較試験（RCT）と呼ばれる実験です（第2話で詳述します）。一見難しい言葉ですが、その本質は小学校の理科で行う実験と同じです。大豆の発芽などで条件を1つだけ変えて発芽するかどうかを検証するという実験です。ただし、主たる実験対象が大豆ではなく人間になります。

人間の場合、ある程度の人数の研究参加者を募った上で、コイントス（またはそれに類するコンピューターによる割付け）で表が出れば介入群、裏が出れば対照群とすれば、いわば神の見えない手に導かれて両群はほぼ属性が同じ集団になり、比較可能になります。これがランダム化です。上記の健康診断の例ではコイントスで決まった介入群にのみ健康診断を受けてもらえば効果検証が可能になります。健康診断のRCTについては第5話で述べます。

政策効果の検証のための社会実験に熱心なアメリカ

RCTは主として医療で行われてきました。重要なきっかけは、サリドマイドという医薬品が胎児の奇形を引き起こすという薬害事件を踏まえて、アメリカが1962年に行った立法措置です。この立法では、製薬会社による新しい医薬品の市場への投入を食品医薬品局（FDA）が承認するのに先立って、効果や安全性についての実質的な証拠を考慮することをFDAに求め、この証拠を決める上で

RCT が行われるべきことが固まってきました⁸。このことは製薬会社が RCT を行う強力なインセンティブとなり、多数の RCT が行われることとなりました⁹。これは後のエビデンスに基づく医療（EBM, Evidence-Based Medicine）の基礎となりました。

1960 年代は EBM だけでなく EBPM の実質的な開始時点とみることもできます¹⁰。当時のアメリカではジョンソン政権により「偉大なる社会」という壮大な構想が進められ、貧困から脱出することを目指す様々なプログラムが実施に移されました¹¹。これらのプログラムの効果を評価することが重要とされ、多くのプログラム評価が行われ、その手法として RCT が用いられました。

RCT による効果検証の結果がその後の政策のあり方につながった例として、1960 年代に行われたペリー就学前プロジェクトがあります。この研究では幼児教育を受けることが就業後の所得増加につながることを明らかにしており、その後の幼児教育プログラムの推進に影響を及ぼしています¹²。

1960 年代終わりに開始された負の所得税（所得が基準額を下回ると、基準額と実際の所得の差額に応じて政府により所得が補填されるもので、給付付き税額控除の原型）の影響を調べる目的で行われた複数の RCT では、途中段階の結果を暫定的に示すべきという圧力が 1970 年代初頭に政治家から研究者にあり（要望に応じて示したら今度は「時期尚早だ」と批判されたそうです）、当初は人々の関心も高かったものの、結果が出る 1980 年頃には人々はこれらの RCT に対する関心を失っていました¹³。

アメリカの社会福祉における RCT は 1980 年代に進展しました¹⁴。代表例として、母子家庭に対する支援措置のあり方についての RCT があります。母子家庭に対する支援措置は措置が作られた頃は母親が未亡人である場合が多く支援措置は金銭的援助によって行われていました。ところが、離婚や未婚による母子家庭が増えて支援を受ける人々の数が増大するにつれて、従来のやり方に疑問が呈されるようになりました。RCT によって新しい試みを実験する州に対して保健福祉省が一部の法律の適用を免除することとしたこともあって（ウェイバ

⁸ Meldrum (2000)

⁹ Baron (2018)

¹⁰ 津田・岡崎 (2018)

¹¹ ワイス(2014)では、プログラムは「人々、団体、組織やコミュニティの厚生（well-being）を向上させるための構造化された介入行為」（P440）と定義しています。

¹² Elango et al. (2016)

¹³ Coyle and Wildavsky (1986); Ventry (2000); Widerquist (2019)

¹⁴ Baron (2018); Bernecker et al. (2021); Gueron (2017); Moffitt (2004); 小林 (2024)

一制度)、多くの州で実験が行われました。それらによって、支援期間に限定がない従来型の支援措置よりも、支援期間を限定して仕事を見つけることを促す方がその後の雇用の確保、収入の高まり、福祉への依存度の低下につながる事が明らかになってきました。このことは1996年の立法によって母子家庭への支援を福祉型から就業促進型へと転換することへの後押しになりました。

ただ、多くの場合にはプログラム評価の結果は埋もれてしまって政策立案者には利用されず、政策形成には貢献しなかったことが指摘されています¹⁵。第9話で改めて触れる予定です。

1990年代の転機

医療においては、医薬品についてはRCTによる事前のチェックが行われていたものの、実際の診療においては、医師を始めとする専門家の経験を偏重するところがありました。これに対して、現時点における最良のエビデンスを医師や患者の意思決定に役立てようという動きが1980年代頃からあり、これが1992年にエビデンスに基づく医療(EBM)として宣言されました¹⁶。

EBMに含まれる「エビデンスに基づく(evidence-based)」という言葉は、医療にとどまらず、教育、犯罪予防など範囲が広がっていき、さらには政策一般にも及ぶこととなりました。1997年に発足したイギリスのブレア政権において「エビデンスに基づく政策(Evidence-based Policy)」という発想が登場し、「重要なのは効果があること(“What matters is what works.”)」という標語が唱えられるようになりました¹⁷。“what works”はEBPMを貫くキーワードです¹⁸。

アメリカではプログラム評価の結果が体系的に使われないという時代が続きましたが、21世紀に入るところから政策形成におけるエビデンス重視という発想が進展してきたように思います。2001年に始まるブッシュ政権で働いた経験のある経済学者のジョン・リストによれば、同政権において行われてきた正しいことでほとんど評価されていないことの1つとして、同政権はエビデンスに基づく政策を推進しており、意思決定を支援する研究が求められたそうです¹⁹。

また、21世紀に入ってからアメリカでは教育におけるエビデンスの取り込みが進むようになりました。2002年1月に制定された「子供を一人も落ちこぼ

¹⁵ Weiss (1988); 益田(2019)

¹⁶ Evidence-Based Medicine Working Group (1992)

¹⁷ Sanderson (2002)

¹⁸ “what works”の訳として、本稿では「効果があること」、「何に効果があるか」など「効果」という言葉を用いています。

¹⁹ リスト(2023)

れにしないための 2001 年法」(NDLB 法)に基づいて科学的根拠のある研究に基づく施策の採用が求められることとなり、what works clearinghouse (WWC) が作られました²⁰。WWC は教育分野における広範囲のプロジェクトについて効果のエビデンスを提供しています²¹。また、1998 年の議会からの要請に基づいて、低所得層の幼児を対象とした教育などの包括的な支援プログラム(ヘッドスタート)の大規模な RCT による効果検証が行われており、2010 年に最終報告が出されました²²。

オバマ政権における EBPM

2009 年からのオバマ政権においては EBPM が積極的に推進されました²³。オバマ政権内には RCT の推進者が多く、オバマ大統領自身が、社会プログラムは厳密なエビデンスに基づくものを中心に行われるべきという考え方の持ち主でした。私が考える EBPM の本質はオバマ大統領が 2014 年に行った以下の演説に示されています。

世の中に出回っている政策プログラムの中には、正しい意図と優れた洞察に基づいて作られていて、一見すると良くできているものの、実際には効果のないものが数多くあります。私たちには、効果のないものに投資するだけの十分なお金も時間も資源もありません。従って、あることに効果がなければそれを行うことはやめようと、頑なに言わなければなりません。効果のあることを行いましょう。それが民主党のプログラムなのか、共和党のプログラムなのか、あるいは、何らかの信条に基づくプログラムなのかは顧慮するべきではありません。効果があるのであれば、私たちはそれを支持すべきです。もしも効果がないのであれば、支持すべきではありません²⁴。

²⁰ Mosteller and Boruch (2002); 田辺(2006)

²¹ U.S. Department of Education (2023)

²² U.S. Department of Health and Human Services (2010)

²³ Haskins and Margolis (2014); Juras et al. (2019); 津田・岡崎(2022)

²⁴ 原文は以下のとおりです。"There are a lot of programs out there that sound good, are well-intentioned, well-inspired, but they're not actually having an impact. We don't have enough money or time or resources to invest in things that don't work, so we've got to be pretty hard-headed about saying if something is not working, let's stop doing it. Let's do things that work. And we shouldn't care whether it was a Democratic program or a Republican program, or a faith-based program or—if it works, we should support it. If it doesn't, we shouldn't."

<https://obamawhitehouse.archives.gov/the-press-office/2014/02/27/remarks->

オバマ政権では、十代の妊娠予防や幼児教育の推進など6分野のプログラムについて重点的に **EBPM** が推進されました。エビデンスに基づく補助金付与という観点から、同政権では階層別の補助金というアプローチがとられました。既にエビデンスがあるプログラムに対して金額の多い補助金を優先的に配分する一方で、改めて効果検証を実施することを求めました。ただ、このやり方を貫くとまだエビデンスがないものの有望なプロジェクトに補助金がつかなくなるため、イノベーションを促進する観点から、ロジックの堅固なプログラムには補助金をつけることとして、効果がなければ改善するか、それでもだめなら廃止するという方針がとられました。

新型コロナウイルスと **EBPM**

第1次トランプ政権下ではオバマ政権の反動とも見られる動きも見られましたが、2019年に **Evidence Act (Foundations for Evidence-Based Policymaking Act)** が制定されて議会と連邦政府の双方が **EBPM** を推進する動きも見られています。2021年からのバイデン政権は就任早々に発した文書において、**EBPM** を推進することを明確にしました²⁵。

ただ、この時期は新型コロナウイルスの蔓延時期と重なり、**EBM** や **EBPM** においては受難の時期であったと思えます。

コロナ禍は、エビデンスを重視するという評判のあった **FDA** の評価に悪影響を及ぼすことになりました。たとえば、**FDA** は抗マラリア薬のヒドロキシクロロキンを新型コロナウイルス対策薬として使用することについて緊急使用許可を出しました²⁶。トランプ大統領はあやふやな根拠に基づいてこの薬に効果があると喧伝していたのですが、**RCT** によるエビデンスが存在しないにも関わらず **FDA** が許可を出したことに批判が集中しました。後に **FDA** は、ヒドロキシクロロキンに新型コロナ対策としての効果がないとしてこの許可を撤回しました。このような **FDA** の対応はエビデンスを軽視して政治の圧力に屈したと捉えられかねず、**FDA** の評判を損なうことにつながりました。

また、コロナ禍ではエビデンスよりも支持政党やイデオロギーが重視されるという事態がしばしば見られました。ワクチン接種への態度が典型的で、効果の有無というエビデンスの問題とは別に、保守的な人々はワクチン接種をためら

president-my-brothers-keeper-initiative

²⁵ <https://bidenwhitehouse.archives.gov/ostp/news-updates/2024/09/30/biden-harris-administration-strengthens-federal-scientific-integrity-practices/>

²⁶ Lenzer (2020); Fuleihan (2023)

い、リベラルな人々はワクチン接種に積極的であるという傾向が見られました²⁷。ソーシャルディスタンスやフェイスマスクの着用でも似たようなことが見られ、共和党支持者はこれらの措置に消極的な傾向があり、民主党支持者はこれらに積極的な傾向がありました²⁸。

因果推論の進展とビッグデータ時代の到来

1990年代がEBPM推進の転機であることは上述したとおりですが、この頃から生じた特記すべき動きとして、経済学などのアカデミアやIT企業などにおける因果関係の検証に対する関心の高まりがあります。

一つ目は因果推論と呼ばれる分野の研究の進展です。RCTのような実験に頼らなくても因果関係の検証を行うための動きが学際的に進みました。因果革命²⁹や信頼性革命³⁰という言葉で語られるような、因果関係を明らかにするための研究の進歩が1990年代頃から見られました。たとえば、回帰不連続デザイン(RDD)、差の差の分析(DID)、操作変数法など様々な分析手法が進化していき、統計ソフトの進歩とも相まって、政策介入の効果検証を目的として、多くの研究者がこれらの手法を使った分析を行うようになりました(第2～4話で述べる予定)。この分野を主導したアングリスト・インベンス・カードは2021年にノーベル経済学賞を受賞しています。

また、21世紀に入る頃から国際開発分野におけるRCTを通じたエビデンス構築が加速しています³¹。貧困アクションラボ(J-PAL)、IPA(Innovations for Poverty Action)、3ie(International Initiative for Impact Evaluation)が設立され、RCTやシステマティックレビューの実施を通じて、開発途上国におけるエビデンスの蓄積やその共有が進められています。この分野を主導したバナジー・デュフロ・クレマーは2019年にノーベル経済学賞を受賞しています。

テックカンパニーと呼ばれるIT企業の因果推論の活用も目立っています。テックカンパニーはRCT(しばしばA/Bテストと呼ばれます)を活用することによって広告活動や価格戦略などの効果を検証してその改善につなげています。テックカンパニーの多くは因果推論に関する研究開発も進めており、たとえば、CausalImpactという統計分析手法はGoogleの研究者によって開発され

²⁷ Cowan et al. (2021)

²⁸ Allcott et al. (2020); Howard (2022)

²⁹ Pearl and Mackenzie (2018)

³⁰ Angrist and Pischke (2010)

³¹ 佐々木(2020)

ています³²。Uberのように経済学者と連携しながら RCT による効果検証を行う事例も見られます³³。

これらの動きと並行して、ビッグデータという言葉で象徴されるように、分析に利用可能なデータが増大しています。背景としては ICT の進歩が挙げられますが、それと合わせて、各国政府が保有する業務情報の研究目的での利用が拡大していることもあります（第6話で述べる予定）。このことは効果検証に不可欠なデータの取得の容易化を通じて EBPM を後押しすることとなりました。

以上の結果、開発途上国の大規模な RCT によるエビデンスが新たに得られることになったことに加えて、RCT のような実験に頼ることなく、行政機関が既に保有している業務情報など既存のデータを活用して政策介入の効果に関する因果関係を検証できる場合が増えてきており、新たなエビデンスを提供する研究は着実に進展しています。

日本における EBPM の導入と和風化

2016 年頃に EBPM という言葉が本格的に日本に持ち込まれました。GDP の統計の正確さを巡る政府内の議論に端を発して、EBPM という名の下で統計改革が進められることとなり、更に EBPM の「三本の矢」という名の下で、経済・財政再生計画の点検・評価、政策評価における取り組み、行政事業レビューにおける取り組みが進められました³⁴。

ただ、ここでいう EBPM は今まで述べてきたものとは異なっており、日本の EBPM は独自のものとなりました³⁵。私はこれを和風 EBPM と呼んでいます。

アメリカの EBPM は社会実験による効果検証の結果を政策立案の参考にしようという考え方に基づくものですし、イギリスも *what works* という言葉に示されるとおりどのような政策介入に効果があるかを探求していくものです。これに対して、日本では統計改革のために EBPM という言葉が使われ、また、政策評価法に基づく政策評価³⁶と行政事業レビュー³⁷という既存の枠組み

³² Brodersen et al. (2015)

³³ Athey, Castillo, and Chandar (2025)

³⁴ 杉谷(2022)

³⁵ 小林(2025); 杉谷(2024); 田辺(2020)

³⁶ 政策評価法に基づく政策評価は、2010 年頃からは数値目標のモニタリングを中心とする業績管理となっており、本来の EBPM に近いプログラム評価とは乖離したものになっています(杉谷, 2024)。

³⁷ 内閣官房の行政改革推進本部のホームページでは「行政事業レビューとは、各府省庁が、予算や基金を用いて行う原則すべての事業を、毎年度、自ら点検

の中に **EBPM** が吸収されることになりました³⁸。

EBPM として唯一新しく日本国内で広まったのはロジックモデルの作成ですが、ロジックモデルは後述するセオリー評価に基づくもので³⁹、**RCT** を始めとする因果推論を基礎とする **EBPM** とは厳密には異なる上に、仏を作って魂を入れないようなロジックモデルが量産されています。

もともと、アメリカやイギリスに見られた因果推論を基礎とする **EBPM** も日本に持ち込まれました。日本政府からアメリカやイギリスに留学生として派遣された若手の行政官の中には留学時の経験によって **EBPM** に関心を抱いた人々がいて、彼らを中心として研究者も巻き込んで日本にも因果推論に基づく **EBPM** を根付かせようとする動きがありました⁴⁰。この動きは、私が在籍する独立行政法人経済産業研究所（RIETI）の活動に反映され、2017年2月に **EBPM** の研究を目的とする研究プロジェクトが立ち上げられて、本稿執筆時点においても続いています（リーダーは当初は山口一男氏、後に大竹文雄氏）⁴¹。

日本では **EBPM** が推進されていることになっていますが、上記の事情により、**EBPM** とは何かという肝心な点についての理解が人々によって異なり、本稿執筆時点でもその状況は続いています。大きく分けると、**EBPM** について、①行政の現場にいる人がロジックモデルを作ることによって政策の合理化を図ることだと考えている人々、②政策介入に効果があるかどうかについての因果関係を統計解析によって明らかにしてその分析結果を用いて政策の合理化を図ることだと考えている人々、③データや統計を整備したり見える化したりすることだと考えている人々、の3つの人々になります。①の中心は政策評価を研究対象としてきた研究者（行政学者など）や、総務省や行政改革推進本部など日本政府内で **EBPM** を推進しようとする人々です。②の中心は統計的因果推論についての知見を有する経済学者を中心とする研究者や一部の若手官僚です。③の中心は日本の諸官庁で働く人々の多くです。

もう一つ指摘しておくべき点として **EBPM** 推進への行政の現場の反発があります。政策評価法に基づく政策評価や行政事業レビューなどで、意味があるかないかわからないような評価活動に忙殺された上に、**EBPM** まで担がせられることに対しては行政の現場で抵抗感があり、「評価疲れ」と呼ばれることに

し、その点検結果を公表する取組です。」となっています。

³⁸ 杉谷(2022)

³⁹ 龍・佐々木 (2004)

⁴⁰ 代表的なものとして、シカゴ大学公共政策大学院に留学した日本人による有志の勉強会があります(近藤, 2022)。

⁴¹ 大竹・内山・小林編著 (2022)

なりました。加えて、担当する政策介入の本当の効果が明らかにされることへの不安がしばしば行政の現場から本音として示されました。

EBPM2.0 と新・三本の矢

本稿では EBPM とは政策評価の一種で、狭義の EBPM と準 EBPM の 2 つから成り立っていると整理します。狭義の EBPM は、実際に行われている政策介入がその目指す目的に近づくことに貢献しているかどうかという因果関係を検証して、その分析結果として明らかにされたエビデンスに基づいて、政策のあり方を検討していくことです。こちらは基本的には政策介入が実施に移された後の事後評価になります。狭義の EBPM は因果推論における近年の学際的な研究成果を反映したものです。

準 EBPM は、政策介入の立案段階において、政策資源（予算、人員など）の投入が最終的な目標の改善に結びつくかどうかについて、合理的で説得力のある仮説が形成されているかどうかをできるだけ厳密に明らかにしていくことです。準 EBPM はセオリー評価と呼ばれるもので⁴²、インプット→アクティビティ→アウトプット→アウトカム→インパクトという一連の流れを明らかにするとともに、この矢印がどうして（why）、どのようにして（how）存在するかを明らかにすることを指します。セオリー評価におけるセオリーとは仮説のことです⁴³。仮説そのものは証拠にはなりえないので、セオリー評価に基づく政策立案は EBPM とはせずに、準 EBPM としました。

EBPM を構成するもう 1 つの要素として、行政機関が保有する業務情報へのアクセスの拡大があります。狭義の EBPM はデータがないと絵に描いた餅になります。準 EBPM においても仮説形成に当たってはデータに基づくファクトの確認が重要になります。多くの場合、データは存在するにも関わらず役所の中に眠っていて、研究者が利用できない状況になっており、その点を改善していく必要があります。

以上をまとめると、2016 年頃に日本で登場して推進されるようになった EBPM を EBPM1.0 としたときに、新しい EBPM である EBPM2.0 は、新・三

⁴² 龍慶昭・佐々木亮（2004）

⁴³ セオリー評価におけるセオリーの定義は、龍・佐々木（2004, p26）では「原因と結果が連鎖上に連なる「仮定」としてしています。ロッシ・レプセイ・フリーマン（2005, P89）ではプログラム理論について、「そのプログラムが採用してきた戦略や戦術と、期待される社会的便益との関係に関する一連の仮説のことである」としてしています。このようなセオリー（理論）の定義は、セオリー評価が属するプログラム評価におけるもので、多くの学術分野における「理論」とは意味することが異なっている可能性に留意する必要があります。

本の矢から構成されており、それらは①因果推論に基づく政策効果の厳密な検証と検証結果の政策への反映、②セオリー評価に基づく政策の確からしさの向上、③業務情報へのアクセスの拡大からなります。

引用文献

- Allcott, H., Boxell, L., Conway, J., Gentzkow, M., Thaler, M., & Yang, D. (2020). "Polarization and public health: Partisan differences in social distancing during the coronavirus pandemic," *Journal of Public Economics*, 191, 104254.
- Angrist, J. D., & Pischke, J.-S. (2010). "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics," *Journal of Economic Perspectives*, 24(2), 3-30.
- Athey, S., Castillo, J. C., & Chandar, B. (2025). "Service quality on online platforms: Empirical evidence about driving quality at uber," *Management Science*.
- Baron, J. (2018). "A Brief History of Evidence-Based Policy," *The ANNALS of the American Academy of Political and Social Science*, 678(1), 40-50.
- Bernecker, A., Boyer, P. C., & Gathmann, C. (2021). "The Role of Electoral Incentives for Policy Innovation: Evidence from the US Welfare Reform," *American Economic Journal: Economic Policy*, 13(2), 26–57.
- Brodersen, K. H., Gallusser, F., Koehler, J., Remy, N., & Scott, S. (2015). "Inferring causal impact using Bayesian structural time-series models," *Ann Appl Stat*, 9, 247-274.
- Cowan, S. K., Mark, N., & Reich, J. A. (2021). "COVID-19 Vaccine Hesitancy Is the New Terrain for Political Division among Americans," *Socius*, 7, 23780231211023657.
- Coyle, D. J., & Wildavsky, A. (1986). "Social experimentation in the face of formidable fables," *Conference Series ; [Proceedings]*, 30, 167-193.
- Elango, S., García, J. L., Heckman, J. J., & Hojman, A. (2016). 4. Early Childhood Education. In A. M. Robert (Ed.), *Economics of Means-Tested Transfer Programs in the United States, Volume II*(pp. 235-298). Chicago: University of Chicago Press.
- Evidence-Based Medicine Working Group. (1992). "Evidence-based medicine. A new approach to teaching the practice of medicine," *JAMA*, 268(17), 2420.

- Fuleihan, C. (2023). "Shattering the Mirage: The FDA's Early COVID-19 Pandemic Response Demonstrates a Need for Reform to Restore Agency Credibility," *American Journal of Law & Medicine*, 48(4), 307-342.
- Gueron, J. M. (2017). Chapter 2 - The Politics and Practice of Social Experiments: Seeds of a Revolution. In A. V. Banerjee & E. Duflo (Eds.), *Handbook of Economic Field Experiments* (Vol. 1, pp. 27-69): North-Holland.
- Haskins, R., & Margolis, G. (2014). *Show me the evidence: Obama's fight for rigor and results in social policy*: Brookings Institution Press.
- Howard, M. C. (2022). "Are face masks a partisan issue during the COVID-19 pandemic? Differentiating political ideology and political party affiliation," *International Journal of Psychology*, 57(1), 153-160.
- Juras, R., Kelsey, M., Lipsey, M., Steinka-Fry, K., & Layzer, J. (2019). *Meta-Analysis of Federally-Funded Teen Pregnancy Prevention Programs: Final Report*: U.S. Department of Health and Human Services.
- Lenzer, J. (2020). "Covid-19: US gives emergency approval to hydroxychloroquine despite lack of evidence," *BMJ*, 369, m1335.
- Meldrum, M. L. (2000). "A Brief History of the Randomized Controlled Trial: From Oranges and Lemons to the Gold Standard," *Hematology/Oncology Clinics of North America*, 14(4), 745-760.
- Moffitt, R. A. (2004). "The Role of Randomized Field Trials in Social Science Research: A Perspective from Evaluations of Reforms of Social Welfare Programs," *American Behavioral Scientist*, 47(5), 506-540.
- Mosteller, F., & Boruch, R. F. (2002). *Evidence matters: Randomized trials in education research*: Brookings Institution Press.
- Pearl, J., & Mackenzie, D. (2018). *The Book of Why: The New Science of Cause and Effect*: Hachette UK.
- Sanderson, I. (2002). "Evaluation, Policy Learning and Evidence-Based Policy Making," *Public Administration*, 80(1), 1-22.
- U.S. Department of Education. (2023). *Using Evidence to Strengthen Education Investments*. <https://www.education.ne.gov/wp-content/uploads/2025/01/titleii-2023-non-regulatory-guidance-evidence.pdf>
- U.S. Department of Health and Human Services. (2010). *Head Start Impact Study: Final report*.

- https://acf.gov/sites/default/files/documents/opre/hs_impact_study_final.pdf
- Ventry, Dennis J. (2000). "The Collision of Tax and Welfare Politics: The Political History of the Earned Income Tax Credit, 1969–99," *National Tax Journal*, 53(4.2), 983-1026.
- Weiss, C. H. (1988). "Reports on Topic Areas: Evaluation for Decisions: Is Anybody There? Does Anybody Care?," *Evaluation Practice*, 9(1), 5-19.
- Widerquist, K. (2019). The Negative Income Tax Experiments of the 1970s. In M. Torry (Ed.), *The Palgrave International Handbook of Basic Income* (pp. 303-318). Cham: Springer International Publishing.
- リスト, ジョン／高遠裕子訳. (2023). 『そのビジネス、経済学でスケールできません』: 東洋経済新報社.
- ロッシ・リプセイ・フリーマン／大島巖・平岡公一・森俊夫・元永拓郎監訳. (2005). 『プログラム評価の理論と方法』: 日本評論社.
- ワイス, キャロル・H／前川美湖・池田満訳. (2014). 『入門評価学: 政策・プログラム研究の方法』: 日本評論社.
- 益田直子. (2019). "評価活動は、何を、どのように、私たちにもたらしうるのか? —「評価の利用・影響」に関する理論研究及び実証研究を振り返る," *日本評価研究*, 19(2), 19-34.
- 近藤清太郎. (2022). 「「真価志向」の政策形成: 因果推論と効果検証の考え方」、大竹文雄・内山融・小林庸平編著『*EBPM: エビデンスに基づく政策形成の導入と実践*』: 日本経済新聞出版.
- 佐々木亮. (2020). "国際開発分野における EBPM の現状," *日本評価研究*, 20(2), 89-103.
- 小林庸平. (2024). "アメリカ連邦政府におけるエビデンスに基づく政策形成の実際," *RIETI Policy Discussion Paper*, 24-P-009.
- 小林庸平. (2025). "日本の EBPM (エビデンスに基づく政策形成) はなぜ機能しないのか?: 日米比較と Evidence-Support System から考察," *日本評価研究*, 25(1), 21-35.
- 杉谷和哉. (2022). 政策にエビデンスは必要なのか: EBPM と政治のあいだ: ミネルヴァ書房.
- 杉谷和哉. (2024). 日本の政策はなぜ機能しないのか?: EBPM の導入と課題: 光文社.
- 大竹文雄・内山融・小林庸平編著. (2022). *EBPM: エビデンスに基づく政策形成の導入と実践*. 日本経済新聞出版.
- 津田広和・岡崎康平. (2018). "米国における Evidence-based Policymaking

- (EBPM) の動向," *RIETI* ポリシー・ディスカッション・ペーパー, 18-P-016.
- 津田広和・岡崎康平.(2022). 米国における *EBPM*. 大竹文雄・内山融・小林庸平 編著 *EBPM : エビデンスに基づく政策形成の導入と実践*. 日本経済新聞出版.
- 田辺智子.(2006). "エビデンスに基づく教育 : アメリカの教育改革と What Works Clearinghouse の動向," *日本評価研究*, 6(1), 31-41.
- 田辺智子.(2020). "エビデンスに基づく政策立案 (EBPM) の推進に向けて : 医療の経験からの示唆," *日本評価研究*, 20(2), 19-31.
- 龍慶昭・佐々木亮.(2004). <増補改訂版> 「政策評価」 の理論と技法. 多賀出版.