

保健・医療・介護（予防）領域における

社会的インパクト 評価ガイドライン 2026

—エビデンスに基づく政策の実現に向けて—

監修・編集

社会的インパクト評価ガイドライン検討会



京都大学大学院医学研究科
社会的インパクト評価学講座



一般財団法人 医療経済研究・社会保険福祉協会
医療経済研究機構

はじめに

本ガイドラインは、国(内閣府、厚生労働省、経済産業省など)によって推進されている保健・医療・介護(予防)などの領域における「根拠に基づく政策形成(Evidence Based Policy Making:EBPM)」に必要な社会的インパクト評価の促進を目的に、その考え方、評価手順、方法・評価研究デザインなどについてまとめた文書です。政策・施策・事業のレベルによってそれらは異なるので、コンパクトに収めるために、対象者を明確に定義できる施策・事業における評価を中心に述べています。

社会問題は複雑化し、産官学民のセクターが協働する必要が高まり、評価においても満足度等の主観的指標だけでなく客観的指標による成果の可視化や説明責任が問われ、施策や事業の「活動の量(アウトプット)」でなく「成果(アウトカム)」の重視への社会的要請が高まっています。これらに対応するためにはロジックモデルを作成・共有し、多様な立場の関係者で議論を重ね、EBPMを促進することが期待されています。それらを進める上で阻害要因の一つになっているのは、評価が参加者への事後アンケートなどに留まり、前後の変化を捉えたアウトカム評価や背景要因を揃えた上で参加(介入)群と非参加(対照)群とで比較する社会的インパクト評価がなされていないことです。さらに評価結果を踏まえて政策等を見直すマネジメントサイクルが回っていないのが実情です。一方で、研究者が評価に関わると、学術的には、現場での実行が困難な実装前研究である無作為化比較試験などが期待され「利用可能な最善のエビデンスを賢明に用いた関係者間の熟議を通じて政策効果を高めるEBPM」とは乖離した状況です。そこで、本ガイドラインは、行政職員と評価に関わる者を主な読者として想定し、内閣府が推進する成果連動型民間委託契約方式(Pay For Success :PFS)における社会的インパクト評価なども含め、行政現場で実施可能で、かつ学術的にもできるだけ妥当性が高いバランスを考えた評価の促進を願ってまとめました。

本検討会は、京都大学大学院医学研究科 社会的インパクト評価学講座(寄附講座)と一般財団法人 医療経済研究・社会保険福祉協会 医療経済研究機構が事務局を担い、国立研究開発法人 日本医療研究開発機構(Japan Agency for Medical Research and Development: AMED)研究班に参加した保健医療福祉分野のプログラム・政策評価研究を専門とする研究者を中心に構成し、合意できる内容(コンセンサス)をまとめました。

Part I では、実施体制づくりと政策目的の設定から社会的インパクト評価までの「社会的インパクト評価の手順」について、Part II では、社会的インパクト評価に必要となる3条件と様々な評価デザインと分析手法、それらからの選択の視点など「社会的インパクト評価の研究デザイン」について述べています。社会的インパクト評価を実施するには、相応の学識経験が必要であることから、本ガイドラインは、社会的インパクト評価について学識経験を持つ第三者に外部委託などをする時に必要となる考え方や手順について理解を得ることを目指してまとめました。

本ガイドラインの編集にあたっては、多くの行政現場で使っていただくことを願い、本文はできるだけコンパクトにまとめ、専門用語などはコラムで説明するようにしました。自治体や中央省庁の担当者・経験者に原稿を読んでいただき、行政の現場で実施可能で、非専門職でも理解できるものになっているか、ご意見を伺いました。さらに検討会に直接参加しなかった10人の外部レビュアーにも原稿を読んで意見をいただき、それらを反映することにより現場で使えて学術的にも妥当性が高い文書となることを目指しました。

本ガイドラインが、政策・施策・事業のアウトカム評価や社会的インパクト評価、EBPM、PFSの普及・推進に役立つことを願っています。

2026 年 6月

京都大学大学院医学研究科 社会的インパクト評価学講座 特定准教授 高木 大資
一般財団法人 医療経済研究・社会保険福祉協会 医療経済研究機構 研究部長 近藤 克則

* * * * * 目 次 * * * * *

概要 本ガイドラインについて	4
Part I 社会的インパクト評価の手順	13
1章 実施体制づくりと実施目的の設定	13
1) 実施体制づくり	13
2) 実施目的の設定	13
2章 ロジックモデルの作成	18
1) ロジックモデルの構成要素	18
2) 施策・事業の目的、ロジックを明確にするための4ステップ	19
3) ロジックモデル作成の手順	21
3章 社会的インパクト予測評価	33
1) インパクト予測評価	33
2) 社会的インパクト予測評価の視点と手順	33
4章 評価計画の立案	35
1) 評価の方法とエビデンス	35
2) 評価計画	35
3) 中間評価の計画と実施	36
4) 評価の計画と実施	39
5章 施策・事業の実施とデータ収集	46
1) 施策・事業の実施とプロセス・アウトカムの測定	46
6章 社会的インパクト評価	48
1) 効果の検証	48
2) 社会的インパクト評価の結果に基づく判断	48
3) データの公表	48
Part II 社会的インパクト評価の研究デザイン	50
1章 「見かけ上でない施策・事業効果」の評価に必要となる3条件	50
1) 比較対照群の設定	51
2) 前後比較で変化を捉える	52
3) 背景要因を揃えて比較	53
2章 評価事例と評価に要する期間	57
3章 様々な評価デザインと分析手法、選択の視点	60
1) 様々な評価デザインと分析手法におけるトレードオフ関係	60
2) 選択の視点	60
参考資料	63
様々な評価デザインと分析手法	63
1) 実験研究デザインと準実験研究デザイン	63
2) 様々な分析手法の概要	64
3) エビデンスレベルと評価デザインの選び方	73

本編に掲載したコラムの一覧を以下に示します。

コラム No.	タイトル	頁
コラム 1	社会的インパクト評価の定義	9
コラム 2	政策と政策評価環境の発展段階（成熟度）	10
コラム 3	政策・施策・事業	10
コラム 4	4 種類の課題	14
コラム 5	4 つのインパクト	16
コラム 6	アウトカムとインパクト	17
コラム 7	ロジックモデルと Theory of Change	18
コラム 8	ロジックモデル作成における 2 つの型	20
コラム 9	エビデンスとは何か	22
コラム 10	データ等のエビデンス（広義）の収集作業	23
コラム 11	PICO の枠組み	23
コラム 12	既存のエビデンスの収集方法	24
コラム 13	ロジックモデルとエビデンスの関係性	25
コラム 14	PFS における測定指標	28
コラム 15	評価の 5 原則	28
コラム 16	ロジックモデルの形式	29
コラム 17	アセスメント (assessment) とエヴァリュエーション (evaluation)	33
コラム 18	ヘルスプロモーション	34
コラム 19	フィデリティ評価	37
コラム 20	施策・事業評価の目的	37
コラム 21	施策・事業評価のアプローチ	38
コラム 22	事前・中間・事後評価と評価の 5 階層	38
コラム 23	効率性評価	39
コラム 24	公平・公正性評価	40
コラム 25	アウトカム指標の設定	40
コラム 26	アウトカム指標が満たすべき条件	40
コラム 27	費用効果	41
コラム 28	費用便益	41
コラム 29	事例 愛知県武豊町／よこはまウォーキングポイント事業を用いた効率性評価	42
コラム 30	無作為化比較試験（Randomized Controlled Trial: RCT）	43
コラム 31	異質性の考慮	44
コラム 32	プロセス評価とモニタリング	47
コラム 33	参加群・非参加群の別名と対照群と対象群の違い	52
コラム 34	「関連」と「因果」	53
コラム 35	交絡	54
コラム 36	効果分析に必要なデータの内容	55
コラム 37	対照群を設定できない場合の評価	56
コラム 38	事例：A 市での PFS による介護予防事業の社会的インパクト評価	57
コラム 39	操作変数法と差の差分分析、傾向スコアマッチングの前提条件	62
コラム 40	評価・分析単位	64
コラム 41	クロスオーバー法、ウェイトイングリストコントロール法	64
コラム 42	クラスター無作為化比較試験（CRCT）	65
コラム 43	Target Trial Emulation や機械学習などの新たな分析手法	73
コラム 44	エビデンスのピラミッドの見直しについて	74

概要

保健・医療・介護（予防）領域における 社会的インパクト評価ガイドライン —エビデンスに基づく政策の実現に向けて—

本ガイドラインについて

本ガイドラインは、行政現場での政策・施策・事業の効果を高める「根拠に基づく政策形成」（Evidence Based Policy Making：EBPM）を進めるために、求められる社会的インパクト評価の手順と研究デザインの考え方をまとめたものである。

背景と目的

EBPM を促進することが期待されている。しかし現状では、多くの評価が事後アンケートに留まり、前後比較や対照群との比較評価が十分ではない。一方、学術的には行政現場で実施困難な無作為化比較試験（Randomized Controlled Trial: RCT）など厳密な因果関係の実証が求められている。保健・医療・介護（予防）等の分野で、「根拠に基づく政策形成」（EBPM）を進めるため、行政現場でも実施可能で、学術的にも妥当性がより高い、政策・施策・事業の効果を高める「社会的インパクト評価」の基本的な考え方や手順、評価研究デザインの選択の視点について簡潔なガイドラインを作成することを目的とした。

対象と適用範囲

中央省庁・地方自治体の行政職員、民間事業者・NPO 等のサービス提供者、第三者評価を担う大学・研究機関の研究者など社会的インパクト評価に関わる者を主な読者として想定した。

適用範囲は、保健・医療・介護（予防）分野における成果連動型民間委託契約方式（Pay For Success：PFS）など、社会的インパクトを期待する政策を念頭にまとめたが、他分野にも適用可能な内容を含むと考える。

社会的インパクトと社会的インパクト評価の定義

定義（コラム 1 社会的インパクト評価の定義）には、その登場の背景の違いから、学術的に厳密なものから広義の一般的な用法までいろいろあるが、本文書では内閣府の定義を用いる。社会的インパクトとは、「短期、長期の変化を含め、当該事業や活動の結果として生じた社会的、環境的な『アウトカム（効果）』」⁽¹⁾。社会的インパクト評価とは、「社会的インパクトを定量的・定性的に把握し、当該事業や活動について価値判断を加えること」⁽¹⁾である。

意義

行政現場で実施可能な社会的インパクト評価の手順や評価研究デザインについて、現場への導入可能性と学術的妥当性を高めるためのコンセンサスを文章化し、EBPM のマネジメントサイクルを回すことで政策・施策・事業の社会的インパクトの最大化に寄与する。

作成体制と作成の流れ

京都大学大学院医学研究科 社会的インパクト評価学講座と医療経済研究機構 研究部とが事務局を担い、保健医療福祉分野の研究者 7 人からなる検討会で原案を作成し、外部の自治体職員をはじめとする 10 人の外部レビューアーの意見を踏まえて改訂して作成した。

書誌情報

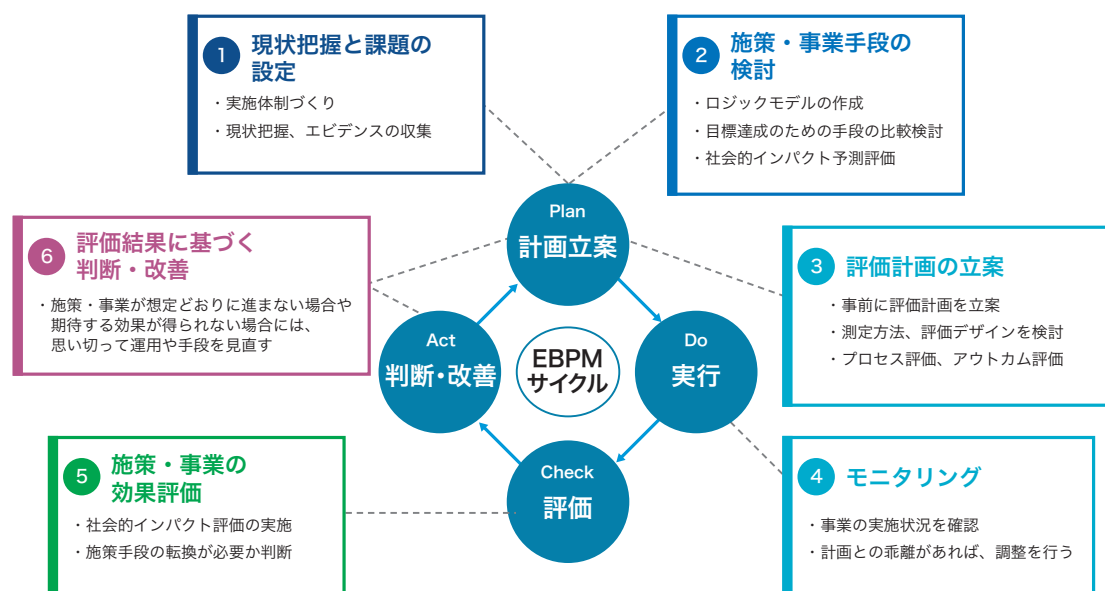
社会的インパクト評価ガイドライン検討会「保健・医療・介護（予防）領域における社会的インパクト評価ガイドライン 2026 ―エビデンスに基づく政策の実現に向けて―」京都大学大学院医学研究科社会的インパクト評価学講座・医療経済研究機構、2026

Part I：社会的インパクト評価の手順

1 章 実施体制づくりと実施目的の設定

関係者を集めた実施体制づくりとそこで施策・事業の目的について議論し合意することから始まる。

- ・実施体制づくり：施策・事業の実施主体、対象集団、事業担当者、学識経験者などの関係者が協議する場をつくる
- ・実施目的の設定：以下 3 要素について関係者で議論し合意する
①現状把握、②課題設定（関連要因の特定）、③アウトカムなどの大枠



（内閣官房行政改革推進本部事務局 EBPM ガイドブック Ver.1.2より引用、改変）

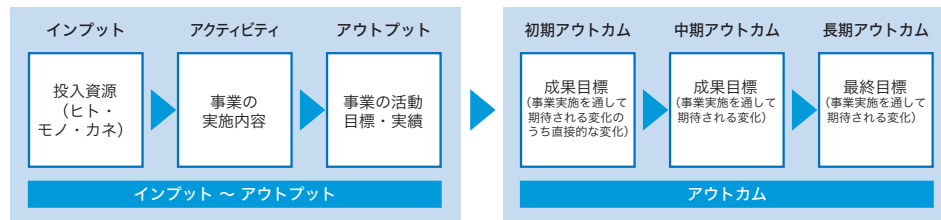
概要 図1 EBPMのマネジメントサイクル

2 章 ロジックモデルの作成

関係者を集めた実施体制づくりとそこで施策・事業の目的について議論し合意することから始まる。

- ・ロジックモデルは、施策・事業課題について「論理（ロジック）」を端的に示した「模式図（モデル）」であり、施策・事業の基本的な枠組みを見える化するツールである。

- ・ロジックモデル作成：インプット→アクティビティ→アウトプット→アウトカムとして図や表で整理し、あわせて測定指標（KPI 等）を選定する。



概要 図2 ロジックモデル

3章 社会的インパクト予測評価

- ・施策・事業導入前に、その影響を事前に予測・評価する。望ましい（正の）影響と（負の）悪影響の両面を検討する。特に負の悪影響を受ける集団への配慮が重要で、悪影響の最小化と望ましい影響の最大化を図る。

4章 評価計画の立案

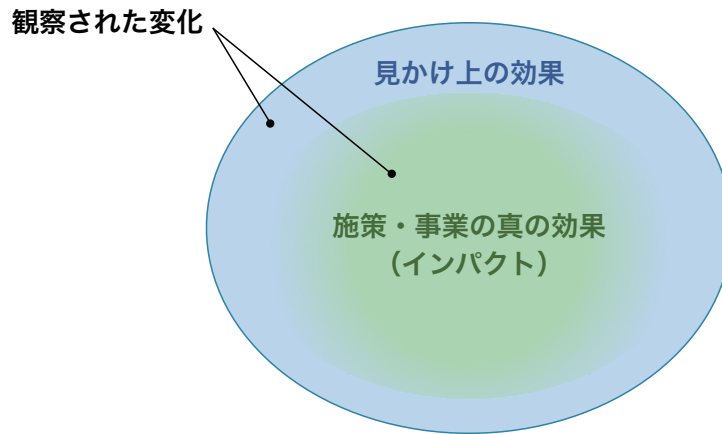
- ・評価計画は、施策・事業の継続・改善・見直しや、拡充の可否を判断するために、実施前に立案する。
- ・評価計画では、目的、対象、時期、評価研究デザイン、必要なデータ、分析方法をあらかじめ整理する。
- ・中間評価では、主にプロセス評価を行い、必要に応じて初期・中期アウトカムも把握し、改善につなげる。
- ・評価結果は、統計的有意性だけでなく実践的意義も踏まえて解釈し、施策・事業の改善や説明責任に活用する。

5章 施策・事業の実施とデータ収集

- ・施策・事業を実施し、参加者数やサービス提供量などのプロセス指標、健康・活動・生活の質などのアウトカム指標など、評価に必要なデータを収集して、運営改善に活かす。

6章 社会的インパクト評価

- ・効果の検証：施策・事業の前後で観察された変化には、施策・事業により生じたものと、他の要因による見かけ上のものを含んでいる。できる限り見かけ上のものを排除した効果の検証を行う。
- ・データの公表：事前の計画に基づき報告書を作成する。①次年度の計画立案に必要な速報値、②年度末の報告書（確定値）、③市民等に向けたプレスリリース資料、などがある。
- ・評価の結果に基づく判断：効果検証の結果に基づき施策・事業の価値判断を行い、継続すべきか、修正すべきか、統合・廃止すべきかなどを決定する。成果連動型民間委託契約方式（PFS）においては、社会的インパクト評価の結果が報酬の支払や契約条件の修正に活用される。

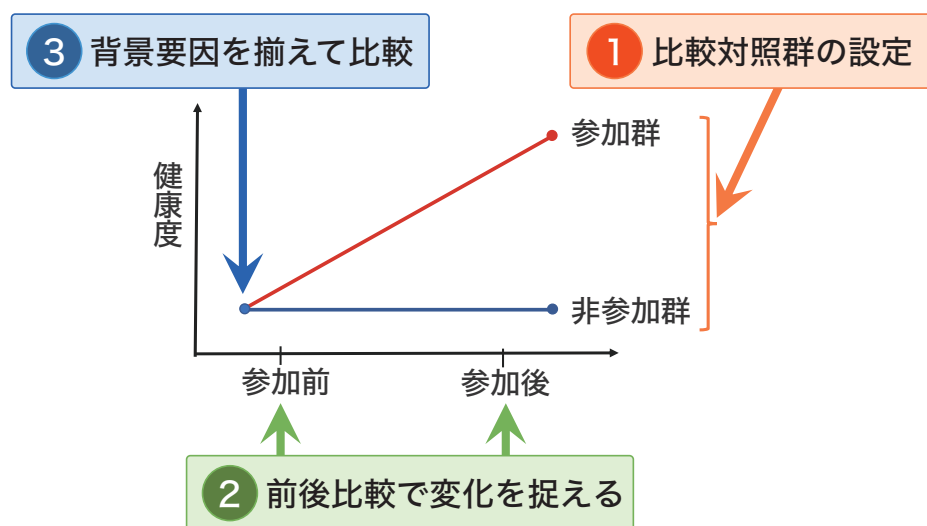


概要 図3 施策・事業の真の効果(インパクト)と見かけ上の効果

Part II：社会的インパクト評価の研究デザイン

1 章 「見かけ上でない施策・事業効果」の評価に必要な 3 条件

- ・観察された変化のすべてが施策・事業の効果とは限らない。見かけ上の効果を除いた効果を評価するには以下 3 条件が必要である。
 1. 比較対照群の設定：参加群と非参加（対照）群を比較し、施策・事業以外の要因の影響をできるだけ差し引く。
 2. 参加前後の変化の把握：1 時点の比較だけでは逆の因果による見かけ上の差を区別できないため、同じ対象の参加前後の変化を確認する必要がある。
 3. 背景要因を考慮した比較：年齢、所得、健康状態などの違いを踏まえ、背景要因の影響をできるだけ調整・考慮して比較する。



概要 図4 見かけ上でない効果評価に必要な3条件

2 章 評価事例と評価に要する期間

- ・実施・評価の準備にかかる期間の一例は、事前評価に数か月から 2 年、中間評価には施策・事業導入の数か月～ 5 年間の事業の場合には 2・3 年、事後評価には施策・事業期間終了後半年（速報値）～ 1 年程度かかることが多い。

3章 様々な評価研究デザインと分析手法、選択の視点

- ・上記3条件を満たすために多くの評価研究デザインが開発されてきた。現場での実施が容易な方法は「見かけ上の効果」が紛れ込みやすくなり、一方、「見かけ上の効果」を排除できる学術的に厳密な評価研究デザインは現場での実施が容易でない、トレードオフ（一方を立てると、もう一方が犠牲になる）の関係がある。そのため、専門家と相談しながら実施可能かつできるだけ妥当性の高い方法を選択する。
- ・エビデンスレベルと評価研究デザインの選び方
 - ▶施策・事業の発展段階によって、実施が望ましい評価研究デザインは異なる。定量的な評価が困難な初期段階では定性的な情報により施策・事業の改善を図ることに重点を置く。関連要因を含む定量データが得られる段階になれば、関連要因の影響を除いた定量的な評価に重点を置く。
 - ▶従来エビデンスレベルが高いとされてきた無作為化比較試験（RCT）は行政現場での実施が容易でない。実施可能性と科学的妥当性のトレードオフを踏まえ、傾向スコアマッチングが至適となることは多い。
 - ▶施策・事業の発展段階や実施体制を踏まえ、適した評価研究デザインを選択する必要がある。必要に応じ専門家と相談し、実施可能で、できるだけ妥当性の高い方法を選択する。

【概要の参考文献】

- (1) 内閣府．社会的インパクト評価の推進に向けて（概要）—社会的課題解決に向けた社会的インパクト評価の基本的概念と今後の対応策について— [Internet]. 2016 Mar [cited 2026 Apr 7]. Available from: https://www8.cao.go.jp/kyumin_yokin/shiryou/sonota/shakaitekijigyou/shakaitekijigyou_04.pdf

本ガイドラインについて

本ガイドラインは、行政現場での政策・施策・事業の効果を高める「根拠に基づく政策形成」(Evidence Based Policy Making：EBPM)を進めるために、求められる社会的インパクト評価の手順と研究デザインの考え方をまとめたものである。

背景と目的

EBPMを促進することが期待されている。しかし現場では、多くの評価が事例報告や事後アンケートに留まり、前後比較や対照群との比較など定量評価が十分ではない。一方、研究者が評価に関わると、現場で実行が困難な実装前研究である無作為化比較試験などが期待され、「利用可能な最善のエビデンスを賢明に用い関係者間の熟議を通じて政策効果を高める「根拠に基づく政策形成」(EBPM)」とは乖離している状況がある。保健・医療・介護（予防）等の分野で、本来のEBPMを進めるため、行政現場でも実施可能で、学術的にも妥当性がより高い、行政現場における政策・施策・事業の効果を高める「社会的インパクト評価」の基本的な考え方や手順、評価研究デザインの選択の視点について簡潔なガイドラインを作成することを目的とした。

コラム 1

社会的インパクト評価の定義

社会的インパクト評価の定義には、登場の背景や立場、重視するものの違いを反映し、いくつかの定義・捉え方がある。一方の極には、インパクト投資などのための評価という文脈で、不確実性を前提とし「完璧な因果推論」ではなく「意思決定に資するエビデンス」を得るという立場がある。行政現場の立場では、政策等の効果を最大化するための一つ的手段としてインパクト評価を位置づけ、学術的な厳密さより、政策・行政現場の実情を反映した最低限のものであっても評価することが重視される。もう一方の極には、見かけ上のものを効果と見誤っては、改善につながらないとして、コストをかけてでも、できるだけ厳密な方法によって「真の効果」の評価をめざすべきとする科学的・学術的な妥当性を重視する立場もある。これらの背景や目的、立場の違いに加え、対象となる政策や評価環境の発展段階（成熟度）の違い（コラム2 政策と政策評価環境の発展段階（成熟度））によっても、実施可能な社会的インパクト評価のあり方は影響を受ける。

本ガイドラインでは、政策等の効果を高めることを重視する行政現場の立場から内閣府の社会的インパクト評価の定義「社会的インパクトを定量的・定性的に把握し、当該事業や活動について価値判断を加えること」の立場でまとめた。ただし、成果連動型民間委託契約方式（Pay For Success：PFS）など、前後変化から見かけ上の効果を差し引いた真の効果の評価が求められる場面が増えつつあることを考慮し、そのために必要となる3条件など、目指すべき評価について説明する。

コラム 2

政策と政策評価環境の発展段階（成熟度）

政策にも、初期段階から成熟段階まである⁽¹⁾。初期段階では、新たに浮上した課題に対応すべきことは合意されたが、政策立案の際に参考にできる先行事例も乏しく、どのような対策に、どれほどの効果を期待できるのかなどの情報も限られている。初期段階では、変化や効果の大きさを評価できるデータベースもなく尺度も確立していないこともあって厳密で定量的なアウトカム評価や因果推論は難しく、政策を改善するための定性的なプロセス評価が評価の中心となる。やがて試行錯誤を経て、知見やデータ、評価手法が蓄積された成熟段階になると、プロセス評価に留まらず、対照群と比較して観察された変化（アウトカム）から、見かけ上の効果を差し引いた定量的な事後評価が可能となる。

日本においては、政策・施策・事業評価が広く普及しているとはいいがたく、行政現場でも第三者評価を担う学術・評価機関でも、データ収集やデータベース整備、評価分析の経験や先行事例などの蓄積途上にある。したがって、学術的に見て理想的な評価が実施可能であることはまれである。できるだけ多くの政策・施策・事業のプロセスやアウトカムなどに対して、できる範囲の評価を組み入れ、得られた科学的根拠（エビデンス）を最大限に活かして効果を高める「根拠に基づく政策形成」（Evidence Based Policy Making：EBPM）を育てる段階にあると思われる。一方、内閣府成果連動型事業推進室が普及を図っている PFS など、従来のプロセスのみの評価からアウトカム重視の評価が前提となる場合には、見かけ上の効果まで報酬の対象としない、施策・事業そのものによる成果の評価に迫ることが必要となる。

コラム 3

政策・施策・事業

一般に政策と呼ばれるものにも、予算や担当課・室が最小単位である「事業」と、いくつかの事業が組み合わさり部局単位などで取り組む「施策」、さらに複数の部局の施策群によって進められる「政策」とを呼び分けることがある。これらのレベルによって、評価対象となる理念や目標、重視される基準、評価指標の抽象度、説明責任を担う担当者（首長・部長・室長など）が異なってくるからである。本ガイドラインで説明する評価は、追跡調査や参加者と非参加者の比較を想定しているので、参加者が特定できる施策・事業レベルを主に想定している。3つのレベルを指すときには政策・施策・事業または政策等と表記することにした。

対象と適用範囲

中央省庁・地方自治体の行政職員、民間事業者・NPO 等のサービス提供者、第三者評価を担う大学・研究機関の研究者などを主な読者として想定した。

主な適用範囲として、保健・医療・介護（予防）分野における PFS などのアウトカムや社会的インパクト評価を念頭にまとめたが、他分野にも適用可能な内容を含むと考える。

社会的インパクトと社会的インパクト評価の定義

インパクトという言葉には、いろいろな用語法がある（本編 Part I コラム 5 4 つのインパクト、P.16）。本ガイドラインでは、行政現場での利用を想定し、内閣府の社会的インパクトの定義「短期、長期の変化を含め、当該事業や活動の結果として生じた社会的、環境的な『アウトカム（効果）』」⁽²⁾を用いる。また社会的インパクト評価とは「社会的インパクトを定量的・定性的に把握し、当該事業や活動について価値判断を加えること」である。

意義

行政現場で実施可能な社会的インパクト評価の手順や評価研究デザインについて、現場への導入可能性と学術的妥当性もできるだけ担保したコンセンサスを文章化し、EBPM とマネジメントサイクルを回すことで政策の社会的インパクトの最大化に寄与する。

作成体制と作成の流れ

本ガイドライン作成主体である京都大学 大学院医学研究科 社会的インパクト評価学講座および医療経済研究機構が事務局を担い、予防・健康づくりに関するエビデンス構築の研究を担う学術系研究者、学術と行政の「両者の立場」を理解する有識者 7 人ならびに内閣府、厚生労働省、経済産業省等の関係行政機関の担当者（オブザーバー）から構成される「ガイドライン検討会」を設置し、原案を作成した。その後、外部の自治体職員をはじめとする 10 人の外部レビューアーの意見を踏まえて加筆修正して作成した。

●ガイドライン作成主体

京都大学大学院医学研究科 社会的インパクト評価学講座
一般財団法人 医療経済研究・社会保険福祉協会 医療経済研究機構

●ガイドライン検討会メンバー

福島県立医科大学 医学部 疫学講座 主任教授 大平哲也
北海道大学大学院 医学研究院 医療政策評価学教室 教授 古元重和
京都大学 大学院医学研究科 社会疫学分野 主任教授 近藤尚己
京都大学 大学院医学研究科 社会的インパクト評価学講座 特定准教授 高木大資
京都大学 大学院医学研究科 社会的インパクト評価学講座 特定助教 石村奈々
医療経済研究機構 研究部長／千葉大学 予防医学センター 特任教授 近藤克則
医療経済研究機構 研究員 藤原聡子

●外部レビューアー 2026 年 3 月時点 ※は現在

青柳光昌 公益財団法人 社会変革推進財団（SIIF）専務理事
伊藤健 同志社大学 政策学部 総合政策科学研究科 客員教授／特定非営利活動法人ソーシャルバリュージャパン※
Social Value Japan（旧称：特定非営利活動法人 SROI ネットワークジャパン）代表理事
大島巖 東北福祉大学客員教授／日本社会事業大学名誉教授／一般財団法人 PBEE 研究・研修センター代表理事※

幸地仁詩 堺市 健康福祉局 長寿社会部 長寿支援課 参事（企画・調整担当）
千葉直紀 株式会社ブルー・マーブル・ジャパン 代表取締役
塚本一郎 明治大学 経営学部 教授
南條正樹 静岡市役所 保健福祉長寿局 地域包括ケア・誰もが活躍推進本部
野口晴子 早稲田大学 政治経済学術院 教授
三善仁 阪急阪神ホールディングス株式会社 グループ開発室 事業開発部（ヘルスケア担当）課長
吉田泰治 株式会社ドリームインキュベータ 執行役員

●オブザーバー

内閣府 成果連動型事業推進室
厚生労働省 政策統括官（総合政策担当）付 政策統括室 政策第5班
経済産業省 商務・サービスグループ ヘルスケア産業課
国立研究開発法人 日本医療研究開発機構（AMED）医療機器・ヘルスケア事業部 ヘルスケア研究開発課

●書誌情報

社会的インパクト評価ガイドライン検討会「保健・医療・介護（予防）領域における社会的インパクト評価ガイドライン 2026 ―エビデンスに基づく政策の実現に向けて―」京都大学大学院医学研究科社会的インパクト評価学講座・医療経済研究機構、2026

【参考文献】

- (1) 一般財団法人 日本民間公益活動連携機構．資金分配団体・実行団体に向けての評価指針 2020 年 7 月改訂版．2020.
- (2) 内閣府．社会的インパクト評価の推進に向けて（概要）―社会的課題解決に向けた社会的インパクト評価の基本的概念と今後の対応策について― [Internet]. 2016 Mar [cited 2026 Apr 7]. Available from: https://www8.cao.go.jp/kyumin_yokin/shiryou/sonota/shakaitekijigyou/shakaitekijigyou_04.pdf

評価は、1. 実施体制づくりと実施目的の設定、2. ロジックモデル作成、3. 社会的インパクト予測評価、4. 評価計画の立案、5. 施策・事業の実施とデータ収集、6. 社会的インパクト評価の順に進めていく。

1 章 実施体制づくりと実施目的の設定

- ・ 施策・事業の実施主体は、行政だけでなく、投資や助成を受ける NPO や民間事業者などが含まれる。
- ・ 社会課題の解決のためには、まず関係者を集めた実施体制づくりが必要である。限られた人員・予算・時間の中で可能な範囲で体制づくりを行う。
- ・ 関係者で、実施目的の明確化に必要な 3 つの要素（①現状把握、②課題設定、③アウトカム）の大枠について議論し合意するところから始める（Part I 2 章参照）。
- ・ 実施主体は、限られた時間内での合意形成に向けて、実施目的やロジックモデル、評価計画などの骨子や原案を準備する。
- ・ 1 章では、事前評価の中でも、ロジックモデル作成に入る前に関係者で議論し合意すべきニーズ評価に関わる点について記述する。

1) 実施体制づくり

- ・ 関係者には、実施主体以外に、事業担当者（候補）、対象となる集団、学識経験者、評価を行う者（候補）などがある。
- ・ 一般に、立場によって、重視する価値や側面、利害などが異なる。
- ・ 政策導入や実施中、事後評価の結果が出てからの立場の違いによる紛糾を避けるためには、施策・事業の形成段階から関係者を特定し、事業者へのサウンディング（事業の発案段階や事業者公募の前に、事業の実現可能性や条件、課題などを探るために民間企業と行政などが直接対話を行うこと）等を通し、社会的インパクト評価の目的と方法（評価対象、評価する側面や指標など）について話し合い、確認と合意を得る場や体制が望まれる。
- ・ 評価は、簡易なものであれば実施主体が自ら行うこともある。しかし、より妥当な評価を行うには、相当の学識経験を持つ者による第三者評価が必要となる。

2) 実施目的の設定

- ・ 実施目的の明確化には、①現状把握、②課題設定（**コラム 4 4 種類の課題**）、③住民・社会の目指す姿（アウトカム）の大枠について、関係者で議論し合意する必要がある。

コラム 4

4 種類の課題

日本語の「課題」には、英語では異なる表現が使われているいくつかの意義があり、同じ用語でも、人によって捉え方・意味するものが異なるので注意が必要である。

a) 観察される「望ましくないこと」「問題」(problem)

例) 閉じこもり高齢者問題、健診未受診者問題などで、ロジックモデルの長期アウトカムに位置づけられるものにあたる。

b) 「問題」を生み出している「原因」あるいは背景にある「関連要因」(cause/determinant)

例) 移動の足がない、うつなど自分の将来の健康に関心を持てない状態などで、ロジックモデルの初期・中期アウトカムに位置づけられるものにあたる。

c) 解決や改善に向けて、満たされることが望ましいニーズ (need)、「対策」

例) 移動支援や、健診受診の動機づけなどで、ロジックモデルのアクティビティ (対策) で満たされるニーズに相当する。

d) 対策が十分でなく、ニーズが満たされていないギャップ (gap)

例) コミュニティバスの便数が少ない、低迷する健診受診率などで、ロジックモデルのアウトプットに位置づけられるものである。

- ・ ①調査による現状値、困難事例などの現状把握を通じて誰にどのような「問題」がおきているのかを特定する。
- ・ ②先行調査研究を収集し「問題」を生み出している「関連要因」(cause/determinant) という意味での「課題」(**コラム 4 4 種類の課題**) を多面的に特定する。
- ・ ③施策・事業によってもたらされる住民・社会の望ましい姿・状態 (アウトカム) は何か、そのために施策や事業で介入すべき対象や、変えられる関連要因、その結果期待する変化などの大枠について、関係者で議論し合意する。
- ・ これらはロジックモデルを作成する中で精緻化されるが (2 章 ロジックモデルの作成)、まずは大枠の合意が必要となる。

合意に向けての留意点

- ・ 以下のような点に留意して、議論し、合意することが望ましい。

(1) 現状把握・対象

- ・ どのような現状や課題があるのか、大枠での共通認識が必要である。認識のずれを減らすために、現状や課題の多面性をリストアップし共通認識にする。
- ・ 問題が生じている対象や政策による介入対象、あるいは関連要因として情報収集をすべき対象には、個人、集団、地域、環境など、複数のレベルがある。
- ・ できる場合には、重視すべき主な対象と副次的な考慮すべき対象とを区別する。例

例えば、要介護者を主な対象とし、家族介護者を副次的な対象とする政策や、ハイリスクな個人を主な対象とし、地域住民集団を副次的な対象とする政策もある。

(2) 課題設定

- ・ 対策を考える上では、a)「問題」だけでは不十分である。b) その問題を生み出している「原因」あるいは背景にある「関連要因」の特定が必要である。例えば「閉じこもり」「引きこもり」という問題には、本人の成育過程や自己効力感、地域に出かける先がない、そこまでの移動の足がないなど、多くの関連要因があり、対策を講じるのはそれらのうち一部に留まることが多い。
- ・ c) 解決や改善に向けて、満たされることが望ましいニーズや「対策」まで掘り下げると、それが施策・事業の中身となりうる。
- ・ 一方で、d) 対策が十分でなく、ニーズが満たされていないギャップだけを課題として設定することは望ましくない。それ以外の対策の方が優れている可能性を見逃すことになるからである。

(3) アウトカムの定義

- ・ 保健・医療・介護（予防）に関わる施策・事業によって直接的・意図的にもたらされる望ましい姿には住民個人の身体的・心理的・社会的に良好な状態であるウェルビーイングからウェルビーイングなまちづくり、社会保障費用の抑制、CO₂排出量の抑制まで、地域レベルや長期的、多面的で間接的なもの（クロスセクター・ベネフィット：部門を超えた便益）までいろいろある。
- ・ 個人レベルに着目する施策・事業もあれば、集団や地域、社会のレベルに着目する政策もあり、立場によって重視するレベルや側面が異なるので、合意形成が必要になる。
- ・ 社会が多面的、多層的であることを共有した上で、以下について合意することが望ましい。
 - 政策・施策・事業の直接的・短期的な特定のアウトカムをどう定義するのか。間接的・長期的・多面的な社会的波及効果のどの範囲まで視野に入れるのか。
 - どこまでを定量的・定性的に、どのようなデータを用いて、どのような基準で評価するのかという評価の方法。

コラム 5

4つのインパクト

「インパクト」には4つの使われ方がある。

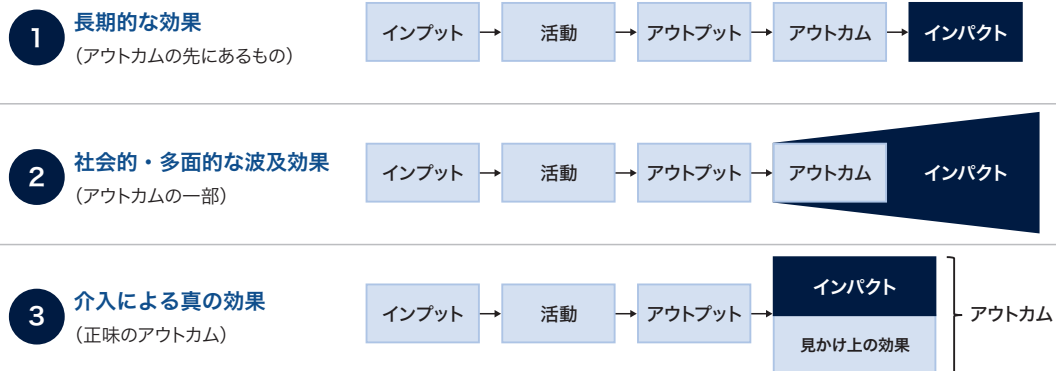
1つ目は、国際援助などの分野で政策等による長期的な効果を指す。

2つ目は、インパクト投資などの分野で多面的・社会的な波及効果・変化として捉えられるものである（例：環境保全事業においてCO₂削減効果が長期的な影響として期待できるが、厳密な測定や因果関係の立証は難しい）。どちらも「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、インパクトはアウトカムの後に起きるものとして表される。長期的・多面的な効果については、データの制約も多く、全ての側面について定量データに基づく因果関係の立証まで求められていない場合が多い。

3つ目は、学術的なプログラム評価研究の文脈で、政策や介入の実施後に観察される変化（アウトカム）のうち、他の影響による見かけ上の効果を取り除いた施策・事業で意図した真の効果を意味する。評価計画に基づき必要なデータ収集を行い定量評価が期待される。

4つ目は、上記の3つをすべて含むやや広い意味で使われているインパクトである。

本文書では、用語法の統一のため、4つ目に相当する 内閣府の定義「社会的インパクトとは、短期、長期の変化を含め、当該事業や活動の成果として生じた社会的、環境的な『アウトカム（効果）』をいう」⁽¹⁾を採用し、1つ目を長期的効果、2つ目を（多面的）波及効果、3つ目を見かけ上の効果を除いた効果と表記することとした。



(ソーシャルセクターのための事業マネジメント実践ガイドー混沌から「チームの道」を見出し、社会的インパクトを創出する⁽²⁾を一部改変)

図1 4つのインパクト

(4) アウトカム

- ・観察された変化のすべてが施策・事業の効果とは限らない。観察された変化には、見かけ上の効果や偶然・測定誤差が含まれている。
- ・アウトカムとは施策・事業によって期待される施策・事業対象に起きる直接的な変化である（コラム6 アウトカムとインパクト）。施策・事業によって誰が（主語）+どう変化するのか記述する。
- ・施策・事業によって起きる変化（アウトカム）には、初期的な変化から、数年かかる変化まである。

- ## コラム 6

2章 ロジックモデルの作成

- ・ロジックモデルとは、施策・事業の「論理（ロジック）」を端的に示した「模式図（モデル）」であり、基本的な枠組みを見える化するツールである。
- ・ロジックモデルの基本的な目的として、①施策・事業の形成・ブラッシュアップ、②対外的なコミュニケーション、③モニタリング・効果評価計画の立案がある⁽¹⁾⁽²⁾。

1) ロジックモデルの構成要素

ロジックモデルは以下の構成要素から成る⁽³⁾⁽⁴⁾。図3は基本的なロジックモデル(事業レベル)を示しており、構成要素を矢印でつなぎ、インプットからアウトカムまでの因果関係や目的-手段の関係を図で示している。

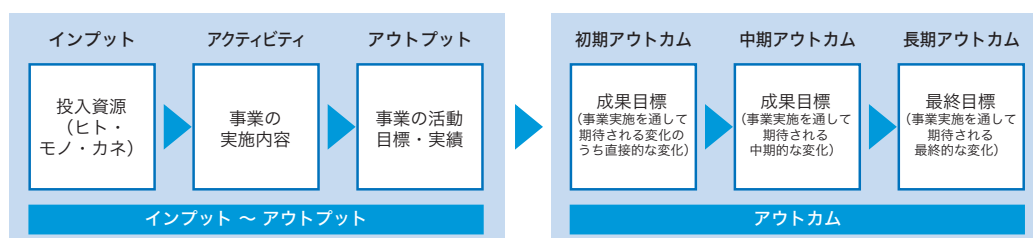


図3 事業のロジックモデル（単線フローチャート形式）⁽³⁾⁽⁴⁾

- インプット : 投入する資源（事業実施に必要なヒト・モノ・カネ等）
- アクティビティ : 活動（事業の実施内容等）
- アウトプット : 施策・事業による活動目標・実績（事業を通じて直接産出される、実施者が直接コントロールできる実績）
- アウトカム : 成果目標・実績（事業の実施を通じて期待される受益者の状態として現れる変化）

※アウトカムは、初期（短期）・中期・長期あるいは直接・中間・最終アウトカムといった時間軸で整理される。

初期アウトカム：アウトプットから直接的に影響を受ける変化

中期・長期アウトカム：初期アウトカムを経て得られる変化

施策のロジックモデルにおいては、他事業や他要因の影響を含めた複合的な変化でも良い。

コラム 7

ロジックモデルと Theory of Change

事業において、どのような資源を投入し、どのような活動を行い、それがどのようにアウトカムにつながるかという因果の流れを整理・共有するツールとして、ロジックモデルのほかに Theory of Change が用いられることがある。

ロジックモデルは、事業の評価や管理を目的として、投入から成果に至る主要な因果関係を、原則として線形の流れのように整理するなど、意図的に単純化された表現をとる。

一方、Theory of Change は、分岐や循環といった非線形の因果構造を含む場合もあり、因果関係が成立するための前提条件や仮定を明示する点に特徴がある。Theory of Change は、事業や施策が社会的変化を生み出すと想定される理論や仮説を、より明確に示すための枠組みである。

一般に、ロジックモデルは事業の構造や評価指標を整理する際に用いられ、Theory of Change は社会的変化が生じるメカニズムや前提条件を明示し、事業の背景にある理論を共有する際に用いられることが多い。

2) 施策・事業の目的、ロジックを明確にするための 4 ステップ

効果的・効率的な手段の選択と適切なエビデンスの設計・収集・分析を行うためには、施策・事業の目的とロジックを明確にすることが重要であり、そのためにロジックモデルを作成する。

ロジックモデルの作成にあたっては、まず(1) 施策・事業の目的を明確にするために3つの要素(①現状把握、②課題設定、③アウトカム)を整理する。次に、(2) 施策・事業の目的から手段に至るまでの論理的なつながり(ロジック)を明確にするため、4つの要素(③アウトカム、④アウトプット、⑤アクティビティ、⑥インプット)を整理する。これらは、バックカスティング法(未来起点・逆算方式)(**コラム 8 ロジックモデル作成における 2 つの型**)により組み立てられることが多い⁽¹⁾。

その後、(3) 作成したロジックモデルの妥当性を検討し、必要に応じて修正を行う。最後に、(4) 測定指標を設定する。

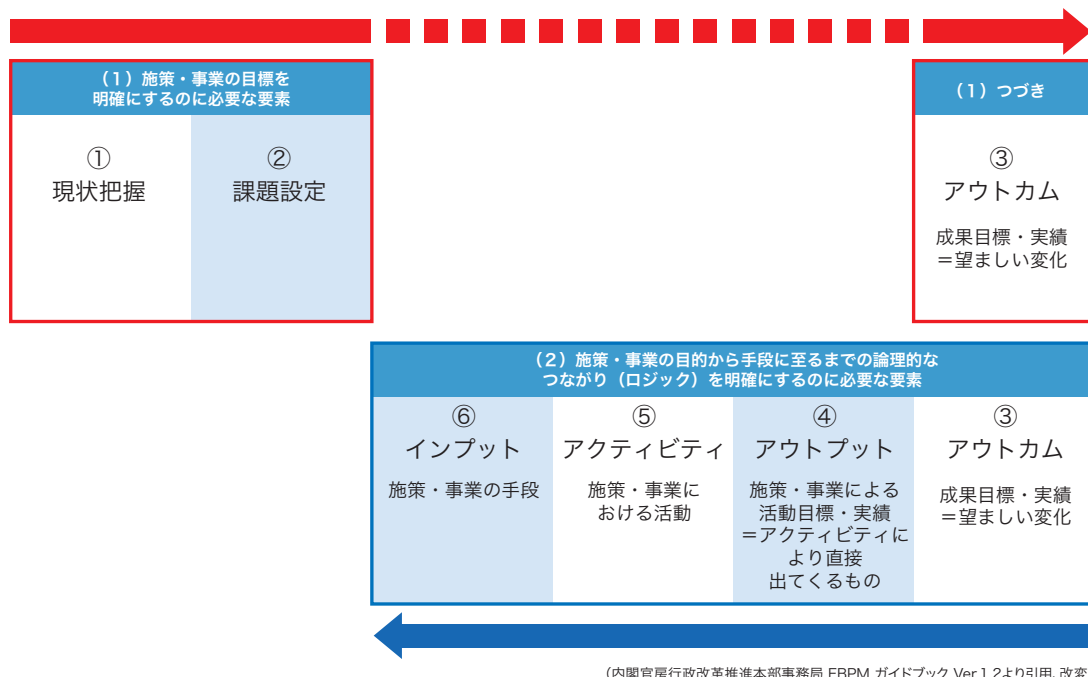


図4 ロジックモデル作成の流れ

施策・事業の目的とロジックを明確にするため、以下の4ステップの作業を行う。

1. 施策・事業の目的を明確にする

3つの要素（①～③）を繰り返し行ったり来たりしながら、定性的・具体的に検討を重ね、施策・事業の目的を明確にする。

2. ロジックを明確にする

4つの要素（③～⑥）を繰り返し行ったり来たりしながら、定性的・具体的に検討を重ね、施策・事業の手段から目的に至るまでの論理的なつながり（ロジック）を明確にする。

3. ロジックのブラッシュアップを行う

ロジックの裏付けとなるエビデンスを可能な限り追求し、各要素（①～⑥）を行ったり来たりしながら検討を重ねる。手段から目的までの論理的なつながり（ロジック）について妥当性を検討し、必要に応じて修正する。

4. 測定指標を設定する

ブラッシュアップが一定程度完了した後、③アウトカム、④アウトプットについて、測定指標の設定（具体化、定量化）を行う。

コラム 8

ロジックモデル作成における2つの型

ロジックモデルの作成は、現在から将来を予測するフォアキャスト方式（現状起点・予測方式）と、将来を起点として現在にさかのぼるバックキャスト方式（未来起点・逆算方式）のいずれか、または両方を用いて行う⁽³⁾⁽⁴⁾⁽⁵⁾。

ロジックモデルの左側（インプット）から右側（アウトカム）に向かって、時間の流れに沿って順番に整理する方法（フォアキャスト方式）は理解しやすい方法である。しかし、施策を構成する事業の内容や事業間の関係性、さらに施策の目指す長期アウトカムについて、あらかじめ明確なイメージを持っていないと作成が難しい。

一方、ロジックモデルの右側（アウトカム）から左側（インプット）に向かって作成する方法（バックキャスト方式）は、自治体の達成したい長期アウトカムが明確であり、それを実現するためにどのような活動が必要かを検討する場合に有効である。多くの場合、この方法が用いられる。

左側から右側、あるいは右側から左側といういずれかの一方向だけではなく、両方を併用してロジックモデルを作成する。因果関係の連鎖を上流から下流へ、また下流から上流へと、両方向から同時並行的に検討することで、ロジックモデルをより適切に構築することができる。
《主な手順》⁽³⁾

【フォアキャスト方式のロジックモデル作成】

手順1：生じている問題とその根本原因について構造的な整理を行い、課題を特定する。「なぜ問題が生じているのか」「現状はどうなっているのか」等を問い、各種調査や研究等も参照しながら、問題が生じるメカニズムを把握する。いわゆるロジカルシンキングの考え方に基づき、「なぜ」を繰り返しながら、漏れや重複がないかを吟味して整理する。

手順2：整理した要因のうち、どの課題を優先して取り組むべきかについて、エビデンス等を参考に検討する。あわせて、課題解決のために取り得る手段の選択肢を洗い出し、エビデンス等を踏まえて吟味・選択する。

手順3：ロジックモデルを作成する。施策手段を通した対象者に現れる変化について「アウトカム」に記入し、手順2で検討した手段を「アクティビティ」に記入する。さらに、アクティビティに必要な予算や人員等を「インプット」に、インプットとアクティビティによって施策の実施主体が直接的にコントロールできる産出物・成果物を「アウトプット」に記入する。→その後、最終手順に進む。

【バックキャスト方式のロジックモデル作成】

手順1：問題は一般に「理想の状態 (To Be) から現在の状態 (As Is) を差し引いた差分」として捉えられるが、バックキャスト（未来起点・逆算方式）では「理想の状態 ≡ 長期アウトカム」から逆算して考える。まず、施策を取り巻く背景について、エビデンス等を参考にしながら、「理想の状態」と「現在の状態」を整理する。

手順2：「理想の状態」から逆算し、それを実現するためには何が必要かを、エビデンスを参照しながら検討する。検討中の施策・事業がある場合は、それがどの部分に位置づけられるのかを考える。

手順3：「理想の状態」からの逆算と、施策・事業から期待される変化を行ったり来たりしながら吟味し、論理を精緻化する。

手順4：施策・事業案の妥当性について検討する。「その他にどのような手段があるか」「他の手段と比べて何が最適か」等を問いながら整理する。

手順5：これまでの検討事項を踏まえ、理想の状態に至るまでの段階を「長期アウトカム」から「中期アウトカム」「初期アウトカム」に遡り、記入する。その上で、施策手段を「アクティビティ」に、アクティビティに必要な予算や人員等を「インプット」に、産出物・成果物を「アウトプット」に記入する。→その後、最終手順に進む。

《最終手順》

アウトプットおよびアウトカムの状況を適切に把握できる指標を設定する。その際には、①取得可能なデータに基づく指標とすること、②「事業が適切に実施されているか」、「成果につながっているか」を確認できる評価指標とすることが重要である。また可能な限り、③想定する分析手法も念頭において定性・半定量・定量指標を設定することが望ましい。指標を設定した後は、各要素間の論理の流れについて再度吟味し、必要に応じて修正を行う（4章評価計画の立案を参照）。

3) ロジックモデル作成の手順

(1) 現状把握・エビデンスの収集

特定された社会的ニーズ・課題を踏まえ、施策・事業（社会的ニーズ・課題を解決する介入行動）を実施する背景や現状を把握する。そのためには、エビデンスの収集が必要である。具体的には、ステークホルダー（国民、民間事業者、有識者、地方自治体、他の行政機関など）との丁寧なコミュニケーションを通じた聞き取りに加え、政府統計、行政記録情報、民間データ（ビッグデータ、オルタナティブデータ等）など、

多様なデータを活用する。

データの活用にあたっては、正確性・速報性・網羅性のバランスを考慮することが重要である。また、各府省庁の職員による分析や専門家による高度な分析を踏まえ、現状の把握に加えて将来の状況についても検討する。

さらに、施策の立案や状況変化に応じた柔軟な見直しを行うためには、ステークホルダーと円滑に意見交換を行い、フィードバックを得られる仕組みを整えることが重要である。立案段階からコミュニケーションを図りながら、施策の目的や内容に応じて、施策実施における役割分担や運用方法等について検討する⁽¹⁾。

コラム 9

エビデンスとは何か

エビデンスとは、施策・事業や介入によって生じる結果について、因果関係を裏付ける根拠を指す。狭義には、施策・事業手段の有効性を示す根拠をいい、ロジックモデルにおいては各構成要素間の「矢印（因果関係）」が成立する根拠となる。広義には、施策の必要性（ニーズ）を示す根拠（データ、ファクト等）を含む⁽¹⁾。

	種類	内容	困窮世帯の子供に対する支援策における具体例
エビデンス (広義)	データ、 ファクト等	データ、ファクト	統計データ、ヒアリング等による 現状把握のための情報
		将来予測	現状のまま推移した場合等の 将来予測
	エビデンス (狭義)	定量的な因果効果	ランダム化比較試験によって明らかになった 教育プログラムの効果 生活習慣の良い子どもと悪い子どもの学力の比較
		定性的な因果効果	フォーカスグループインタビューに基づいて 明らかになった生活習慣改善プログラムの効果

図5 エビデンスの種類と具体例⁽¹⁾

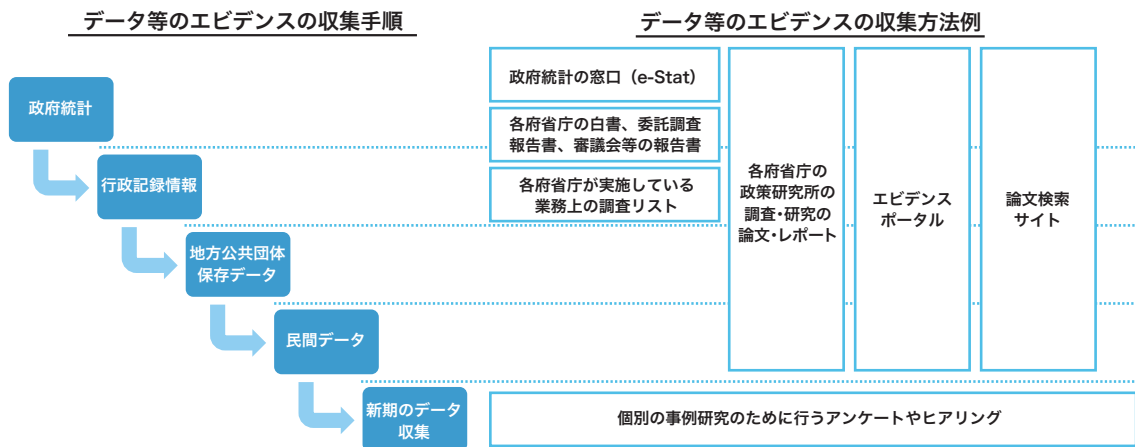
データ・ファクト等＝施策の必要性（ニーズ）を示す根拠

エビデンス（狭義）＝施策・事業手段の有効性を示す根拠

コラム
10

データ等のエビデンス（広義）の収集作業

現状把握および効果検証を行うためには、データ等のエビデンスの収集が不可欠である。アンケート調査などの新たなデータ収集を実施する前に、まずは既存のエビデンス（広義）の収集を行うことが重要である⁽¹⁾。

図6 エビデンスの収集手順と収集方法例⁽¹⁾コラム
11

PICO の枠組み

施策・事業の目的を実現するためには、候補となる施策・事業手段が有効なものかどうかについて、既存のエビデンス（施策手段の有効性の判断根拠となる研究成果や海外事例等）を収集・分析し、検討することが重要である。

保健医療分野においては、PICO の枠組みに沿って収集する情報を整理することで、既存のエビデンスを効率的に収集することができる。

PICO の枠組み	
① 施策手段の対象 (Population)	施策手段の対象となる個人や事業者は誰か？
② 施策手段の内容 (Intervention)	施策手段の具体的な内容は何か？
③ 施策手段の対象との比較対象 (Comparison)	効果検証にあたって、施策手段の対象ではない誰と比較するのか？
④ 施策手段の成果目標 (アウトカム) (Outcome)	施策手段により改善したい定量的な成果目標指標 (アウトカム指標) は何か？

(内閣官房行政改革推進本部事務局 EBPM ガイドブック Ver.1.2より引用、改変)

図7 PICO の枠組み⁽¹⁾

既存のエビデンスの収集方法

PICO の枠組みに従って収集する情報を整理し、それを基に、例えば A ～ G の方法で既存のエビデンスを収集する。その際、公表されているマクロデータのみならず、個人情報保護に留意しつつ、非公表のマイクロデータ（調査票情報等）の利用も検討する⁽¹⁾。

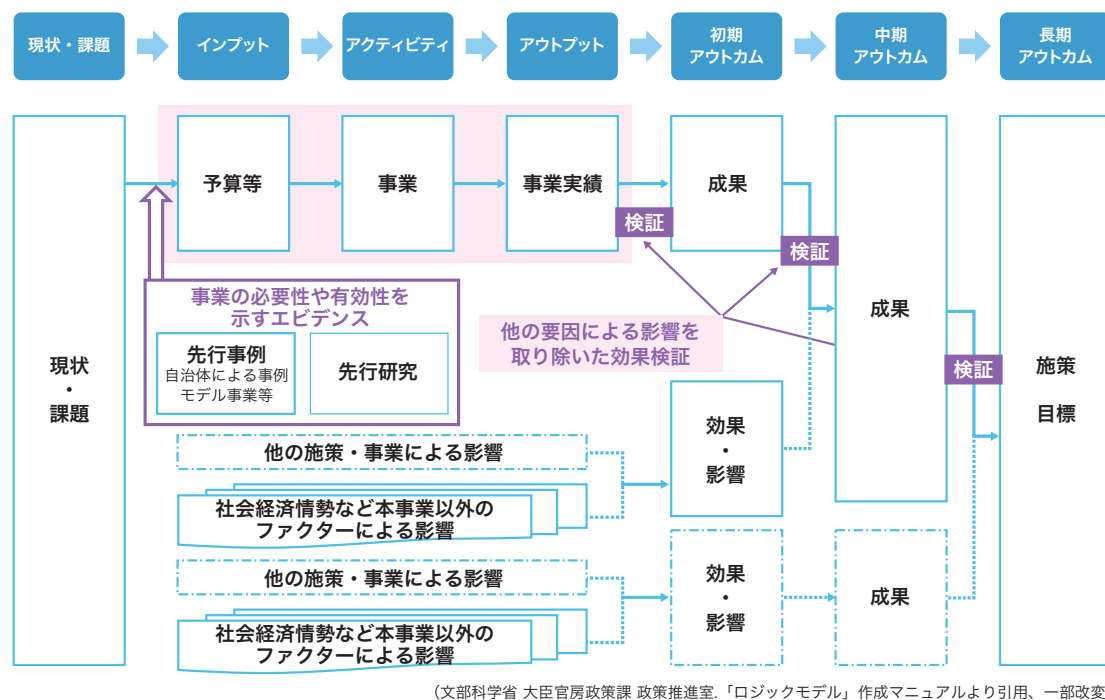
データ等のエビデンスの収集方法例	
A	政府統計の窓口(e-Stat) (https://www.e-stat.go.jp/)
B	各府省庁の白書、委託調査報告書、審議会等の報告書
C	各府省等が実施している業務上の調査リスト 内閣官房行政改革推進本部事務局等が構築した、各府省等(所管する独立行政法人も含む)が行う調査等を改善するための恒常的な仕組みの下、整備された調査等リスト。 各府省とりまとめ部局が、①逐次更新される各府省等が実施している業務上の調査等リスト、②内閣官房行政改革推進本部事務局等により半期ごと(5月、11月)に更新される全府省版の調査等リストを把握。
D	各府省庁の政策研究所の調査・研究の論文・レポート エビデンスポータル 既存のエビデンスを整理し、概要や質などを一覧できる形で公表しているプラットフォームや、エビデンスを新たに構築・整理し公表しているプラットフォーム。ただし、ポータルが作成されているトピックはまだ少なく、カバーされている取組も限られている。 ○経済・財政一体改革エビデンス整備プラン2022 (https://www5.cao.go.jp/keizai-shimon/kaigi/special/reform/ab1/20220411/shiryoku21.pdf) 新経済・財政再生計画改革工程表の改善に向けて、①改革工程表の「政策目標」、「KPI」等のつながり(ロジック)の検証、②必要なエビデンスの構築、のための取組・工程等を示すもの。経済財政諮問会議の下に設置された経済・財政一体改革推進委員会EBPMアドバイザリーボードが取りまとめ、毎年度改定。 ※人材育成・職業訓練、社会保障、社会資本整備等、地方行政改革等、文教・科学技術 ○経済・財政と暮らしの指標「見える化」データベース (https://www5.cao.go.jp/keizai-shimon/kaigi/special/reform/mieruka/index.html) E 経済・財政一体改革を着実に進めるため、自治体別および時系列に整備した各種データ・指標の比較により経済・財政と暮らしに係る様々な地域差を「見える化」できるようにしたデータベース。47都道府県別/1741市区町村別および1975年～のデータを利用できる。 ○内閣府エビデンスシステム(e-CSTI) (https://e-csti.go.jp/) 大学等の研究機関における「研究」、「教育」、「資金獲得」に関するエビデンス*を収集し、インプットとアウトプットの関係性を「見える化」するための各種分析機能を開発し、関係省庁や国立大学・研究開発法人等の関係機関に対して分析機能・データを共有するための内閣府が運営するプラットフォーム。 ※科学技術関係予算、国立大学・研究開発法人等の研究力、大学・研究開発法人等の外部資金・寄付金獲得、人材育成に係る産業界ニーズ、地域における大学等の目指すべきビジョン ○EBPM データベース(https://cyberagentailab.github.io/EBPMDB/) 学術研究等から得られたエビデンスを、中央省庁や自治体などの政策現場で活用できる形でサマリー記事として収集・共有するプラットフォーム。民間企業や大学等研究機関の研究者が自主的に運営(開発・運営:株式会社サイバーエージェント、運営協力:三菱UFJリサーチ&コンサルティング株式会社)。 論文検索サイト 学術論文の検索に特化した検索サイト。記述研究や理論研究などあらゆる内容の学術論文を対象としているため、広範な論文検索ができる一方、自ら効果検証型の研究を選り分ける必要がある。また、学術論文といえども、質の高いものとそうでないものが混在しているため、質の吟味も重要になる。 ○CiNii Articles(https://support.nii.ac.jp/ja/cia/cinii_articles) 大学共同利用機関法人情報・システム研究機構国立情報学研究所が運営する、学協会刊行物・大学研究紀要・国立国会図書館の雑誌記事索引データベースなど、学術論文情報を検索の対象とする論文データベース・サービス。 F ○researchmap(https://researchmap.jp/) 国立研究開発法人科学技術振興機構が運営する、研究者が業績を管理・発信できるようにすることを目的とした、データベース型研究者総覧。 ○Google Scholar(https://scholar.google.co.jp/schhp?hl=ja) Google が提供する、幅広い分野・国の学術文献の検索サービス。
G	学会のホームページ、学会誌、研究者のホームページ
H	事例研究 政府統計、行政記録情報、地方公共団体保有データや民間データといった既存のデータ等(非公表のマイクロデータ(調査票情報等)の二次利用を含む)では必要十分なデータ等のエビデンスが得られない場合、定量的なエビデンスを補強するために行う個別のアンケートやヒアリングによる事例研究。

図 8 データ等のエビデンスの収集方法例⁽¹⁾

コラム
13

ロジックモデルとエビデンスの関係性

課題を解決するための事業内容を検討する際には、先行事例や先行研究から得られたエビデンスを収集し、インプットからアウトプット、アウトカムに至る因果経路の妥当性を検討する必要がある⁽³⁾。また、計画・実施された事業については、アウトプットおよびアウトカムの効果検証を行い、その結果を公表することが重要である。これにより、当該事業の成果が新たなエビデンスとして蓄積され、今後の事業設計や施策立案に活用されることが期待される⁽¹⁾。



(文部科学省 大臣官房政策課 政策推進室、「ロジックモデル」作成マニュアルより引用、一部改変)

図9 ロジックモデルとエビデンスの関係性⁽¹⁾

(2) 課題設定

現状把握を行いながら、施策・事業課題に関する仮説を構築し、「望ましい姿」（長期アウトカム）と「現状」とのギャップを明確にした上で、解決すべき社会課題を設定する。この際、解決策をあらかじめ埋め込まず、まずは問題の説明にとどめること、要望と課題を区別して整理することが重要である。

さらに、対象とする社会課題と現状把握の内容を組み合わせ、社会課題の発生メカニズムの背後にある根本原因を追究する。その上で、根本原因に対処する施策を立案するのか、表面的な関連要因への対応を行うのかについて検討する。

(3) アウトカム設定

まず、施策・事業内容を検討するにあたり、現状を踏まえつつ、住民・社会の望ましい姿を長期アウトカムとして設定する。検討の初期段階では、その問題解決が、元

のあるべき状態に戻すものなのか、それとも新たな望ましい状態に到達するためのものなのかが整理できていれば十分である。併せて、いつの時点までに到達を目指すのかについても設定することが望ましい。検討を進めロジックモデルをブラッシュアップしていく過程で、これらの内容は適宜見直しを行う⁽¹⁾。

次に、段階的なアウトカムを設定する。

アウトカムは、原則として受益者の視点から、複数の段階に分けて設定する⁽¹⁾。段階の分け方は様々であるが（初期・中間・最終、短期・中期・長期、初期・中長期など）、先に設定した長期アウトカムから順にさかのぼり、初期（短期）アウトカムは、アウトプットにより最初に生じる受益者の状態の変化を設定する。この際、変化の対象となる主体（主語）と変化の内容を明確に記述することが重要である。

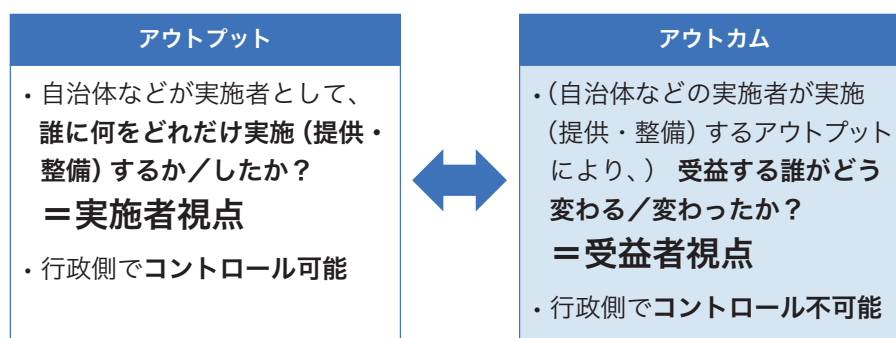


図 10 アウトプットとアウトカムの違い⁽¹⁾

施策・事業のアウトカム（成果目標）は原則として、①明確であること（Specific）、②測定可能であること（Measurable）、③達成可能であること（Achievable）、④現実的であること（Realistic）、⑤期間が明確であること（Time-limited）という、SMART の原則を満たすことが重要である⁽¹⁾。ただし、②測定可能であることと、③達成可能であること、さらに施策・事業の真の目的を適切に反映することとの間にはトレードオフが存在する。

厳密に測定可能な指標は、短期的かつ限定的なものになりやすく、施策・事業が目指すより広範な社会変化を十分に反映しない場合がある。そのため、データを収集しにくいという理由のみで、代替的に測定しやすいアウトカムを設定し、施策・事業のロジックを歪めることは避けるべきである。一方で、達成可能性を欠く過度に理想的なアウトカム設定とならないよう留意する必要がある。

例：「健康長寿社会の実現」を測定可能な形で具体化すると「健康寿命の延伸」となり、例えば「2040 年までに 80 歳」といった具体的な数値目標を設定することができる。詳細は、P.29 ～ 32 のロジックモデルの例や P.31、32 の具体的な事例を参照。

(4) アウトプット設定

アウトプットは、原則として実施者の視点から設定する⁽¹⁾。実施者は自治体、民間事業者、NPO 等の可能性がある。

例：アウトプット：「地域において健康教室が提供される」

⇔ アウトカム：「住民の健康に対する関心が高まる」

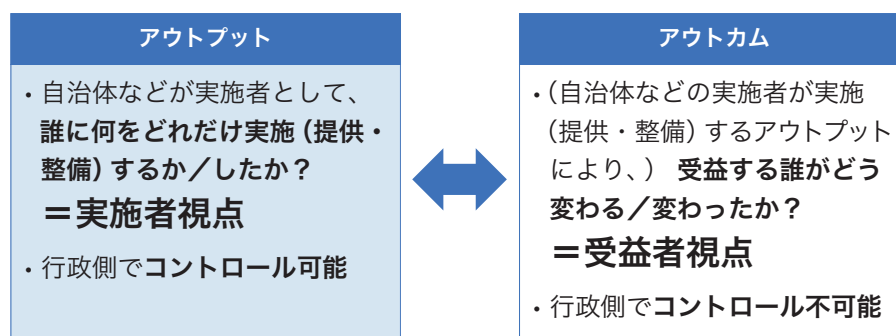


図 11 アウトプットとアウトカムの違い⁽¹⁾

(5) アクティビティ設定

実施者を具体的に設定した上で、実施者の視点からアクティビティ（具体的な事業内容）を設定する⁽¹⁾。

例：健康教室の開催、健康に関する広報誌の発行

(6) インプット設定

インプット（施策・事業手段）を設定する。これには予算、税制措置、規制の強化・緩和、人的資源・物的資源、制度・計画などが含まれる。施策目的を実現するために効果的かつ効率的な手段を選択することが重要である。「予算ありき」で施策を設計するのではなく、ロジックに矛盾や飛躍がないかを検討しながら、適切な施策手段を設定する⁽¹⁾。

(7) 測定指標設定

施策目的から施策手段に至るまでの論理的なつながり（ロジック）を明確にした上で、成果目標（アウトカム・インパクト）および活動目標（アウトプット）について、達成度が客観的に判断できる測定指標を設定する⁽¹⁾。少なくともその一部については定量的な指標を設定する。KPI（Key Performance Indicator: 主要業績評価指標）とは、事業の進捗状況や成果を客観的に測定するための重要な指標である。

また、施策・事業が立案段階で想定したとおりに実施されているかを確認し、モニタリングを通じてボトルネックを早期に発見するため、短期間で測定できる指標を設定し、初期・中期など複数の段階にわたって測定指標を設定することが望ましい。特に、アウトプット→初期アウトカムへのつながりを検証できる指標を設定することは重要である。

コラム 14

PFS における測定指標

PFS では、成果に応じて支払額が増減する「成果指標」のほか、支払額とは直接連動しない「測定指標」が設定される場合がある⁽⁵⁾。後者は、参加者の属性把握、施策と効果の因果関係の検証のための中間アウトカム指標、医療費や介護費などの社会的便益の算出などに用いられる。

成果指標を設定する際には、何を最終的に評価するのか（費用抑制、受益者の便益、より上位の公共目的など）を明確にすることが重要である。これが曖昧な場合、手段の目的化、対象者の選別、望ましくないインセンティブの発生など、政策決定プロセスに歪みが生じる可能性がある。そのため、(i) 最終目的（公共目的）と成果指標の整合性を確認すること、(ii) 副作用・公平性・選別の可能性を検証すること、(iii) 目的の矮小化を回避すること、(iv) 契約前に評価設計とデータアクセスを確定しておくことなどに留意する必要がある。

なお、指標数が増えるとデータ収集コストやモニタリング負担、支払算定の複雑化などの負担が増加するため、支払額と連動する成果指標は一般に2～5個程度に設定されることが多い。また指標によって変化の確認に要する時間が異なるため、1年後から最終年度までのどの時点でどの測定指標を用いるのかを考慮して設定する必要がある。

コラム 15

評価の5原則

施策・事業の評価は、以下5つの原則に基づいて行うことが重要である⁽⁶⁾。

1. 多様な関係者の参加、連携、協働：対象となる組織、事業に関わる多様な関係者の幅広い参加、連携、協働により、目指す価値や評価目的について合意形成を目指す。
2. 信頼性：信頼できる方法で収集するなど適切な情報を使用し、過剰な成果を主張するような評価報告を避ける。
3. 透明性：実施過程の活動状況や調査方法、成果などについて、正確かつ誠意ある情報開示・説明・報告を行う。
4. 重要性：事業の実施や意思決定にとって重要な事項を選択し、重点的に評価する。
5. 比例性：評価の目的やニーズ、利用可能な資源（予算、人員など）を考慮し、組織や事業の規模に比例して、過度の負担がかからない評価方法や情報開示の方法を選択する。

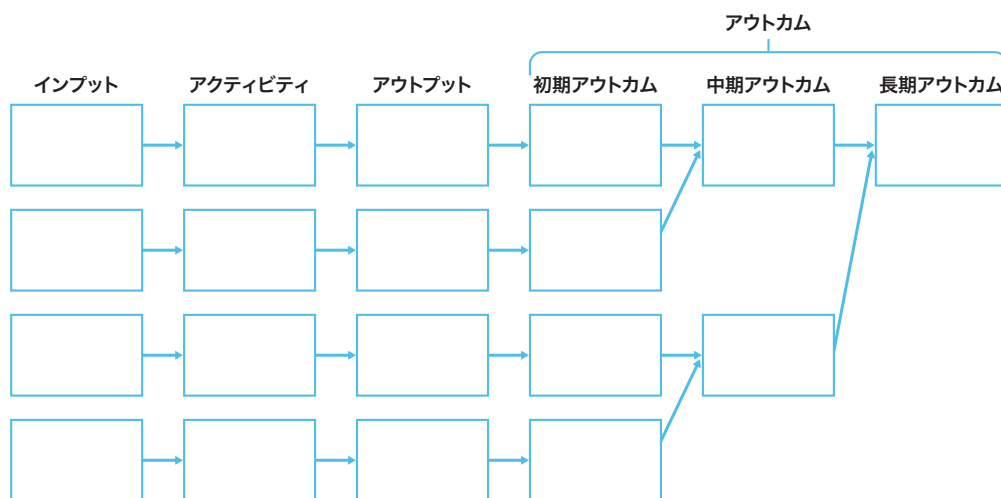
コラム
16

ロジックモデルの形式

施策・事業における活動が成果（アウトカム）につながるまでの過程を何段階に分けるか、また各段階をどのような名称で呼ぶか、どのような形式で示すかについては様々な方法がある。ロジックの整理と明確化が可能であれば、ロジックモデルの様式は自由である。

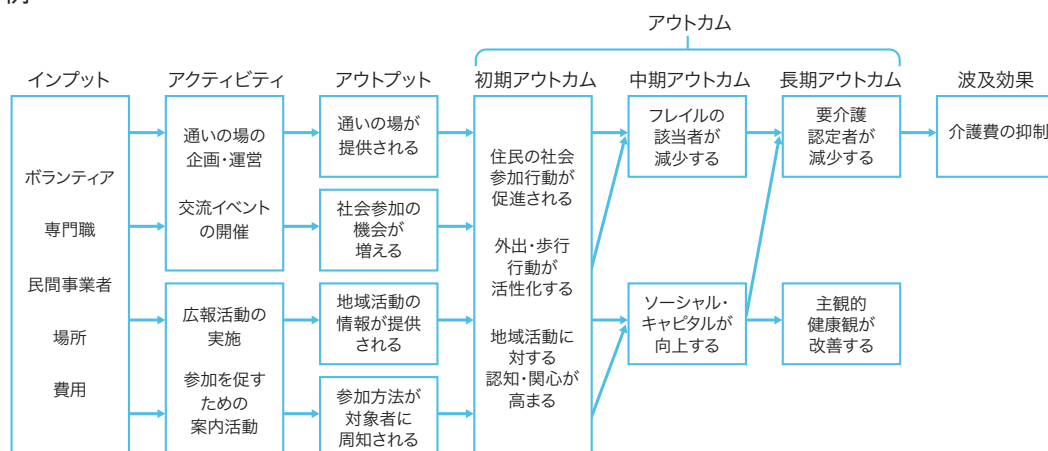
代表的なロジックモデルとして、1) に示した「事業のロジックモデル（単線フローチャート形式）」の他、施策間の関係性を示すものとして以下のようなものがある⁽⁴⁾。

(1) 施策のロジックモデル①（複線フローチャート形式）

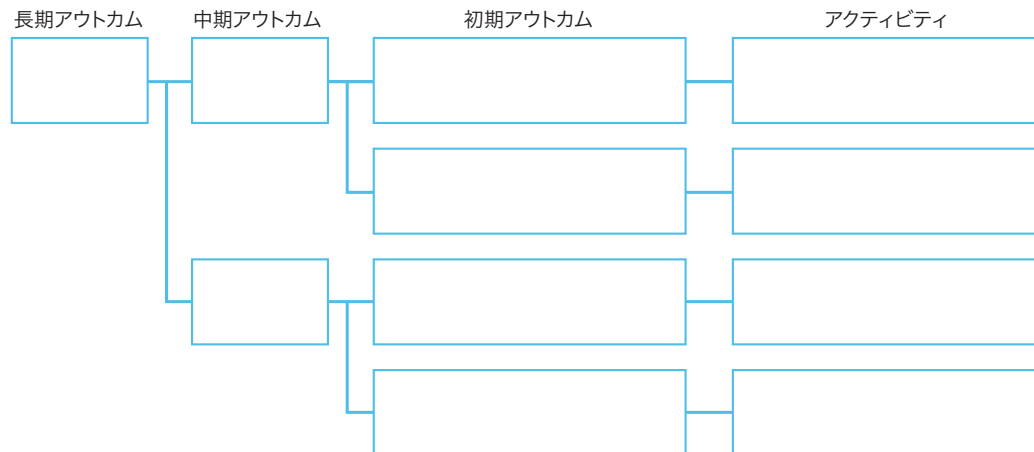


インプットから長期アウトカムまでの流れを、施策単位で示すことができる。施策や機能が複数存在し、因果関係が分岐・合流する場合に適している。構造が複雑になりやすいため、作成には一定の時間を要するが、計画の初期から中期にかけての探索段階において有効である。

例：

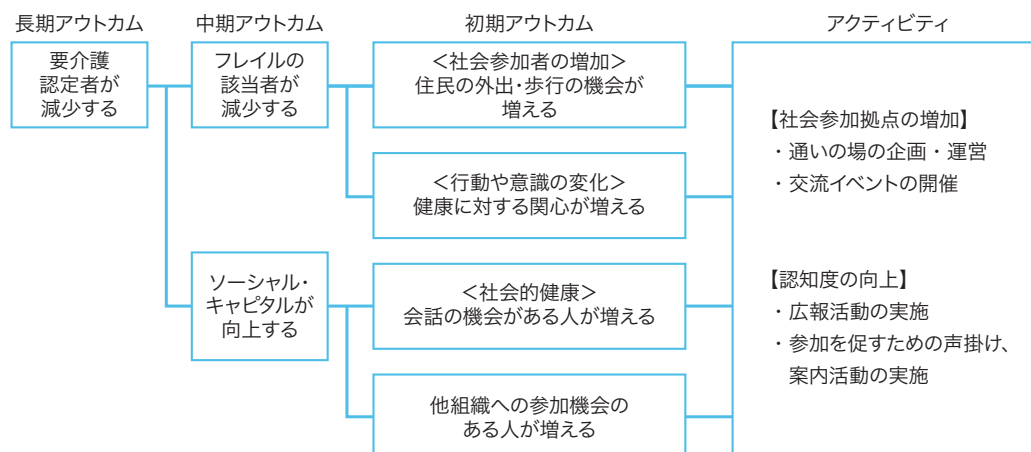


(2) 施策のロジックモデル②（体系図形式）



「目的から出発して必要な手段を検討する」という発想に基づく形式であり、施策の全体像や構造を整理するのに適している。ステークホルダー間の合意形成や説明資料の冒頭などで用いられることが多い。因果関係の強弱や時間軸、検証の視点の表現にはやや弱い面があるが、シンプルで直感的に理解しやすく、上位目的との整合性を示しやすい。必要に応じて、初期アウトカムとアクティビティの間にアウトプットを配置することも可能である。

例：



(3) 表形式のロジックモデル

インプット	アクティビティ	アウトプット	アウトカム		
			初期	中期	長期
インプットA	アクティビティA	アウトプットA	初期アウトカムA	中期アウトカムA	長期アウトカムA
インプットB	アクティビティB	アウトプットB	初期アウトカムB	中期アウトカムB	
インプットC	アクティビティC	アウトプットC	初期アウトカムC		

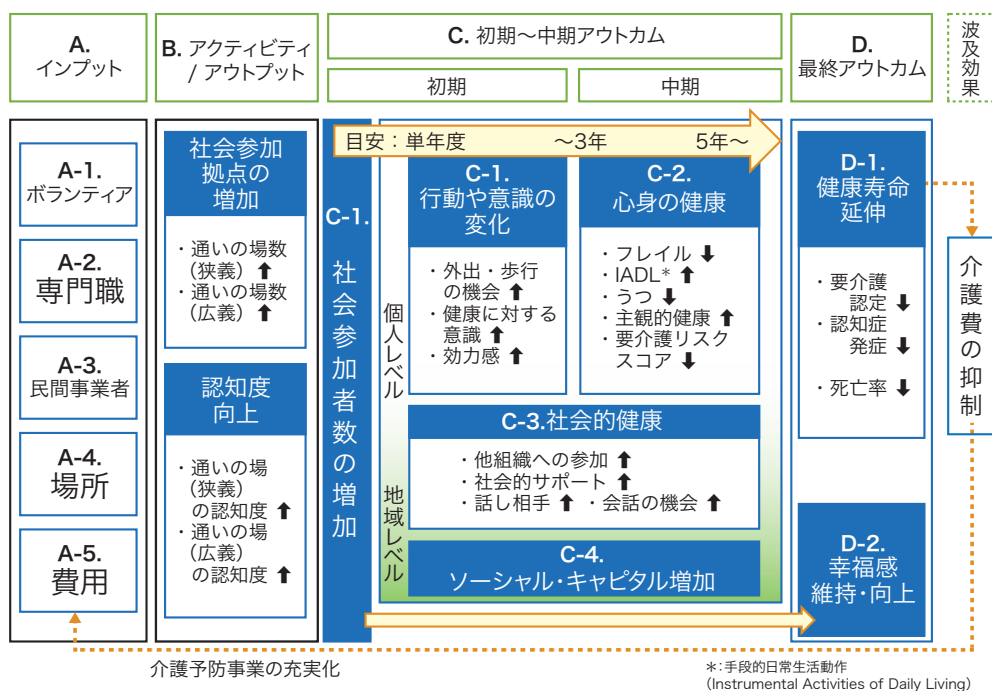
各項目に具体的な内容を記入することで作成できるため、比較的簡単に作成できる。一方で、項目間の論理的関係が矢印や線で示されないため、因果関係がやや分かりにくくなる場合がある。そのため、計画や進捗管理、レビュー作業などに適しており、実行段階や報告資料として活用されることが多い。

例：

インプット	アクティビティ	アウトプット	アウトカム		
			初期	中期	長期
ボランティア	通いの場の企画・運営	通いの場が提供される	住民の社会参加行動が促進される	フレイルの該当者が減少する	要介護認定者が減少する
専門職	交流イベントの開催	社会参加の機会が増える	会話の機会が増える	ソーシャル・キャピタルが向上する	主観的健康観が改善する
場所・費用	広報活動の実施	地域活動の情報が提供される	地域活動に対する認知・関心が高まる		

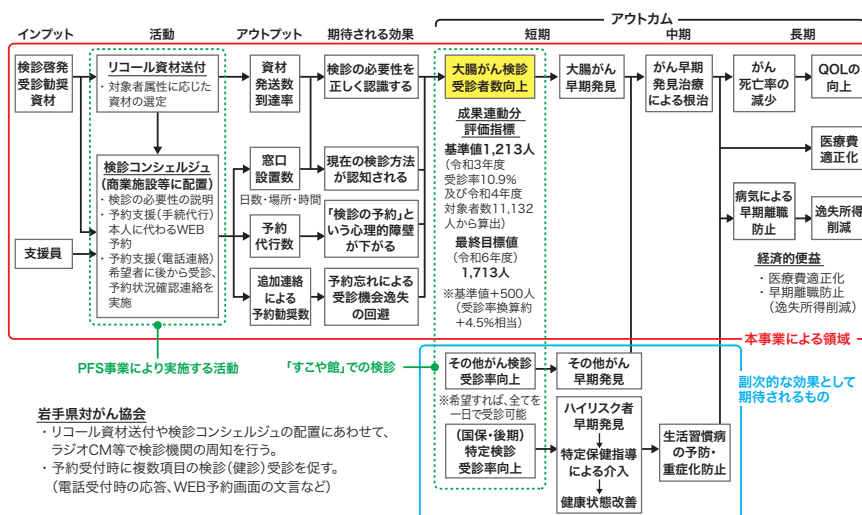
具体的な事例①

本ロジックモデル⁽⁷⁾は、通いの場事業（アクティビティ）を行うことで、地域における「社会参加者数の増加」や「社会的・身体的・精神的健康の向上」を促し、最終的に「健康寿命の延伸」や「幸福度の維持・向上」を目指す取り組みの全体像を示したものである。事業の投入資源（インプット）からアクティビティ/アウトプット、短期～中期アウトカム、そして最終アウトカムへと連なる因果関係を整理している。



具体的な事例②

本ロジックモデル⁽⁸⁾は、がん検診受診率の向上を通じて「がんの早期発見・早期治療」を実現し、最終的に「がん死亡率の減少」と「QOL（生活の質）の向上」を目指す取り組みの全体像を示したものである。検診啓発や受診勧奨を支える資源（インプット）を基盤に、リコール資料送付や検診コンシェルジュ設置などの活動を行い、その成果として検診受診者数や検診方法の理解度が向上する。これにより短期的には大腸がん検診の受診率上昇と早期発見が促進され、中期的にはがんの早期治療・死亡率低下へとつながる。最終的には、疾病負担の軽減を通じて県民のQOLが向上し医療費の適正化へとつながる因果関係を整理している。



【参考文献】

- (1) 内閣官房行政改革推進本部事務局. EBPM ガイドブック Ver.1.2 2023.
Available from : https://www.gyokaku.go.jp/ebpm/img/guidebook1.2_230403.pdf
- (2) 内閣官房行政改革推進本部事務局. 行政事業レビューシート作成ガイドブック Ver.1.2 2024.
Available from : https://www.gyokaku.go.jp/review/img/R07sakusei-guidebook_ver1-2.pdf
- (3) 文部科学省 大臣官房政策課 政策推進室. 「ロジックモデル」作成マニュアル 2023.
Available from : https://www.mext.go.jp/content/20260108-mxt_kanseisk01-100000155_01.pdf
- (4) 佐藤 徹. エビデンスに基づく自治体政策入門～ロジックモデルの作り方・活かし方. 公職研 ; 2021.
- (5) 内閣府成果連動型事業推進室 成果連動型民間委託契約方式（PFS: Pay For Success）共通ガイドライン 2021. Available from : https://www8.cao.go.jp/pfs/r6_guidelines.pdf
- (6) 一般財団法人 日本民間公益活動連携機構. 休眠預金活用における社会的インパクト評価. Available from: https://www.janpia.or.jp/_assets/download/dormant-deposits/ongoing/guideline/impact_hyoka.pdf
- (7) 井手 一茂, 近藤 克則. 介護予防の効果—医療経済学的な立場から—. 老年社会科学. 2023;44(4):392-398.
- (8) 内閣府. 成果連動型民間契約方式（PFS）による大腸がん検診の受診率向上プロジェクト事業（岩手県 矢巾町）2025. Available from : <https://www8.cao.go.jp/pfs/jirei/yahabachor6.pdf>

3章 社会的インパクト予測評価

- ・日本語では同じ「評価」と呼ばれるものにも「予測評価」「プロセス（中間）評価」「事後評価」の3つがある（コラム14 アセスメントとエヴァリュエーション）。
- ・本章で取り上げるのは、政策を決定・導入する前に、それを導入した場合の結果や影響を、事前に予測して評価する「予測評価（アセスメント）」である。（藤野・近藤、2011）

コラム 17

アセスメント (assessment) とエヴァリュエーション (evaluation)

日本語では、同じ「評価」と訳されるが、英語では元の単語が異なることがある。assessment は、「査定」や「事前評価」「予測評価」など、契約や介入などの前に評価する時に用いられる。一方、evaluation は、介入が始まった後のプロセス評価や事後的な評価を意味する。ただし文脈や慣用などのために、例外はある。

1) インパクト予測評価

- ・日本ですでに実施されている「予測評価」として、大規模開発前に実施が義務づけられている「環境影響アセスメント」がある。
- ・他に、広範な社会的インパクトを評価する社会的インパクト予測評価（Social Impact Assessment：SIA）、環境も合わせて評価するときには環境・社会インパクト予測評価（Environmental and Social Impact Assessment：ESIA）、健康面への影響のアセスメントを、健康影響予測評価（Health Impact Assessment：HIA）⁽¹⁾と呼ぶ。WHO は HIA のことを「政策、施策、事業が潜在的に集団に与える健康影響や、集団中の 影響の違いなどについて判断するための一連のプロセス、方法、およびツール」と定義している⁽²⁾。
- ・環境や健康に限らず、社会的インパクトを広く事前（予測）評価することを、ここでは社会的インパクト予測評価と呼ぶこととする。

2) 社会的インパクト予測評価の視点と手順

- ・政策等による望ましい（正の）影響と（負の）悪影響の両面を、下記のような事項について検討する。
 - 社会的に不利な状況にあり、影響を受けやすい集団を把握する。新しい運動施設開設の例では、近隣に比べ遠隔地の住民、運動団体にすでに参加している者に比べ参加していない者などで、利用機会が少なく、恩恵が得られにくい、など。
 - 新しく提案された政策が、既知の正と負の影響を助長させる可能性について検討する。新しい運動施設開設の例では、身体活動や住民の交流の機会の増加などの

正の影響と、イベント開催時の渋滞やゴミ問題などの負の影響、など。

- ・新しく提案された政策が、特定の集団に新たな正と負の影響をもたらす可能性について検討する。新しい運動施設開設の例では、近隣住民に新たな正と負の影響が及びやすい。一方で、遠方の住民にはアクセスが悪いことによって利用機会が少ないために、どちらの影響も及びにくい、建設・運営コストは住民税を通じて負担を求められる、など。社会的インパクト予測評価では、定量データだけでなく、参加型アプローチなどの質的手法も含め、利用しうる範囲で最上のエビデンスを提供する。
 - ・望ましい正の影響と負の悪影響の両面を捉え必要な修正を議論する。健康格差を拡大せず縮小するよう、施策・事業の決定プロセスに反映する推奨意見を提出する。
 - ・社会的インパクト予測評価の効用は、社会的インパクト予測評価によって提示される推奨意見のみではなく、その実施プロセスによっても得られることが指摘されている。
- 負の悪影響を受けると予測される集団が、社会的インパクト予測評価の過程に関与することで意見を述べる機会が提供され、情報が速やかに開示されることで集団の自己効力感を高め、それが意思決定への参加機会の増加をもたらし、より良い関係での交渉の機会を作り出すことが期待される。

コラム 18

ヘルスプロモーション

ヘルスプロモーションを提唱したオタワ憲章(3)においても、健康の前提条件として、平和、住居、教育、食料、収入、安定した環境、持続可能な資源、社会的正義と公正が謳われている。HIAはこの社会環境モデルに基づいて、広範な政策を評価の対象とする。

【参考文献】

- (1) 日本公衆衛生学会 公衆衛生モニタリング・レポート委員会．健康影響予測評価（Health Impact Assessment）の必要性と日本公衆衛生学会版ガイダンスの提案．日本公衆衛生雑誌．2011;58(11):989-92.
- (2) World Health Organization (WHO). Health Impact Assessment 2004. Available from: <http://www.who.int/hia/en/>.
- (3) WHO. Ottawa Charter for Health Promotion（島内憲夫訳：21世紀の健康戦略2 ヘルスプロモーション，垣内出版，1990）．1986.

4 章 評価計画の立案

- ・ 施策・事業の効果評価を元に PDCA（Plan-Do-Check-Act）サイクルを回して効果を最大化するためには、実施前に評価計画を立案することが重要である。
- ・ 事前評価については、1 章、2 章で説明したので、本章では中間・事後評価の計画立案に焦点を当てる。

1) 評価の方法とエビデンス

- ・ 評価は、施策・事業の継続・改善・見直しに加え、対象地域・対象者の拡大や実施規模の拡充（次年度の予算措置を含む）といった意思決定に資する情報を得ることを目的とする。評価・施策・事業のステージ（**コラム 2 政策と政策評価環境の発展段階（成熟度）**）の違いによって評価の力点を変える必要がある。
- ・ 評価のエビデンスには、定量データ（参加者数、実施回数、指標の変化等）と、定性データ（インタビュー、観察、自由記述、ケース記録等）がある。定量データは、施策・事業の実施量や成果の程度、変化を数量として示すことができる。定性データは、利用者や関係者が経験した変化、その意味づけ、成果や課題が生じた背景・理由など、数量だけでは捉えにくい側面を示すことができる。実施状況や成果は、定量データと定性データのいずれからも捉えることができ、両者を組み合わせることで、施策・事業を多面的に評価できる。
- ・ 評価方法は、評価の目的や明らかにしたいこと、施策・事業の段階、利用可能なデータ、予算・期間等の条件を踏まえて選択する。データの収集には、質問紙調査、既存データの利用、インタビュー、観察等の方法があり、収集したデータの種類や評価目的に応じて分析方法を選ぶ。成果の変化や施策・事業による効果を検討する場合には、可能な範囲で実施前後の比較や比較群との比較を行う。必要に応じて定量的な方法と定性的な方法を組み合わせ、成果の程度だけでなく、その内容や背景、実施上の課題等も含めて評価する。小規模事業や導入初期など、十分なデータを確保することが難しい場合には、事業の段階や取得可能なデータに応じて、現実的な評価方法を選択する。

2) 評価計画⁽¹⁾

- ・ 評価計画の一般的な手順を表 12 に示す。使える資源に応じて、具体化を図る必要がある。その際に、(i) 評価目的と実施上の条件（倫理・予算・期間等）を確認する。その上で、(ii) その条件の中で実施できる評価方法や評価研究デザイン（Part II 3 章参照）の候補を挙げる。つぎに、(iii) 必要なデータ（指標、収集方法、取得可能性、品質）を整理する。これにより、評価目的に照らして現実的で妥当な評価設計を選択できる。

表 12 評価計画の一般的な手順

手順	検討内容
1. 評価目的	・評価目的（施策・事業改善か説明責任か）の明確化
2. 評価対象	・評価対象（人・物・事象）の特定
3. 評価者の選定	・評価者（内部評価、または外部専門家への委託など）の検討
4. 評価時期の設定	・評価実施および公表時期の設定 ・変化・改善が見込まれるまでの期間の考慮
5. 評価研究デザイン	・データの収集時期・方法（定量・定性） ・分析の粒度（個票／制度単位／地域指標） ・比較対照群（同一自治体内の対照群との比較、または全国・他自治体との比較など）、指標および分析手法の選定など
6. データ収集・分析	・計画に基づくデータ収集および分析の実施
7. 評価結果の解釈と価値判断	・定量・定性データに基づく、結果の意義（実践的な意味）の解釈 ・関係者（現場、住民、自治体等）との確認・熟議による合意形成
8. 評価結果の報告・公開と活用	・報告内容：報告対象（現場、市民など）・時期（年度末、随時など）・媒体・分量の設計 ・評価結果に基づき、発注者・資金提供者（自治体等）や事業運営者が事業の方針を決定：①予算・支払判断（契約で事前合意した指標に基づきどの支払段階にあたるかを判定）の検討、②施策・事業改善・再設計の検討、③予算の確保や規模拡大（スケールアップ）の検討 ・活用先に応じた詳細設計：施策改善、プログラム選択・再設計、目的へのフィードバック、目的に合わせた分析の深度、関係者、意思決定の場（担当課、企画部門、首長・議会等）、成果物（改善提案、代替案比較、政策提言）公表範囲の設計

3) 中間評価の計画と実施

(1) 中間評価とは⁽²⁾

- ・中間評価では、主には、プロセス評価を行うが、中間年において、初期・中間アウトカムの評価を行う場合もある。
- ・「その事業がどのように運営されているのを記録し、意図した活動が計画通りに実行されているかを検証する」ものである。
- ・施策・事業がうまく運営できていない場合、原因が「目的」、「設計」、「実施」のいずれにあるかを特定し、見直しを図る必要がある。
- ・エビデンスに基づく実践プログラム（施策・事業）を構築する上で不可欠であるフィ

デリティ評価（コラム 19 フィデリティ評価）によるモニタリング手法が世界的に提案されている⁽¹⁾。

- ・ ロジックモデルに基づきプロセスに関わるインプットやアウトプットに関するデータを継続的に測定し、状況に応じて事業に修正を加える。
- ・ 一般的にはサービス利用の状況と実施組織の機能を評価する。

1. サービス利用の評価

施策・事業がターゲット集団に届いているか、その集団に偏りがいないかを評価する。

2. 組織機能の評価は、以下のような内容を考える。

- ▶ 実施計画が守られているか、そうでない場合の理由は何か、変更が必要か。
- ▶ 実施体制は適切か。ステークホルダーとの連携は十分か。

コラム 19

フィデリティ評価

フィデリティ評価とは、施策・事業（プログラム）が想定したやり方どおりに実施されているかを確認する指標である。事業の効果が出ないとき、原因が事業内容にあるのか、実施が計画どおりでなかったのかを区別できないと、改善につながらない。そこで、効果につながると考えられる実施の重要点（回数・時間・内容・対象者・担当者・手順など）を整理し、チェック項目（必要に応じて点数化）として、実施状況を継続的に確認する。詳細は「社会的インパクト・マネジメント・イニシアチブ『フィデリティ尺度』」（<https://simi.or.jp/info/7782>）⁽³⁾を参照。

(2) プロセス評価の計画で決めておくこと

- ・ KPIを確認する。例えば、「1年間での利用者数増加率」「継続参加率」「満足度80%以上」など、目的に応じた具体的かつ測定可能な指標を事前に設定することが望ましい。評価におけるKPIの活用により、進捗の可視化や目標達成度の定量的評価が可能になる。
- ・ 月報などの定期報告システムを構築し、計画と実績の乖離を早期に発見できる仕組みを作る。

コラム 20

施策・事業評価の目的

施策・事業評価の目的には、主に施策・事業の改善と説明責任がある⁽¹⁾⁽²⁾。

施策・事業の改善を目的とする評価は、施策・事業の円滑な実施と、より効果的な施策・事業の形成と実施に活用されるべき情報を提供する。より良い施策・事業にするための改善点を見つけ、実施中に軌道修正する。

説明責任を目的とする評価は、施策・事業に投入された資源（評価者投入財、人件費、費用、時間、情報など）を使って、望ましい成果（アウトカム）が実現したかを評価し、確認する。

コラム 21

施策・事業評価のアプローチ

施策・事業評価のアプローチには、形成評価と総括評価の2つがある⁽¹⁾⁽²⁾。形成評価は、施策・事業の改善に役立てることを目的とし、設計段階や継続的改善、あるいは戦略的選択のために行われる評価である。総括評価は、施策・事業の介入後に行うもので、成果や効率的な資源の活用がなされたかを評価の中心とする。説明責任を目的とした評価に役立つ。

コラム 22

事前・中間・事後評価と評価の5階層

評価実施時期によって事前評価・中間評価・事後評価に分けられる。一方、評価の目的・対象によって、ニーズ評価、セオリー評価、プロセス評価、アウトカム評価、効率性、公平・公正性評価の5つの階層に整理することもできる（表13）。事前評価では主にニーズ評価やセオリー評価を行い、中間評価ではプロセス評価やモニタリング（Part I 5章参照）と複数年度事業の中間年度における初期・中期アウトカム評価、効率性評価を実施することがある。事後評価は終了時や終了後も追跡したアウトカム評価や効率性、公平・公正性評価などが行われる。

表 13 評価の5階層と主な評価内容

評価の階層	主な評価内容
①ニーズ評価	施策・事業の実施により充足しようとしているニーズは何か
②セオリー評価	プログラムがどのように組み立てられているか
③プロセス評価	プログラムが意図されたとおりに実施されているか
④アウトカム評価	プログラムの成果が上がっているか
⑤効率性、公平・公正性評価	プログラムが効率的、公平・公正に実施されているか

（「プログラム評価ハンドブック：社会課題解決に向けた評価方法の基礎・応用」p32表2-2を一部改変）

4) 評価の計画と実施

(1) 評価研究デザインの決定

- ・ 中間・事後評価をするために、比較対照（前後比較・対照群）の設定（Part II - 1 章参照）や測定時期（1 時点または複数時点など）、評価基準（効果、効率性（**コラム 23 効率性評価**））、公平・公正性評価（**コラム 24 公平・公正性評価**）を決める評価研究デザインが必要である。
- ・ 評価研究デザインは、主に Part II 表 p58 に示したような手法がある。目的や条件に合うものを事前に決める。これらは専門知識を要するため、必要に応じて外部の専門家に相談する。
- ・ 施策・事業が目指す変化が生じているかを評価するには、導入前後の変化を捉えるアウトカム評価が必要になる。さらに“見かけ上の効果”をできるだけ取り除き、本当に施策・事業がもたらした効果だけを正しく評価することが望ましい（Part II 1 章参照）。そのため、プログラムのロジックに基づいてアウトカム指標を選択（**コラム 25 アウトカム指標の設定、コラム 26 アウトカム指標が満たすべき条件**）し、実施前後のデータ収集計画や参加群と非参加群とを比較する評価デザインが必要となる。
- ・ 施策・事業がうまく運営できていない場合、原因が「目的」、「設計」、「実施」のいずれにあるかを特定し、見直しを図る必要がある。

コラム 23

効率性評価

- ・ 効率性評価とは、プログラムのインプット（例えば、人件費・機器・運営費など）に対して、アウトプットやアウトカムがどの程度得られたか（費用と効果の関係）を、複数の施策・事業（または何も対策を打たない場合）との比較により評価することである。
- ・ 効率性評価を示す方法には、簡易な方法として、例えば、参加者 1 人あたりのコスト（総事業費÷参加者数）などもある。厳密な手法として、費用効果（Cost-Effectiveness：効果を金銭的価値に換算せず定量指標で評価する）や、費用便益（Cost-Benefit：効果を金銭的価値に換算して費用と比較する）などがある（**コラム 27 費用効果、コラム 28 費用便益**）。
- ・ 医療・福祉・教育のように、人への介入が多いプログラムでは、効果評価の実施後に評価することが多い。効率性評価の算定イメージとして、愛知県武豊町の介護予防事業（「通いの場」（サロン））および、よこはまウォーキングポイント事業の例を、**コラム 29 事例 愛知県武豊町／よこはまウォーキングポイント事業を用いた効率性評価**に示した。

コラム 24

公平・公正性評価

同じ施策・事業であっても、すべての人に同じように利用や効果がみられるとは限らず、対象者によって利用や効果に差が生じることがある。公平性評価では、対象者全体の施策・事業利用割合や成果だけでなく、年齢、性別、教育歴、所得などによって利用や成果が偏っていないかを捉え、「誰に効果があり、誰が効果を受けにくかったのか」を明らかにする。

コラム 25

アウトカム指標の設定

- ・アウトカム指標は、短期・中期・長期にわたる多段階での測定を行うことが望ましい。例えば、介護予防プログラムにおいては、短期（プログラム終了直後～3か月）では身体機能（握力、歩行速度、バランス能力）や社会参加の頻度、中期（6か月～1年後）では日常生活動作（ADL）の自立度、転倒発生率、外出頻度、主観的健康感、長期（2～3年後）では要介護認定率や介護度の進行状況などを測定する。
- ・施策・事業によって得られる変化（アウトカム）を設定し、それを客観的に評価するための測定手法を決める。
- ・最終的な目標達成に向けた進捗状況を確認するための時期別マイルストーンを設定し、その進捗を追跡するための方法を定める。

コラム 26

アウトカム指標が満たすべき条件

アウトカム指標が満たすべき条件は、信頼性と妥当性があり、アウトカムの変化を検出できる感度が求められる。「信頼性」とは、ほぼ同時期に繰り返し、あるいは別の主体が測定しても同じ結果が得られる「再現性」「安定性」のことである。「妥当性」とは、その指標が本当に測りたいものを正しく測れているかということである。妥当性には、内容的妥当性、基準関連妥当性、構成概念妥当性などがある。詳細は「予備的な分析と信頼性・妥当性」ウェブサイトを参照されたい⁽⁴⁾。指標における「感度」とは、測定する変化や差異があるときに、対応してその指標の値が変化や差を捉えられる程度のことである。また、データ収集の手間と費用も重要である。信頼性・妥当性が高くても手間と費用が高い指標は使えない。

信頼性と妥当性が検証されている例として、「GDS-15-J（日本版高齢者期うつ尺度短縮版）」がある。この指標は、高齢者のうつ症状を評価する指標で、信頼性と妥当性、感度が検証されている⁽⁵⁾。一方、「とても良いと思う割合」のような指標は、回答者ごとに判断基準が異なりやすく、同じ状態でも答えが変わる可能性がある。そのため、必ずしも安定して対象を測れるとは限らず、よい指標とは言い切れない。

コラム
27

費用効果 (Cost-Effectiveness)

費用効果とは、プログラムに要する費用に対して、プログラムによってどれだけの効果（アウトプットやアウトカム）が得られたかを示す。費用効果分析は、この関係を比較し、効率性を評価する方法である。費用は、投入した資金として金銭的に測定される。効果は、通常、プログラムの目的達成に関連した定量的な指標を用いる。

なお、効率性評価では、効果を金銭的価値に換算する費用便益（Cost-Benefit）と、金銭価値に換算せず評価する費用効果（Cost-Effectiveness）がある。費用効果分析では、例えば、介護予防プログラムによる健康状態などを効果指標として用いる。費用を効果で割った「費用対効果比」を算出し、「最も少ない費用で、最大の効果を得られるプログラムはどれか」を選択するために役立つ。

医療・保健分野では、効果指標として QALY（Quality-Adjusted Life Year：質調整生存年）が海外では広く用いられている。QALY は、生存年数に生活の質（QOL）を掛け合わせた指標であり、「完全な健康状態で 1 年生存すること」を 1QALY とする。例えば、QOL が 0.8 の状態で 5 年生存した場合は 4QALY（ 0.8×5 年）となる。QALY を用いることで、異なる疾患や介入方法を横断的に比較することが可能となる。

コラム
28

費用便益 (Cost-Benefit)

費用便益とは、プログラムに要する費用と、プログラムによってもたらされるアウトプットやアウトカムを金銭的価値で単位換算した便益との関係によって、効率性を評価する方法である。費用便益分析では、便益（金銭換算した効果）が費用を上回るかどうか、また費用に対して便益がどの程度得られるかを判断する。

費用便益の示し方は、①費用を便益で割った「便益あたりの費用（C/B）」、②便益を費用で割った「費用あたりの便益（B/C）」を用いる方法がある。一般に、B/C が 1 より大きい（または C/B が 1 より小さい）場合は、便益が費用を上回る。また、便益と費用の差（B-C）を算出して、費用便益差（純利益）と示すこともある。

1. 愛知県武豊町の介護予防事業（憩いのサロン）

武豊町では、2007 年より地域の「通いの場」（サロン）を中心に、住民の参加を基盤とした介護予防の取り組みを進めている。ここでは、事業の効率性を検討するための試算例として、要支援・要介護認定率の低下がサロン事業によるものと仮定し、介護サービス利用者の減少数と介護給付費の抑制額を概算する。武豊町の後期高齢者における要支援・要介護認定率が2015年から2020年までに5.8%ポイント低下した。武豊町の後期高齢者は約5,500人であったのでこの低下は約320人（ $5,500 \times 0.058$ ）に相当する。当時の全国統計（令和2年4月審査分）では、認定者数のうちサービス受給者の割合は約75%⁽⁷⁾であったことから、介護サービス利用者の減少は約240人（ 320×0.75 ）と推計される。厚生労働省「介護給付費等実態統計の概況（令和2年度）」⁽⁷⁾に示される令和2年4月審査分の受給者1人あたり費用額（172.6千円／月）を年換算すると、約207.1万円／年（ $172.6 \text{ 千円} \times 12 \text{ か月}$ ）となる。したがって、受給者が約240人減少したと仮定した場合、介護給付抑制額は、 $207.1 \text{ 万円} \div \text{年} \times 240 \text{ 人} = \text{約} 4.97 \text{ 億円} \div \text{年} \approx \text{約} 5.0 \text{ 億円} \div \text{年}$ と概算される。一方、事業の実施に要した費用として、8か所のサロンの運営委託費は合計約270万円であり、運営を支援した町の職員と研究者の人件費で数百万円が見込まれた⁽⁸⁾。以上の仮定に基づくと、介護給付費の抑制額は、把握できた範囲の事業費を上回り、費用に対して大きな成果が得られた可能性が示される。

2. よこはまウォーキングポイント事業（YWP）

「よこはまウォーキングポイント（YWP）」は、市民が歩数計やスマホで歩数を記録し、ポイント特典で歩行を促す健康増進プログラムである⁽⁶⁾。横浜市在住の要介護認定を受けていない65歳以上を対象に、横浜市とJAGES（日本老年学的評価研究）が共同で行った調査では、YWPに参加した場合、参加していない場合に比べて、1日あたりの歩行時間が約3.6分増加し、身体機能低下と抑うつが抑制されることが示された⁽⁹⁾。

この歩行時間の増加（約3.6分）は、歩数に換算すると1日あたり約360歩の増加に相当する。国土交通省の報告⁽¹⁰⁾では、「1日1,500歩増えると1人あたり年間約3万5,000円の医療費が抑制に相当する」とされており、これを1歩あたりの医療費抑制額に換算すると年間約23.3円（ $35,000 \text{ 円} \div 1,500 \text{ 歩}$ ）となる。YWPによる「1日360歩の増加」にあてはめると、参加者1人あたり約8,400円（ $23.3 \text{ 円} \times 360 \text{ 歩}$ ）の抑制効果が期待される。これをYWP参加高齢者約15万人に適用すると、全体で年間 $8,400 \text{ 円} \times 150,000 \text{ 人} = \text{約} 12.6 \text{ 億円}$ の医療費抑制効果が見込まれる。

一方、YWPの年間予算額⁽¹¹⁾は3億7,900万円である。ただし、公表されているデータには高齢者のみの予算額が示されておらず、この予算額は高齢者以外の世代も含む全体予算である。この全体予算を用いて費用便益比を算出すると、 $12.6 \div 3.79 = \text{約} 3.32$ 、費用便益差は、 $12.6 \text{ 億円} - 3.79 \text{ 億円} = \text{約} 8.81 \text{ 億円}$ と推計される。

(2) データ収集と分析の計画⁽¹⁾⁽²⁾

- ・評価レベルに応じた分析単位の設定
 - ▶ 評価目的に応じ、個人レベルか、地域・自治体レベルかを明確化し、個人の変化を追うのか、地域・自治体単位の集計データを用いて地域と比較するのかといった分析の粒度を決定する。
- ・収集方法と必要サンプル数の設計
 - ▶ データの種類（アンケート、インタビュー、健診記録、既存の行政データ等）および収集手法（郵送、オンライン、対面等）を選定する。
 - ▶ 必要に応じて、過去事例や文献を参考に、効果検出に必要なサンプル数を算出する。
- ・対象の異質性を踏まえた項目設定（コラム 30 無作為化比較試験、コラム 31 異質性の考慮）
 - ▶ 人によって効果が異なること（異質性）を考慮する。
 - ▶ 成果指標に加え、基本属性（性、年代、居住地）、社会経済状況（所得、教育、就業）、健康状態（既往歴、現在の健康度）等も収集項目として含めて設計する。
- ・分析手法の選定と指標の考案
 - ▶ 量的なデータの分析手法には、大別してデータの特徴を数値で示す「記述統計」と、統計学的な手法を用いて変数間の関連や値を推定するなどの「分析・推計統計」がある。どのようなデータを用いて、どのような分析をするか、分析の方向性を見据えて、指標の考案・選定をする。
- ・法遵守とデータ管理体制の整備
 - ▶ 個人情報の保護に関する法律第 69 条等を踏まえ、適切に、個人情報の利用及び提供の制限の手続きを行う。
 - ▶ 安全なデータ管理・共有システムを構築する。

コラム 30

無作為化比較試験（Randomized Controlled Trial: RCT）

実験的な研究デザインである RCT は因果推論の精度は高いが、行政の現場では適用が困難な場合が多い。また RCT は条件を厳密にそろえて検証するのに適している一方で、参加対象を絞り込む傾向があるため、対象から除外される複雑なあるいは例外的な属性をもつ対象者（例えば 90 歳以上、認知症や要介護度が重症など）を含む多様な地域や人々を対象とした取り組みには適さない場合がある。このような場合は、実施可能な範囲で妥当性を確保できる他の手法（例えば、傾向スコアマッチングなど）を選択することが重要である。

※参考資料：様々な評価デザインと分析手法、p●参照

同じ介護予防プログラムに参加した場合でも、人によって効果の大きさが違うことがある。これを異質性と言う。例えば、同じ運動教室でも、もともと活動的な人よりも、これまであまり運動してこなかった人の方が、効果が大きく出ることがある。また、ゲーム依存症のように、子どもや若者には悪影響がある一方で、高齢者では少ないなど、年齢などによって影響が変わることもある。このように、「誰にとって」「どのような条件の人にとって」効果が出やすいのか／出にくいのかを明らかにすることは重要である。効果が十分に得られていない層に対して、適切な介入が行えるよう、設計段階から評価フレームに異質性の分析（例えば、年齢、性、経済状況の検討など）を組み込む必要がある。

(3) 評価結果の解釈と活用

- ・ 結果の解釈では、統計的有意性（その差が偶然や誤差による可能性が低いと判断されるかどうかの指標）だけでなく、施策・事業がアウトカムに与えた影響の大きさも踏まえて、実践的意義を考慮して評価することが重要である。なお、この「影響の大きさ」を専門的には効果量（Effect Size）と呼ぶ。また、観察された変化のうち、住民・患者が「意味がある改善／悪化」と感じる最小の差は MCID（Minimum Clinically Important Difference：臨床的に意義がある最小変化量）と呼ばれる。統計的有意性は参加者のサンプル数に影響を受けやすく、サンプル数が大きいとわずかな差でも統計的に有意となる。一方、サンプル数が少ない場合は統計的に有意でなくても、実践上は意味のある差である可能性がある。そのため、結果を示す際には、統計的有意性だけでなく、効果量や MCID も踏まえて解釈する。

(4) 第三者評価機関への委託

- ・ 評価研究デザインの選定や必要サンプル数の計算、評価計画の策定、結果の解釈等には相当の学識経験を必要とするので、第三者評価機関への委託を検討も考慮する。委託する際は、計画段階で仕様書に記載する内容（例：分析内容に関する取り決めと予算など）も決定する。

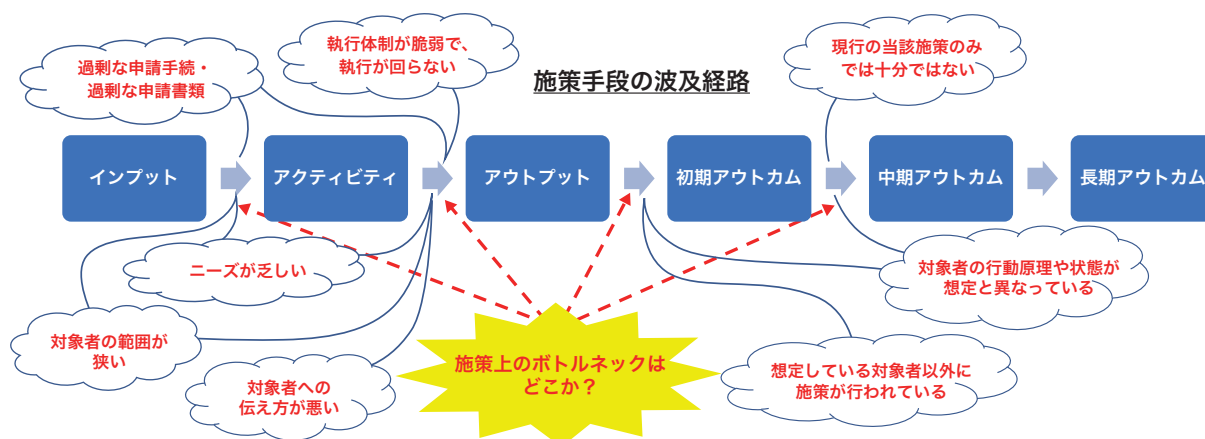
【参考文献】

- (1) 源 由理子, 大島 巖, 山谷 清志. プログラム評価ハンドブック: 社会課題解決に向けた評価方法の基礎・応用. 晃洋書房; 2020.
- (2) Rossi PH, Lipsey MW, Freeman HE, (監訳) 大島 巖, 平岡 公, 森 俊, et al. プログラム評価の理論と方法: システマティックな対人サービス・政策評価の実践ガイド. 日本評論社; 2005.
- (3) 社会的インパクト・マネジメント・イニシアチブ. フィデリティ尺度. Available from : <https://simi.or.jp/info/7782>
- (4) Bunji K. 4 予備的な分析と信頼性・妥当性 [Internet]. 神戸大学大学院経営学研究科 [2022 Jun 7] Available from : https://www2.kobe-u.ac.jp/~bunji/files/lecture/MVA/html/chapters/04_prealysis.html# 信頼性と妥当性の考え方
- (5) Sugishita K, Sugishita M, Hemmi I, Asada T, Tanigawa T. A Validity and Reliability Study of the Japanese Version of the Geriatric Depression Scale 15 (GDS-15-J). Clin Gerontol ; 2017;40(4):233-40.
- (6) 平井 寛, 竹田 徳則, 近藤 克則. まちづくりによる介護予防: 「武豊プロジェクト」の戦略から効果評価まで. ミネルヴァ書房; 2024.
- (7) 厚生労働省. 令和2年度 介護給付費等実態統計の概況 2020. Available from : <https://www.mhlw.go.jp/toukei/saikin/hw/kaigo/kyufu/20/index.html>.
- (8) 近藤克則. 保健・医療・介護における効果・質・格差の評価—到達点と課題—. フィナンシャル・レビュー. 2015; 123: 133-157.
- (9) 藤原 聡子, 辻 大士, 近藤 克則. ウォーキングによる健康ポイント事業が高齢者の歩行時間, 運動機能, うつに及ぼす効果: 傾向スコアを用いた逆確率重み付け法による検証. 日本公衆衛生雑誌. 2020;67(10):734-44.
- (10) 国土交通省. 『医療費抑制効果の見える化 (原単位の試算)』. 2017. Available from : https://www.mlit.go.jp/report/press/toshi07_hh_000107.html
- (11) 横浜市. 『平成 28 年度予算案について — 横浜の未来の鍵を握る年 —』 2016. Available from : https://www.city.yokohama.lg.jp/city-info/zaisei/jokyo/yosan/h28/28yosan.files/0076_20180713.pdf.

5章 施策・事業の実施とデータ収集

1) 施策・事業の実施とプロセス・アウトカムの測定⁽¹⁾⁽²⁾

- ・計画書に基づき、施策・事業を実施する。社会状況の改善を効率的・効果的に得るためには、計画どおりに施策が実施されることが望ましいが、実際の運用では様々な要因により想定どおりに進まないことがある（例：関係者の合意形成の遅れ、予算の変更、人事異動、感染症流行等による施策環境の変化など）。
- ・実施にあたっては多くのステークホルダーとの調整が必要であり、その過程で計画の修正や妥協を余儀なくされる場合がある。その結果、当初意図した施策の内容と実施された内容との間に乖離が生じることがある。
- ・そのため、事前の計画に基づき実施のプロセスを測定・モニタリングし、どこまでが想定どおりに実施されているのか、どの段階で想定との乖離が生じているのかを把握し、施策上のボトルネックを特定する。より効率的・効果的な施策とするため、必要に応じてステークホルダーと協議し、施策内容の修正を行う⁽¹⁾⁽²⁾。
- ・また、施策の最終的な目標は社会状況の改善にあるため、ロジックモデルに基づき、適切なアウトカム指標を用いて成果を測定する。このとき、短期・中期・長期等の多段階に分けて、測定（モニタリング）を行う。



(内閣官房行政改革推進本部事務局.EBPM ガイドブック Ver.1.2より引用、一部改変)

図 14 施策実施におけるボトルネックの具体例

コラム
32

プロセス評価とモニタリング

プロセス評価は事業活動がどのように実施されたかという実施過程（プロセス）の妥当性を検証するものであり、事後的に実施される場合もある。インプットからアクティビティ、アウトプットに至る実施過程に関するデータを分析し、施策がどのように実施されたか、またその結果がアウトカムにつながる構造は適切であったかを検討することで、より効果的な施策の形成・改善に役立てる。

一方、モニタリングはプロセス評価と重なる部分もあるが、主として計画どおりに施策が実施されているかを継続的に把握する進捗管理の意味合いが強い⁽²⁾。

プロセス評価：『なぜうまくいっているのか／いないのか』を分析する

モニタリング：『計画どおりに実施されているか』を確認する

【参考文献】

- (1) 内閣官房行政改革推進本部事務局. EBPM ガイドブック Ver.1.2 2023.https://www.gyokaku.go.jp/ebpm/img/guidebook1.2_230403.pdf
- (2) Rossi PH, Lipsey MW, Freeman HE, (監訳) 大島 巖, 平岡 公, 森 俊, et al. プログラム評価の理論と方法：システムティックな対人サービス・政策評価の実践ガイド. 日本評論社; 2005.

6章 社会的インパクト評価

1) 効果の検証

- ・ プロセスおよびアウトカム指標の測定のみでは、観察された変化が施策・事業により生じたものなのか、それとも他の要因による影響なのかを明確に区別することができない。そのため、可能な限り見かけ上の効果を取り除いた効果（アウトカム、社会的インパクト）の評価を実施することが望ましい⁽¹⁾⁽²⁾⁽³⁾。
- ・ 社会的インパクト評価を行う際には、事業の実施プロセスや利用状況（参加者数や参加者の構成）に関する情報に加え、関連する人口動態や経済状況などの外部環境の変化についての情報も収集し、それらを踏まえて効果を検証する必要がある。
- ・ 評価は、事前に策定した計画に基づき、可能な限り厳密な方法で実施し、これにより施策・事業が標的集団や社会にどのような効果・影響を与えているのかを明らかにする。また、ロジックモデルに基づいた検証を行い、施策・事業の因果関係に齟齬がないかを確認する。
- ・ 評価研究デザインとしては、実験的デザインや準実験的デザインが用いられる（Part II 3章 様々な評価研究デザイン 参照）。これらの方法では、参加（介入）群と非参加（非介入 / 対照）群との前後比較を通じて、施策・事業の効果をより適切に推定することを目指す。ただし、使用した手法の限界を理解した上で結果を解釈することが重要である。

2) 社会的インパクト評価の結果に基づく判断

- ・ 収集された情報（社会的インパクト評価の結果）に基づき、施策・事業をそのまま継続するのか、修正すべきか、あるいは統合・廃止するのかについて判断を行う⁽¹⁾⁽²⁾。
- ・ PFS 事業においては、この評価結果が、報酬の支払額の決定や事業継続の是非の判断、契約条件の見直しなどに活用される。
- ・ 評価結果が必ずしも期待どおりでない場合でも、それを単なる失敗と捉えるのではなく、学習の機会として位置づけることが重要である。評価結果を踏まえて実施方法を見直し、次の改善や再設計につなげていくことが重要である。

3) データの公表

- ・ 社会的インパクト評価の内容が整理された後は、事前の計画に基づき報告書を作成し、公表する。例えば、①次年度の事業計画立案に必要な速報値、②年度末までの事業報告書（確定値）、③市民等に向けたプレスリリース資料、などが考えられる。
- ・ これらのエビデンスは、意思決定に関わるステークホルダーと共有することで、特に新たな事業計画の立案や予算請求の場面において有効に活用される。

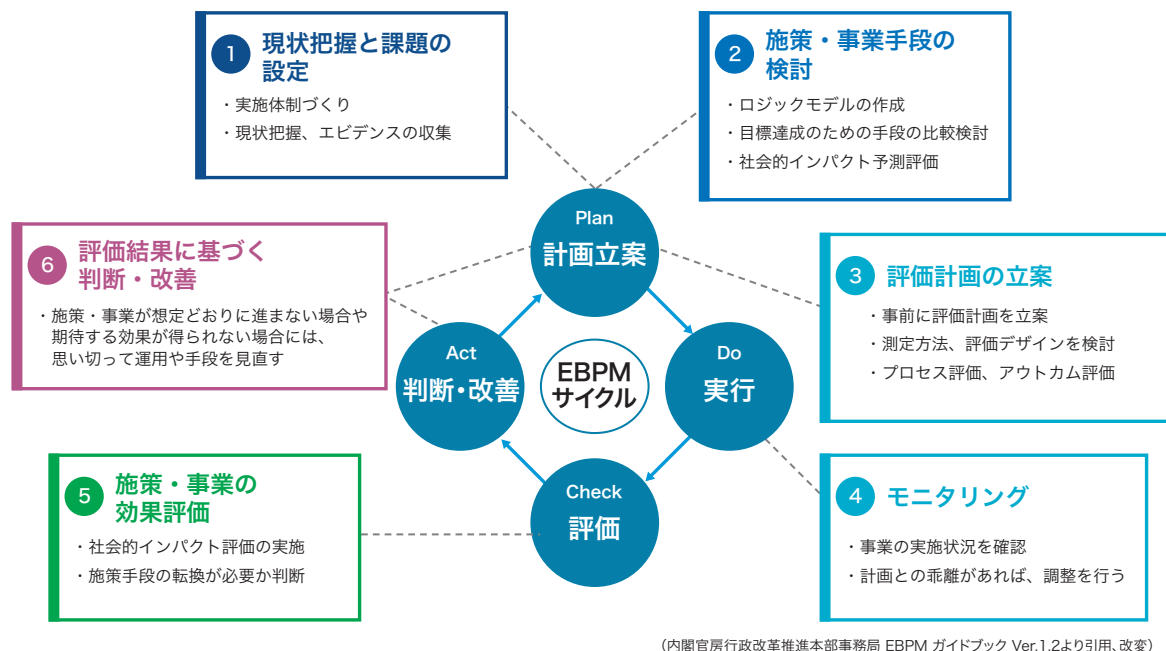


図15 EBPMのマネジメントサイクル

【参考文献】

- (1) 内閣官房行政改革推進本部事務局. EBPM ガイドブック Ver.1.2 2023.
Available from : https://www.gyokaku.go.jp/ebpm/img/guidebook1.2_230403.pdf
- (2) 源 由理子, 大島 巖, 山谷 清志. プログラム評価ハンドブック: 社会課題解決に向けた評価方法の基礎・応用: 晃洋書房; 2020.
- (3) Rossi PH, Lipsey MW, Freeman HE, (監訳) 大島 巖, 平岡 公, 森 俊, et al. プログラム評価の理論と方法: システマティックな対人サービス・政策評価の実践ガイド. 日本評論社; 2005.

Part II

社会的インパクト評価 の研究デザイン

Part I では、社会的インパクト評価の目的や手順などについて紹介した。Part II では、社会的インパクト評価に必要となる 3 条件—1) 比較対照群の設定、2) 前後比較で変化を捉える、3) 背景要因を揃えて比較（図 16）—について、その必要性を説明する。さらに、評価事例や評価に要する期間、これらの条件を満たすために開発されてきた評価研究デザイン、評価研究デザインを選択する際の考え方について紹介する。

これらの条件を満たすための評価研究デザインの設計やデータ収集、データ分析には、専門的な知識や経験が必要となる。そのため、本書では、こうした作業を学識経験者や評価機関等と連携して進めることを想定している。行政職員が専門家に評価を依頼したり、関係者との調整を行ったりする際に役立つよう、評価に関する基本的な考え方や知識、専門用語を、できるだけわかりやすく解説する。より詳しい内容は、必要に応じて参照できるよう、参考資料にまとめている。

1 章 「見かけ上でない施策・事業効果」の評価に必要となる 3 条件

本文書では、施策や事業の実施前後で見られる対象における変化を「アウトカム」と呼ぶ。ただし、その変化のすべてが必ずしも施策・事業の効果によるものとは限らない。

例えば、参加者の健康状態が改善したとしても、その要因が「事業への参加」によるものか、あるいは「季節や社会環境の変化など、他の要因」によるものかを区別する必要がある。このように、見かけ上の効果を含んだ観察された変化が「アウトカム」である。効果を評価するためには、このアウトカムから、見かけ上の効果を差し引く必要がある。

社会的インパクト評価では、この「観察された変化」から、実際には施策・事業によらない見かけ上の効果（専門的には「逆の因果」や「交絡」などと呼ばれる）をできるだけ取り除き、施策・事業がもたらした効果だけを正しく評価することを目指す。

しかし、この見かけ上の効果を差しひくことは容易ではない。そこで本書の Part II では、見かけ上でない施策・事業の効果評価に必要な 3 つの条件、

- 1) 比較するグループ（対照群）を設けること
 - 2) 事業の前後で変化を比較すること
 - 3) 背景となる条件（年齢・性別・所得など）を揃えて比較すること
- について、具体的に説明する。

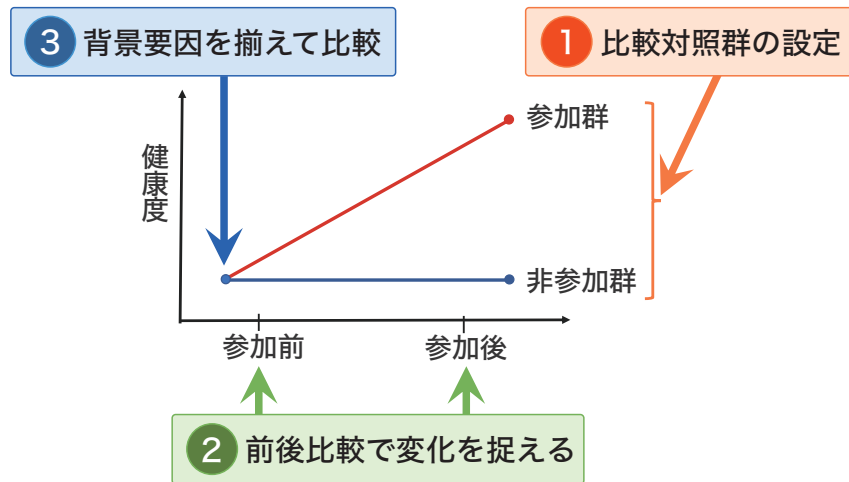


図16 見かけ上でない施策・事業の効果の評価に必要な3条件

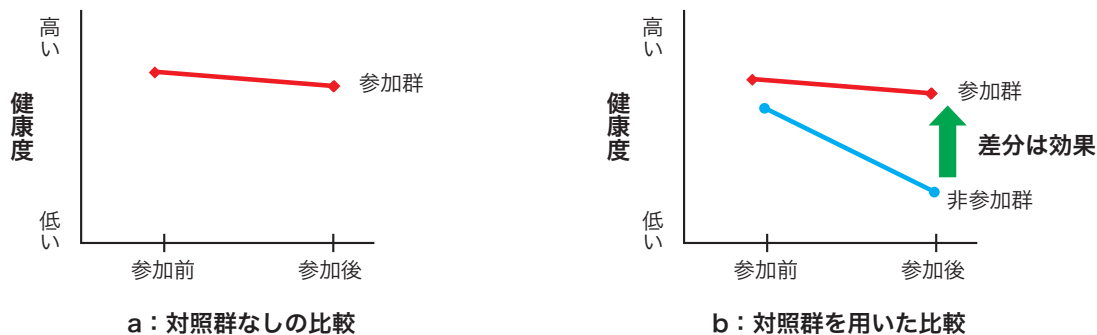
1) 比較対照群の設定

施策・事業への参加者のみに注目すると、真の効果を見誤る恐れがある。例えば、健康教室の実施後に、参加群の健康度が低下した場合を考えてみよう。

図17_aのように、参加群の変化（アウトカム）だけに注目すると、健康度が低下しているため、一見「効果がなかった」、あるいは「むしろ悪化した」と評価してしまう。しかし、この健康教室が新型コロナウイルス感染症流行下（以下、コロナ禍）などに実施された場合、その結果の捉え方が大きく変わる場合がある。

コロナ禍では、感染への不安や社会的孤立などによって健康度が悪化したことが報告されている⁽¹⁾。このような時期に、もし非参加群が参加群よりも健康度が大きく低下していた場合、その健康教室には「健康度の悪化を最小限に食い止めた」という効果があったことになる（図17_b）。

したがって、参加群における変化（アウトカム）を追うだけでは、施策・事業の効果は正しく評価できない。効果を評価するためには、非参加群（比較対照群）を設定し、2群間で比較することが不可欠である（図17_a、b）。



a：対照群なしの比較
教室前より後で健康度が低下し、一見すると「参加には効果がなかった」ように見える。

b：対照群を用いた比較
参加群は現状維持または健康度の低下が少ない一方で、非参加群は悪化。参加が健康度の悪化を予防する効果があったことがわかる。

図17 参加群のみの前後比較と比較対照群を用いた事業効果の評価

2) 前後比較で変化を捉える

ある1時点のデータ（横断データと呼ばれる）だけでは、たとえば「教室の参加者は健康度が高い」という結果が出ても、それが「教室への参加（原因）によって健康になった（結果）」とは限らない。

例えば、施策・事業後の調査で「健康教室の参加群（**コラム 33 参加群・非参加群の別名と対照群と対象群の違い**）は、全国平均（や非参加群）より健康指標が良い」という結果が見られたとしても、「健康教室参加（A）が健康（B）をもたらした」とは限らない。もともと健康度が高い（B）人ほど教室に参加しやすい（A）という「逆の因果」（**コラム 34 「関連」と「因果」**）の可能性があるからである（図 18_c）。

この「逆の因果」の可能性を取り除くには、同じ人を追跡し、参加前後の変化をみる必要がある。一方で、参加群だけの前後の変化をみても、その変化が施策・事業によるものなのか、社会全体の状況の変化（例：コロナ禍）によるものなのかを区別することはできない。

そのため、参加群と非参加群の両方について「参加前後の変化」を比較することが重要である。具体的には、両群が同じ傾きで変化している（施策・事業以外の環境などの影響を受けた）のか、あるいは参加後に差が生じたのかを確認する（図 18_d）。このように、同じ人を複数時点で調査する縦断研究デザインを用いることで、見かけ上の効果を除いた評価が可能になる。

コラム 33

参加群・非参加群の別名と対照群と対象群の違い

同じ集団でも、文書によって別々の呼び方で呼ばれることがある。混乱を避けるため、代表的な言い換えを整理しておく。

- ・参加群（施策・事業に参加した人たち）… 利用群、介入群、処置群、曝露群 と呼ばれる
- ・非参加群（施策・事業に参加しなかった人たち）… 非利用群、非介入群、非曝露群、対照群（コントロール群） と呼ばれる

呼び名が違うだけで、指しているものは同じである。

注意が必要なのが「対象群」と「対照群」である。読みはどちらも「たいしょうぐん」だが、意味は異なる。

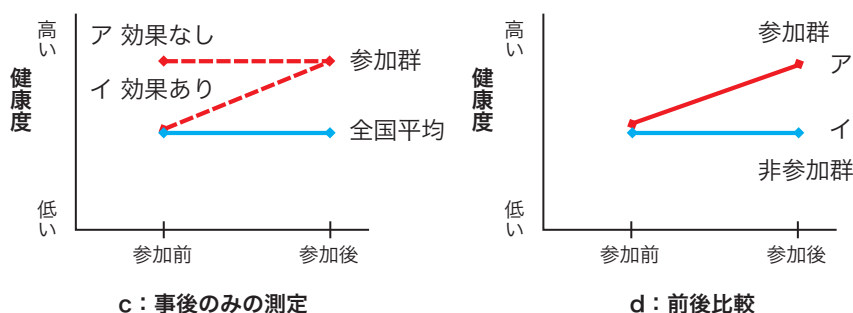
- ・対象群… 評価研究の対象となった人の全体（＝参加群＋非参加群）
- ・対照群… 比較のために置く非参加群

コラム
34

「関連」と「因果」

評価では、事業の実施状況や健康状態といった調査項目を「変数」と呼ぶ。2つの変数の間で、一方が変化すると他方も連動して変化する状態を「関連がある」と言う。

ただし、関連があるからといって、必ずしもAが原因でBが結果（ $A \rightarrow B$ ）であるとは限らない。Bが原因でAが結果（ $B \rightarrow A$ ）という「逆の因果」や、「第3の要因」（コラム33 交絡）によって見かけ上の関連が生じる場合もあるため注意が必要である。



参加後の一時点のデータが全国平均より高くても効果ありとは限らない。アのようにもともと健康度の高い人たちが参加していて変化なしなら効果なし。イのように改善していれば効果あり。事後評価だけでは効果の有無は不明である。

前後比較で、非参加群では（イ）のように変化がないのに対して、参加群でアのような改善が見られた場合に、はじめて効果である可能性が出てくる。

図18 事後のみの測定と前後比較による事業効果の評価

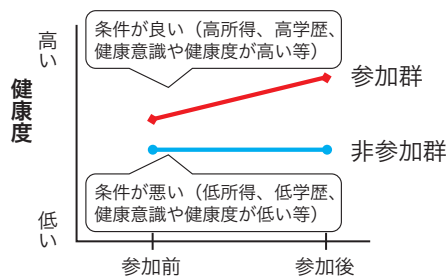
3) 背景要因を揃えて比較

「前後比較」で変化を捉え、非参加群と比較するだけでは、「見かけ上の効果」を完全に取り除くことはできない。参加群と非参加群の背景要因（コラム35 交絡）の違いによる見かけ上の違いの可能性が残っているためである。

図19_eのように、背景要因（特性、属性などとも呼ばれる）が異なる群間で比較すると、それが健康教室による効果なのか、あるいは対象者の背景要因の違いによる見かけ上の違いなのかを区別できない。そのため、見かけ上の効果を取り除いて評価するには、図19_fアのように所得や教育歴、健康度などの背景要因を揃えて比較することが重要である。

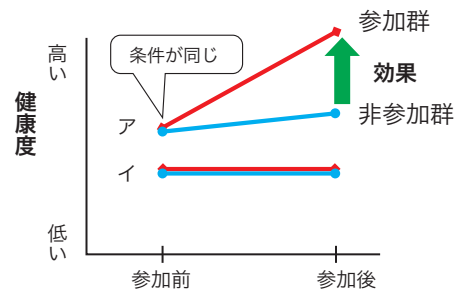
また、個人の背景要因によっては、同じ参加群の中でも効果の現れ方が異なる場合がある。例えば、元々の健康度や教育歴が高い人では教室の前後で改善が見られる一方（図19_fア）、低い人では変化が見られない場合がある（図19_fイ）。

このように背景要因別にグループ分けをして分析（層別分析）することで、どのような人に効果があるのかを詳細に把握できる。より精度の高い評価を行うためにも、背景要因に関する情報は可能な限り収集しておくことが望ましい。



e：背景要因が異なる人との比較

参加群にはもともと高所得、高学歴、健康意識も、参加前の健康度も高いなど背景要因の条件が良い人が多いことがよくある。参加群が非参加群かの違いによる「真の効果」でなく、もともとの背景要因の違いによってもたらされる「見かけ上の『効果』」のことがある。



f：背景要因が同じ人との比較

背景要因を揃えて比較したうえで、アのように非参加群より参加群の健康度が高くなっていれば「真の効果」と考えられる。一方、イのように、もともと健康度が低いなど特定の背景要因を持つグループでは効果が見られないことがある。このように、背景要因によって効果の現れ方が異なることを「異質性」という。

図19 背景要因の違いを考慮した事業効果の評価

コラム
35

交絡

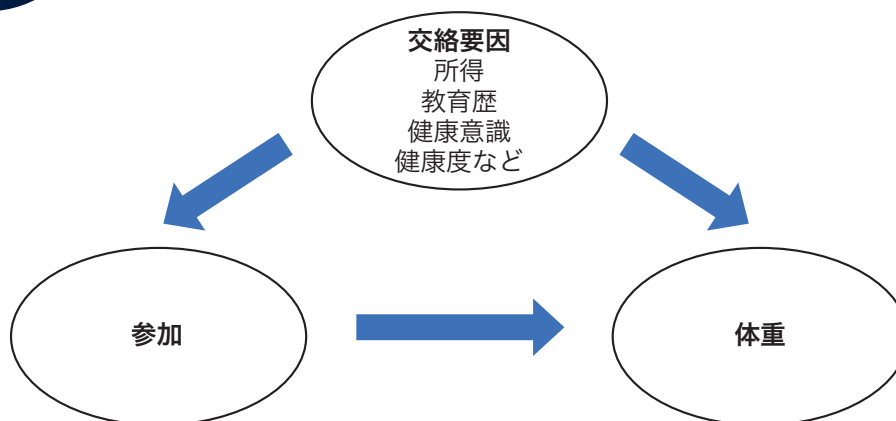


図20 交絡要因による見かけ上の効果

「交絡（こうらく）」とは、施策・事業以外の第三の要因が結果に影響を与えてしまうことを指す。この交絡を見落とすと、実際には効果がなかったのに「効果があったように見える」場合や、逆に「効果がない」と誤って判断してしまうことがある。

例えば、「健康教室に参加した人は体重が2.6kg減り、参加していない人は1.4kg減った」という結果が出たとする（次頁表①）。この数字だけを見ると、「健康教室には1.2kg分の体重減少効果があった」と考えてしまいそうである。しかし、もしも健康教室の参加者には「もともと健康意識が高い人」が多く、参加していない人には「健康にあまり関心がない人」が多かったとしたらどうだろうか（次頁表②）。この場合、体重の減少は「健康教室の効果」ではなく、「健康意識の違い」によるものかもしれない。このような、結果に影響を与えてしまう第3の要因を「交絡要因（または交絡因子）」と呼ぶ。

交絡は専門家でもデータがなければ排除することが難しく、知られていない交絡要因が隠れている場合もある。そのため、研究報告では「効果があった」と断定せずに、「効果がある可能性がある」「示唆された」「かもしれない」といった表現が使われる。

表① 健康意識の違いを考慮せずに比較した場合

グループ	体重変化
参加群 (100 人)	-2.6 kg
非参加群 (100 人)	-1.4 kg
2 群間の差	-1.2kg

表② 健康意識が同じ層で比較した場合

健康意識	参加群		非参加群		真の効果
	人数	体重変化	人数	体重変化	
健康意識が高い層	80	-3.0kg	20	-3.0kg	0.0kg
健康意識が低い層	20	-1.0kg	80	-1.0kg	0.0kg
計	100	-2.6kg	100	-1.4kg	0.0kg

コラム
36

効果分析に必要なデータの内容

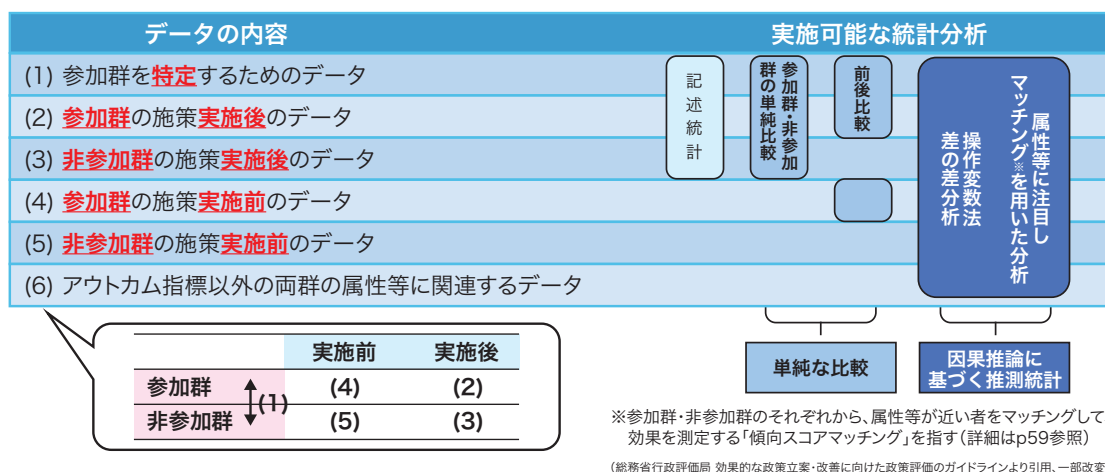


図21 必要なデータと実施可能な統計分析

上図⁽²⁾のように、(1)+(2)+(3)で「記述統計」および「参加群・非参加群の施策実施後のみ（一時点）データの単純比較」が可能となり、(1)+(2)+(4)で「前後比較」が可能となる。さらに(1)～(5)で参加群と非参加群との前後比較（参考資料「差の差分分析」など p68 参照）が可能となり、(6)の参加群・非参加群の背景要因（関連するデータ）が得られることで背景要因を揃えた比較分析（参考資料「傾向スコア（プロペンシティスコア）マッチング」など p65 参照）が可能となる。まずは「記述統計」や「対照群との比較」、「前後比較」を行うが、専門家の助言を受けながら背景要因も含めたデータ収集を計画し、真の介入効果をより適切に評価できるデザインを目指す必要がある。

コラム 37

対照群を設定できない場合の評価

「対照群を設定できない」「全員に施策・事業を適用せざるを得ない」ケースが多くある。そのような場合の代替手法には、回帰不連続デザイン、操作変数法、合成コントロール法などがあり、対照群がない場合でも可能な限り因果推論に近づける工夫が必要である。詳細は＜参考資料＞様々な評価デザインと分析手法を参照。

【参考文献】

- (1) Zeigler Z. COVID-19 Self-quarantine and Weight Gain Risk Factors in Adults. *Curr Obes Rep.* 2021;10(3):423-33
- (2) 総務省 行政評価局．効果的な政策立案・改善に向けた政策評価のガイドライン ～具体的な事例を中心として～．2025. Available from : https://www.soumu.go.jp/main_content/000935597.pdf

2章 評価事例と評価に要する期間

これらの3条件を満たした自治体における社会的インパクト評価の架空の事例をコラム 38 事例：A市でのPFSによる介護予防事業の社会的インパクト評価に示した。これまでの報告から、施策・事業の実施準備には、1年数ヶ月から2年程度かかることが多い。

関係者が全体像を理解し、ロジックモデルを作成し、予算規模に目途を立てるのに数ヶ月から1年程度、9月の次年度向け概算要求に間に合わせたとして、次年度の春から施策・事業導入に向け業者の公募手続きなどが始まり、夏～秋頃から実際の施策・事業導入に至る。ここまでの1年数ヶ月～2年かかる。1年～5年間など施策・事業実施期間の長さによって、中間評価の時期は数ヶ月～2・3年後などとなる。

コラム 38

事例：A市でのPFSによる介護予防事業の社会的インパクト評価

1. 背景：

- ・A市は要支援・要介護認定率が県平均・全国平均と比べて高く、とくに軽度層（要支援）へのアプローチが課題とされていた。
- ・従来型の介護予防事業では、無関心層を取り込めていないという課題があった。民間企業のノウハウを活用し、改善したいという意図があった。
- ・A市の財政部門側から社会保障費抑制の観点での関心もあり、PFSによって、成果に応じて対価を支払う形を導入することが検討された。

2. スケジュール例：

A市のPFS事業例では、下図のように準備、サービス提供、最終評価・支払まで含めて約6.5年間のスケジュールで実施された。

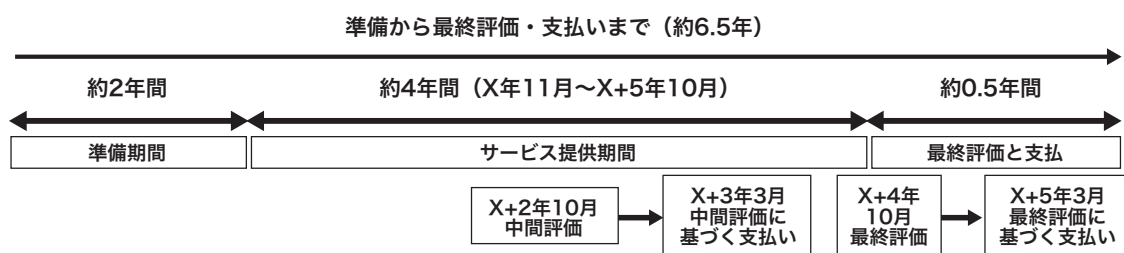
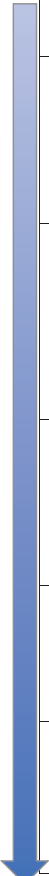


図22 A市のPFS事業における実施・評価・支払いのスケジュール

2.1. 準備期間：体制づくり（約2年）：

- ・準備期間では、次表のような流れで準備・体制づくりが行われた。



段階	実施内容	担当主体・関係者	備考・ポイント
発案・内部検討	庁内にて社会保障費抑制・PFS導入の可能性検討	財務部門・部局、企画部門・部局、福祉部門・部局	まずは市役所内部でPFSの概念理解を共有。次年度予算編成段階でPFS事業検討予算（調査費）を計上。
庁内勉強会・啓発	金融機関や専門家を招いた職員向け勉強会	企画部門・部局、関係部署	担当課以外の部署職員にも理解を広げるため、勉強会により職員間共通理解を醸成。
外部助言ネットワーク構築	PFS有識機関や評価機関との関係構築	有識者、大学・評価機関	成果指標設計・評価手法の助言を得る。A市はPFS/SIB支援実績のある団体の紹介で有識者助言を受けた。
サウンディング調査	民間事業者へのヒアリング	企画部門・部局、外部支援者、事業者候補	民間ノウハウやアイデアを探るとともに、使用要件のすり合わせを行う。
仕様書・提案書設計	成果指標設定、報酬・支払条件案、契約枠組み検討	企画部門・部局、福祉部門・部局、外部助言者、評価機関	提案記載事項を公募前に固める。応募企業を比較しやすいよう設計。
予算化、合意形成	予算要求、庁内質疑対応、議会説明	財務部門・部局、企画部門・部局、関連部署、議会对応	外部助言機関の存在、実施リスクや機体収益性を説明して承認を得た。

このように、事業化検討から実施までに、内部合意づくり→外部関係者巻き込み→仕様整備→予算化というステップを踏み進めた。事業年度に業者公募・選定・契約を経て事業開始となった。

2.2. サービス提供、評価、支払い（約4.5年）

・スケジュール

サービス提供期間：X年11月～X+4年10月

中間評価：X+2年10月

最終評価：X+4年10月

支払い：4年間の事業の2年目の評価に基づく支払い→X+3年3月、最終評価に基づく支払い→令和X+5年3月

・この期間の流れと関係者役割

段階	実施内容	担当主体・関係者	備考・ポイント
公募・事業者選定	公募型プロポーザル実施、応募受付、審査、選定	企画部門・部局、福祉部門・部局、有識者審査会	公募仕様書の明瞭性、中間成果と最終成果の関係性。応募者数確保。
契約・開始準備	契約締結、オリエンテーション、参加者募集、評価手法を織り込んだ事業スケジュール	福祉部門・部局、受託企業、評価機関	個人情報保護、同意取得、事業運営ルール整備、実施マニュアル整備。
プログラム運営	通いの場、交流イベント、行動変容支援メニュー実施	受託企業、地域サポーター、協働団体	参加者募集、メニューバリエーション、脱落防止戦略など。
モニタリングと中間レビュー	参加者数、継続率、行動変容指標、ドロップアウト分析、改善検討	福祉部門・部局、受託企業、評価機関	5年間の事業の3年目の評価前に進捗チェック、必要な軌道修正をかける。
中間評価・支払（部分的）	評価機関が中間成果を検証→条件達成分の支払実施	評価機関、福祉部門・部局	評価手法の透明性、仮説（ロジックモデル）の妥当性、結果説明責任。
継続運営、最終期	プログラム持続、参加者フォロー	受託企業、福祉部門・部局	継続意欲を高める仕組み。
最終評価・支払	評価機関が最終成果を検証→最終成果に基づく支払	評価機関、福祉部門・部局	成果の精査、公表と説明責任対応。評価結果を将来政策にフィードバック。

3. 指標設計・成果連動条項の構成（A市例）：

- ・成果指標（契約上の主要指標）には、総参加人数／累積参加者数、継続参加人数、要介護状態進行遅延人数などが採用された。
- ・契約時には、最低支払（一定の固定額）と成果連動分支払を区分設定。
- ・財政効果の試算では、介護給付費抑制効果を基に試算を行い、支払上限額や返戻率（事業者の報酬率）を検討。例えば、介護給付費抑制効果を前提に支払可能額を見積もって構造化した。
- ・契約書には、4年間の事業の2年目の評価・最終評価のタイミング・基準・評価手法を明記。支払時期（例：4年間の事業の2年目の成果→翌3月、最終成果→翌3月）も定められていた。
- ・A市は評価結果を基に、次期介護予防総合事業計画に「民間委託型成果評価モデル」を明記し、社会的インパクト評価の定常運用を目指す方向を打ち出した。

3 章 様々な評価研究デザインと分析手法、選択の視点

Ⅱ -1 章で説明した 3 条件を満たすため、これまで多様な評価研究デザインや分析手法が開発されてきた。本節では、多くの選択肢の中からどの方法を選択するのかを検討する際の基本的な考え方について説明する。

1) 様々な評価研究デザインと分析手法におけるトレードオフ関係

社会的インパクト評価に用いられる評価研究デザインや分析手法には、下表および参考資料に示すように、多様なものが存在する。一般に、現場での実施が容易な方法では、見かけ上の効果が紛れ込みやすい。一方で、これらの影響を排除できる学術的に厳密な評価方法ほど、行政現場での実施が容易ではないというトレードオフ（一方を立てると、もう一方が犠牲になる）の関係がある。そのため、評価方法の選択にあたっては、学術的妥当性と実務上の実施可能性の双方を踏まえた総合的な判断が必要となる。

主な評価研究デザインと分析手法の一覧

手法	3条件 対照 前後 比較	背景 要因	頁	分析方法	強み	弱み
1. 無作為化比較試験 (RCT)	○	○	○	P.64	施策の参加者（介入群）と非参加者（非介入群）を無作為に振り分けて効果を推定する。	・ 施策の効果を正確に評価することが可能となる。 ・ 実務上、施策の参加者を無作為に振り分けることは困難なことが多い。
2. 回帰不連続デザイン	○	○	○	P.66	施策を受けるかどうか、ある 1 つの基準で決定される場合、閾値（施策を受けられるかどうかの境目）の前後の対象者を比較することで効果を推定する。	・ 閾値（施策を受けられるかどうかの境目）の前後では、施策の効果を正確に評価することができる。 ・ 施策実務を変更しなくても実施できる可能性がある。
3. 傾向スコアマッチング	○	○	○	P.67	施策参加者と非参加者のなかから、特性の似た個人をマッチングして、効果を推定する。	・ 施策を受けるかどうか測定可能な変数により決定されている場合、効果を正確に評価することができる。 ・ 施策実務を変更せずに実施できる。
4. 操作変数法	○	○	○	P.69	アウトカム指標には直接的に影響はしないが、施策には影響を与えるような変数（操作変数）を用いて、効果を推定する。	・ 施策実務を変更せずに実施できる。 ・ 適切な操作変数を見つけることができる。 ・ 場合によっては、施策担当者が適切な操作変数を作り出すこともできる。
5. 差の差分分析	○	○	○	P.70	施策参加者と非参加者のそれぞれについて、施策実施前後のデータを用いることで、トレンド要因を取り除いた上で効果を推定する。	・ トレンド要因を取り除くことができるため、前後比較よりも厳密な分析が可能となる。 ・ 他の手法と組み合わせることも可能。
6. 合成コントロール法	○	○	○	P.71	施策非参加者のデータを合成し、施策対象者が施策を受けなかった場合の仮想的な状況を推計し、施策の効果を推定する。	・ ひとつの集団しか施策対象者がいなかった場合でも分析ができる。 ・ 施策実務を変更せずに実施できる。
7. 回帰分析	○	○	○	P.72	施策の参加者と非参加者について、個人特性などの変数を考慮した回帰モデルを用いてアウトカムを比較し、施策の効果を推定する。	・ 比較的簡便な方法であり、個人特性などのデータがあれば分析ができる。 ・ 観察可能な「第 3 の要因」（交絡要因）を統計的に調整した分析ができる。
8. 一般指標と前後比較	○	○	×	—	施策参加者のアウトカムの変化を、政府統計などの一般的な指標（全国平均や地域平均など）の変化と比較して、効果を推定する。	・ 追加の調査を行わなくても、既存の統計データを利用して比較分析できる。 ・ 社会全体の動向の中で施策参加者の変化を把握できる。
9. 前後比較	×	○	×	—	施策参加者について、施策実施前と後のアウトカムを比較し、変化量をもとに施策の効果を推定する。	・ 観察されていない要因（交絡）や参加者の選定に偏りがある場合、効果を正確に推定できない。 ・ 施策以外の要因の影響を排除できない。

（総務省行政評価局 効果的な政策立案・改善に向けた政策評価のガイドラインより引用、一部改変）

2) 選択の視点

前節で説明した 3 条件をどのように確保するのかについて、評価方法を選択すると

きの考え方の概要を図 23 のフローチャートに示す。

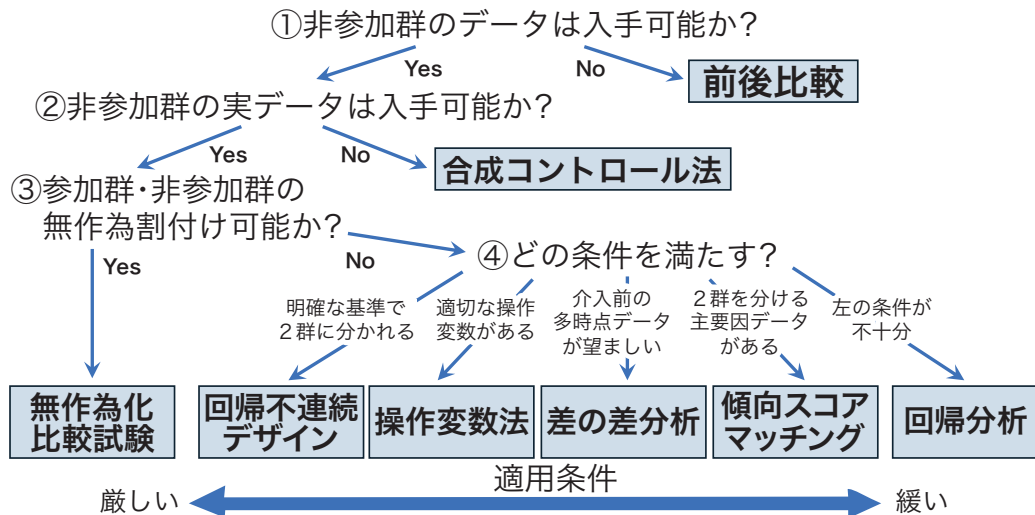


図23 評価デザインの選択フロー

選択の視点

無作為化比較試験（Randomized Controlled Trial：RCT）は、日本でも行政現場で用いられた事例が生まれているものの、実験的な性格が強く、関係者の理解や協力を得るための調整が必要となるなど、現時点では一般的な評価研究デザインとは言えない段階にある。

実験的なデザインを用いず、観察された収集データのみを用いる評価方法も複数存在し、実務上はこれらが適用しやすい場合が多い。しかし、操作変数法（Instrumental Variable：IV）や差の差分析（Difference-in-Differences：DID）などの方法では、適用のために成立している必要のある複数の前提条件が存在する（**コラム 39 操作変数法と差の差分析、傾向スコアマッチングの前提条件** 参照）。これらの前提条件が満たされているかを確認することは必ずしも容易ではない。

RCT 以外の介入研究で広く用いられている手法の一つに、傾向（プロペンシティ）スコアマッチングがある。異なる見解もありうるが、本文書では、操作変数法や差の差分析が前提条件を満たすことが実務上難しい場合が多いことを踏まえ、傾向スコアマッチングは行政実務の観点からも比較的適用しやすく、学術的にも（制約はありつつも）一定（回帰分析以上）の妥当性を有する手法と考える。ただし、この手法にも、観察できない交絡要因には対処できないなどの限界点があることに留意する必要がある。

さらに、プログラムの発展段階や実施体制によって、評価の力点の置き方は変化する。「設計・導入期」は事業が計画どおりに進んでいることを確認することに力点を置き、定性的なアウトカムの把握を中心に評価を行う。「形成・改善期」は実施状況の分析

に力点を置き、定量的なアウトカムの把握に力を入れ始める。そして、「成熟・普及期」にあたる、事業モデルが確立した上でさらなる改善や拡大に取り組む段階では、より厳密な効果検証のためのアウトカムの分析に力点を置く。専門家を交え、上記の方法により厳密な効果検証を行うことが望ましいが、使用可能なデータや人員の制約により、より簡便な手法（回帰分析や前後比較など）を選択することも検討される（P.55 コラム「効果分析に必要なデータの内容」を参照）。

コラム 39

操作変数法と差の差分析、傾向スコアマッチングの前提条件

操作変数法（参考資料 p67 参照）では、次のような前提条件が満たされる必要がある。①操作変数と説明変数（参加群か非参加群か）が強く関連していること、②操作変数は説明変数を通じてのみアウトカムに影響すること、③操作変数の値は各個人にとって偶然的・外生的（原因が個人の内部の要因ではなく外部の要因によって引きこされる）に決まっていること（観測されない交絡因子と独立していること）、④操作変数が説明変数に与える影響の方向が個人ごとに逆転しないこと、である。これらの条件が満たされる場合、操作変数法によって推定される介入効果は理論的には RCT による推定と同等の意味を持つ。しかし、これらの条件を満たす適切な操作変数を見つけることは一般に容易ではない。

差の差分析では、主に次のような仮定が必要となる。①施策実施がなかった場合、参加群と非参加群の変化が同じであること（施策実施前の 2 時点において、参加群と非参加群のアウトカムの変化が同じであることを検証することが多い）、②両群が施策以外の外的要因（社会状況の変化、災害、制度改正など）から同様の影響を受けること、③参加群・非参加群の割り当てがアウトカムの動きと無関係であること、④時間の経過に伴い、参加群・非参加群の構成が大きく変動しないこと、である。

傾向スコアマッチングにも妥当であるための前提条件が存在する。一般に、施策・事業への参加は観測可能な個人特性（この個人特性を傾向スコアに用いる）によって説明され、アウトカムとして用いられる健康状態などによる影響を受けないこと、である。ただし、この条件が完全に満たされているかを検証することはできない。そのため、参加群と非参加群の特性が類似していることを確認し、異なる条件設定による分析（感度分析）を行うことなどにより、推定結果の妥当性や一貫性を検討する必要がある。

【参考文献】

- (1) 総務省 効果的な政策立案・改善に向けた政策評価のガイドライン ～具体的な事例を中心として～ 2025. https://www.soumu.go.jp/main_content/000935597.pdf
- (2) 内閣官房行政改革推進本部事務局. EBPM ガイドブック Ver.1.2. 2023. https://www.gyokaku.go.jp/ebpm/img/guidebook1.2_230403.pdf

様々な評価デザインと 分析手法

- ・ここでは、インパクト評価の代表的な方法について、その違いや特徴を説明する。
- ・これら評価方法の目的は、妥当で意義のある比較の枠組みをつくることである。調査研究においては無数のバイアスが生じることが知られ、それらの影響を取り除き「真の介入効果」を評価することが求められる。

1) 実験研究デザインと準実験研究デザイン

(1) 実験研究デザイン（無作為化フィールド実験法）

- ・インパクト評価に利用できる選択肢の中で「介入効果を示す最も妥当な方法は、無作為化（ランダム化）フィールド実験法である」という視点がある。この方法は、介入と効果の因果関係（真の介入効果）を推定するのに最も理想的な方法といわれる。
- ・施策の対象者は2つ以上の群に（本人の希望とは無関係に）無作為に（ランダムに）割り付けられ、施策介入を受ける「介入群」と施策介入を受けない「非介入群（対照群）」に分け、各群の効果を比較する。無作為に割り付けることにより、因果関係が逆方向である可能性や、第3の要因が影響を与えている可能性を排除でき、施策の効果を正確に測定することができる⁽¹⁾⁽²⁾。

(2) 準実験研究デザイン（非無作為化準実験法）

- ・無作為化フィールド実験法でないインパクト評価は非無作為化準実験法と呼ばれ、参加者と背景要因（年齢、性別など）が同等の非参加者を比較する。無作為割り付けによる実験は行われないため、準実験法と呼ばれる⁽¹⁾⁽²⁾。
- ・準実験法を用いたデザインは一般的に、無作為化フィールド実験法に比べて結果の説得力が弱く、効果推定の妥当性から見ると、無作為化フィールド実験法は特段の選択肢である。しかし社会環境において、無作為化フィールド実験法は実施が現実的でない場合が多く、インパクト評価において準実験法は有用な方法である。

(3) インパクト評価のデザイン選択における考え方

- ・真の効果を追求するインパクト評価において無作為化フィールド実験法は理想的であるが、現場では「最もよい」評価デザインを選択することが難しい状況にしばしば遭遇する。技術的に最良であっても、人を対象とした無作為化フィールド実験法を倫理的に実践することの限界や時間と資源の制約があり、資源投入の正当性は介入の重要度や結果の活用度により異なる。
- ・インパクト評価において常に最良といえるデザインはなく、評価者は結果の潜在的な重要度と各デザインの実用性、実現可能性、そしてそのデザインを用いた場合に有用で信頼できる結果が得られる可能性を考慮した上で、実施可能な方法の中から

最も強力なデザインを選ぶ必要がある。「完璧な」インパクト評価ではなく、「まあ十分」なインパクト評価を目指し、評価研究デザインを構築する⁽²⁾。

コラム 40

評価・分析単位

インパクト評価において、アウトカムが測定される単位を分析単位と呼ぶ⁽¹⁾。分析単位は必ずしも人（個人）とは限らず、個人や家族、近隣の人々、地域、学校や会社などの組織、あるいは市区町村、都道府県といった政治的管轄区域などがある。施策介入の性質とそれが提供される対象集団の単位に基づいて選択する。

コラム 41

クロスオーバー法、ウェイトングリストコントロール法

施策介入を受ける群と受けない群を比較する無作為化比較試験では、倫理上の問題が提起されやすい。新たな社会介入は一般的にはよかれと考えるため、介入を受けない群が相対的に不利益を被ると想像されるからである。（ただし介入の効果は事前にわかるものではなく、副作用がみられる場合もあることから、本質的にはこの指摘は適切でない。）

こうした問題意識と主張は根強く、それに対処するための方法がいくつか提案されている。例えば別のサービスを提供して心理的な均衡をとる、各群について介入を受ける順序を交互にして2群とも介入を受けられるようにする（クロスオーバー法）、開始時期をずらして非介入群にも遅れて介入を行う（ウェイトングリストコントロール法、段階的割り込みクラスターデザイン）などの回避手段がある⁽¹⁾。

2) 様々な分析手法の概要

(1) 無作為化比較試験 (RCT)

- ・ 真の介入効果を推定するのに最も理想的と言われるのが無作為化比較試験 (Randomized Controlled Trial : RCT) と呼ばれる手法である⁽³⁾。
- ・ 施策の対象者を、参加者と非参加者に無作為（ランダム）に振り分けた上で、両者のアウトカム指標を比較する。無作為の割り付けにより、因果関係が逆方向である可能性や、第3の要因が影響を与えている可能性を排除することができ、施策の効果を正確に測定することができる。
- ・ 無作為化比較試験を行うことは、実務上簡単ではないケースが多く、また参加者の条件が限られることから必ずしも結果を他の集団に当てはめられる訳ではないが、1つの理想形として念頭に置きながら、実施可能な方法の中から、より精緻な方法を検討していくことが重要である。
- ・ また、エビデンスに基づく施策・事業の構築のためには、実験室的に統制された環境での取り組みではなく、通常的环境において試験を実施すること、いくつかの施

設（対象者）共同による試験を実施することが重要である。

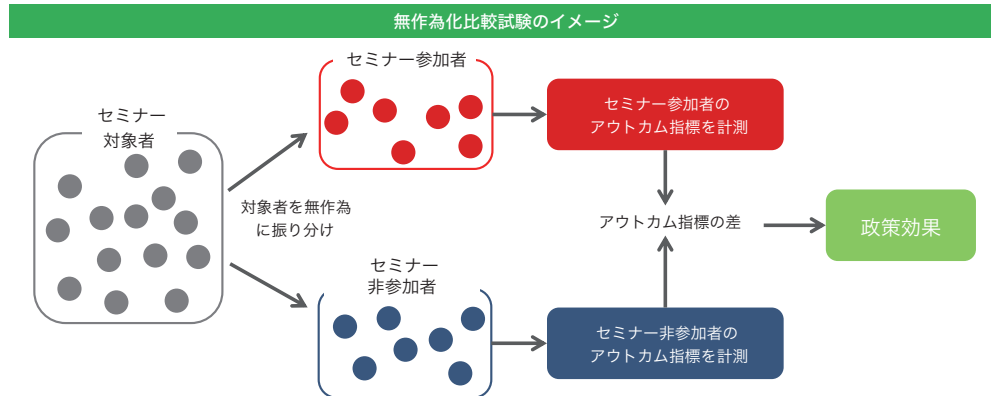


図24 無作為化比較試験の概要⁽³⁾

コラム 42

クラスター無作為化比較試験（CRCT）

- ・無作為化比較試験は個人単位だけでなく、クラスター（集団）単位で行うこともできる⁽³⁾。下図では、演習を取り入れたセミナーと演習を取り入れないセミナーの効果の差を測定することを想定している。
- ・介入の内容により個人単位での無作為化が難しい場合は、クラスター単位で実施内容を無作為に割り当てて評価を行うことも可能である。

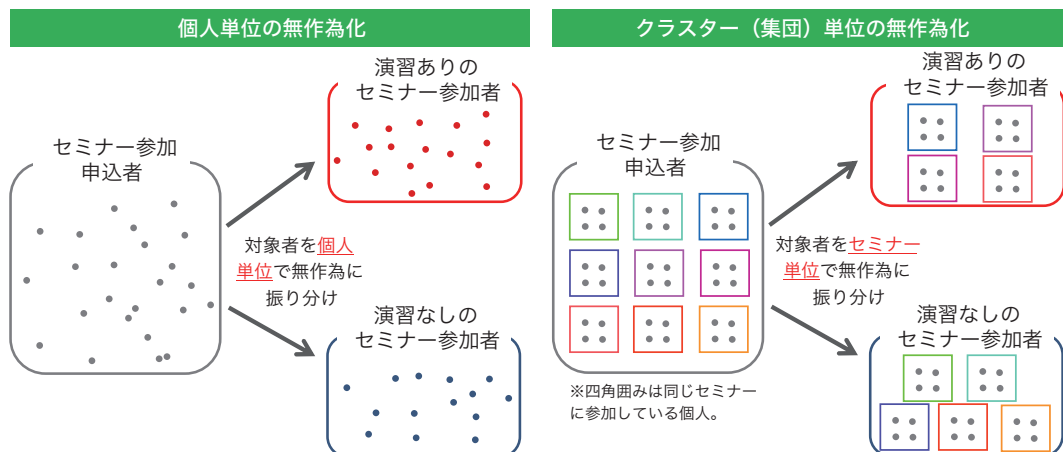


図25 クラスター無作為化比較試験（CRCT）の概要⁽³⁾

国内の先行研究例：

- ▶ Shimoda A, Saito Y, Kondo N. Postal reminder with reduced burden of clinic information seeking for equitable participation in general health check-ups: A randomized controlled trial. *Prev Med.* 2022;157:107009. doi:10.1016/j.ypmed.2022.107009
- 全国健康保険協会（協会けんぽ）加入者 21,017 人を対象に、自宅から最寄りの健診実施医療機関を明示した“負担軽減型リマインダー”の郵送が一般健診受診率に与える影響を検証した RCT。リマインダーの郵送により、健診受診率が向上（3.2% vs 2.1%、オッズ比 = 1.55、95% 信頼区間：1.31–1.85）し、低所得層ほど効果が大きく、所得による受診格差縮小の可能性も示唆された。
- ▶ Fukuma S, Sasaki S, Taguri M, et al. Effect of Nudge-Based Intervention on Adherence to Physician Visit Recommendations and Early Health Outcomes among Individuals Identified with Chronic Kidney Disease in Screens. *J Am Soc Nephrol.* 2022;33(1):175-185. doi:10.1681/ASN.2021050664
- 健診で慢性腎臓病が疑われた住民 4,011 人を対象に、ナッジ入りレターなど 3 種類の通知を無作為に送り、医療機関受診率と早期健康アウトカムへの影響を比較した RCT。ナッジを取り入れた通知は通常のお知らせよりも医療機関受診率を向上させた。一方、腎機能などの早期健康アウトカムには明らかな改善は認めなかった。
- ▶ Kamada M, Kitayuguchi J, Inoue S, et al. A community-wide campaign to promote physical activity in middle-aged and elderly people: a cluster randomized controlled trial. *Int J Behav Nutr Phys Act.* 2013;10:44. Published 2013 Apr 9. doi:10.1186/1479-5868-10-44
- 島根県雲南市の中高年住民をコミュニティ単位で無作為化し、身体活動促進のための地域横断的キャンペーンの効果（歩行・運動の実施）を評価したクラスター RCT。1 年間の地域キャンペーンでは身体活動の明らかな増加は確認されなかったが、身体活動に関する認知度や知識は向上し、行動変容にはより長期の介入が必要である可能性が示された。

(2) 回帰不連続デザイン (RDD)

- ・回帰不連続デザイン (Regression Discontinuity Design : RDD) は、施策・事業の対象者と非対象者がいる 1 つの基準（閾値、年齢や異常値など）で決められる場合に、閾値の前後の対象者を比較することで効果を測定する手法であり、準実験的な手法の中では精緻な方法とされる⁽³⁾。
- ・閾値ぎりぎりに位置する対象者同士の個人特性はかなり似通っていると考えられる

ため、もし下図のような効果量の「ジャンプ」があれば、これを施策の効果とみなして測定することが可能である。また、閾値の前後を比較すれば効果測定が可能であるため、施策・事業実務を変更しなくても実施できる可能性がある。

- ・一方で、閾値から離れた対象者について効果を把握することはできない（＝施策・事業全体の効果を評価することはできない）。また、基準が他の施策・事業と同じであると、他の施策・事業の効果と識別できないため留意が必要である。

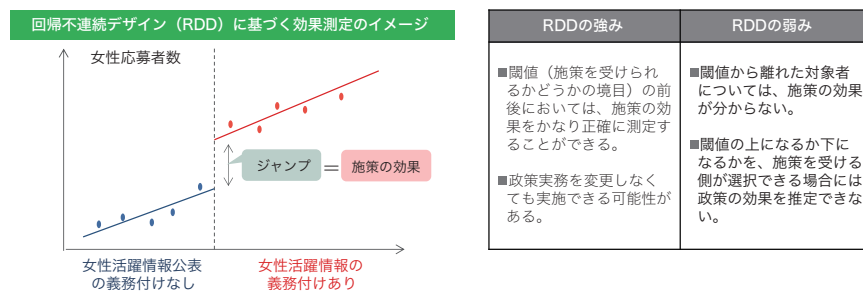


図26 回帰不連続デザイン（RDD）の概要^③

国内の先行研究例：

- Fukuma S, Mukaigawara M, Iizuka T, Tsugawa Y. Impact of the national health guidance intervention for obesity and cardiovascular risks on healthcare utilisation and healthcare spending in working-age Japanese cohort: regression discontinuity design. *BMJ Open*. 2022;12(7):e056996. Published 2022 Jul 29. doi:10.1136/bmjopen-2021-056996

→腹囲の基準により特定保健指導の対象となった就業世代を対象に、カットオフ付近の人々を比較する回帰不連続デザインを用いて、保健指導が医療利用および医療費に与える因果効果を検証した研究。腹囲の基準に基づく特定保健指導により外来受診日数の減少を認めたが、薬剤使用、入院、医療費には明らかな差を認めなかった。

(3) マッチング（傾向スコアマッチング）

- ・ マッチングは、観測可能な変数で条件付けることにより施策・事業の参加者と属性の似通った非参加者をマッチング（＝ペアとして抽出）させ、両群のアウトカムの差を比較することで効果を測定する手法であり、無作為化比較試験や回帰不連続デザインを適用できない場合に、適切な手法となり得る^③。
- ・ マッチングを行う際の具体的な方法として、参加しやすさの確率である「傾向スコア」に基づくマッチングを行う傾向スコアマッチング（Propensity Score Matching: PSM）がある。

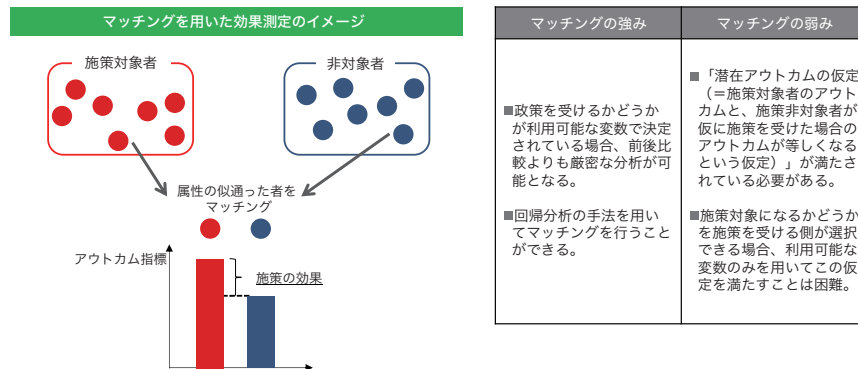


図27 マッチングの概要⁽³⁾

- ・ 個人が施策介入を受ける（参加する）確率は参加群と非参加群で、もともとの背景要因が異なることが多く^(※)、その違いを考慮せずに2群の効果を推定すると、得られた効果量は参加確率の違いの影響（選択バイアス）を含む可能性が高い。（※例えば、健康増進のための政策を考えた場合、健康リスクが高い集団の方が政策を受ける対象になる可能性が高い、といったことが挙げられる。）
- ・ 選択バイアスの影響を取り除くためには、施策介入を受ける確率が同等の集団に限定して、介入による効果を計算・比較すればよい。ここで用いられる手法の1つが傾向スコアマッチングである。
- ・ 具体的には、
 - ①参加する確率（傾向スコア）を計算
 - ②参加群と非参加群（対照群）の中から傾向スコアが近いものを抽出（マッチング）する
 - ③抽出された参加群・非参加群（対照群）を用いて介入の効果を推定する
- ・ 比較的取り組みやすく、学術的に見ても妥当性が高い手法と考えるが、この手法においても、観察できない交絡要因には対処できない、などの限界点があることに留意が必要である。

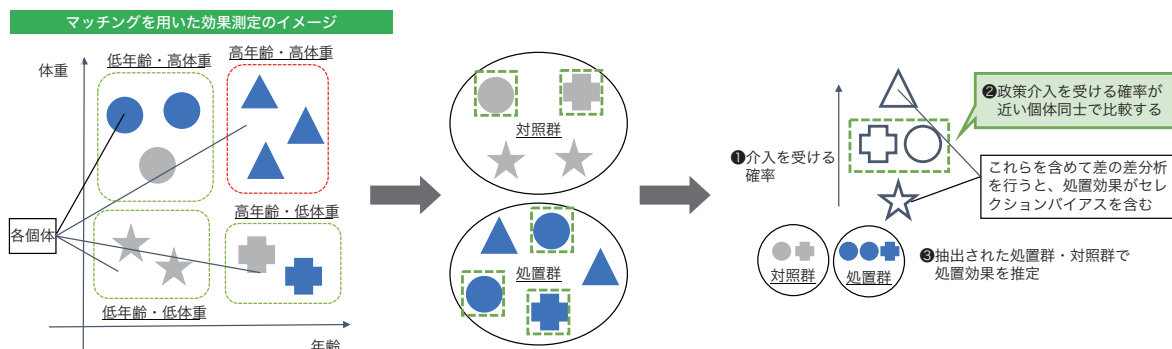


図28 傾向スコアマッチング（PSM）の概要⁽³⁾

国内の先行研究例：

▶ Makizako H, Shimada H, Doi T, Tsutumimoto K, Yoshida D, Suzuki T. Effects of a community disability prevention program for frail older adults at 48-month follow up. *Geriatr Gerontol Int*. 2017;17(12):2347-2353. doi:10.1111/ggi.13072

→地域で暮らすフレイルの高齢者について、介護予防プログラムの参加者と非参加者を傾向スコアマッチングを用いて比較し、身体機能維持や要介護発生リスク低減に寄与するかを検証した研究。介護予防プログラム参加者は、非参加者と比べて48か月間の要介護発生率が低く（11.3% vs 19.8%）、要介護発生リスクも約45%低かった（ハザード比=0.55、95%信頼区間：0.35-0.88）。

(4) 操作変数法

・例えば、家計の所得と子どもの学力との因果関係を明らかにしようとした場合に、それぞれに影響を与える外部要因（例：両親の教育レベル等）があると、家計の所得と子どもの学力の間に相関があったとしても、

① 家計の所得の向上が子どもの学力の向上をもたらすのか

② 両親の教育レベルが家計の所得と子どもの学力を引き上げていて、家計の所得の向上は子どもの学力の向上をもたらすわけではないのか

はわからない。

・そこで、家計の所得とは関係するが、外部要因とは直接関係しない変数を操作したときに子どもの学力が変わるか否かを分析する手法が「操作変数法」である。

→子どもの学力が変われば、外部要因とは無関係に効果が変わった＝家計の所得と子どもの学力とは関係があるといえることができる。

→子どもの学力が変わらなければ、家計の所得と子どもの学力との間の因果を立証することができず、外部要因による影響の可能性を排除できないということになる。

・この方法を用いる際には、優れた操作変数の条件とされる「政策と強い相関をもち、外部要因と一切相関がない変数」を見つけることは一般に困難な点に留意が必要である⁽³⁾。

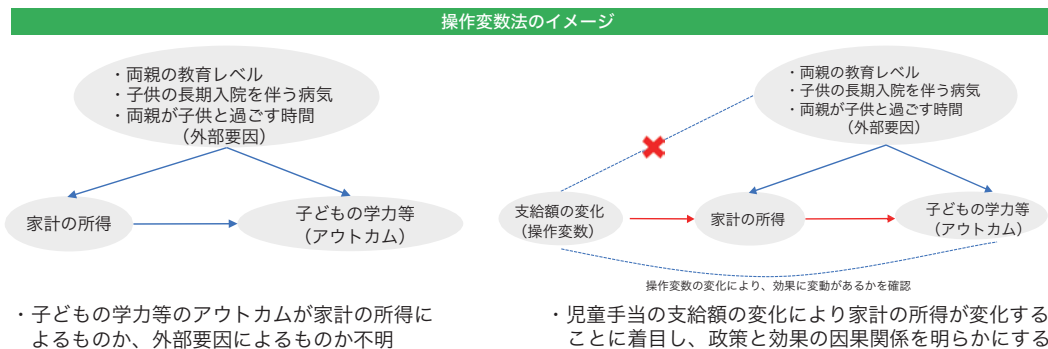


図29 操作変数法 (IV) の概要⁽³⁾

国内の先行研究例：

- ▶ Sato K, Kondo N, Hanazato M, Tsuji T, Kondo K. Potential causal effect of physical activity on reducing the risk of dementia: a 6-year cohort study from the Japan Gerontological Evaluation Study. Int J Behav Nutr Phys Act. 2021;18(1):140. Published 2021 Oct 29. doi:10.1186/s12966-021-01212-w

→日本の地域在住高齢者を対象とした6年間のJAGESコホートにおいて、地域の身体活動環境（積雪量）を操作変数として用い、身体活動が認知症発症リスクに与える潜在的な因果効果を検証した研究。身体活動は認知症発症リスクを低下させ、その効果は少なくとも4年間持続する可能性が示された。

(5) 差の差分析 (DID)

- ・差の差分析 (Difference-in-Differences: DID) は、施策・事業の対象者と非対象者それぞれの施策・事業の実施前後のデータの差を比較する分析手法である⁽³⁾。
- ・DIDでは、参加者と非参加者それぞれの施策実施前後のデータを用いることにより、トレンド要因を取り除くことができるため、一群（参加者のみ等）での前後比較よりも厳密な分析が可能となる。
- ・一方で、DIDを用いるには施策実施前後の両方のデータが必要となる。また、対象者と非対象者の施策実施前後のトレンドが平行になっている（平行トレンドの仮定）確認のためには3時点データが必要があることに留意が必要である。

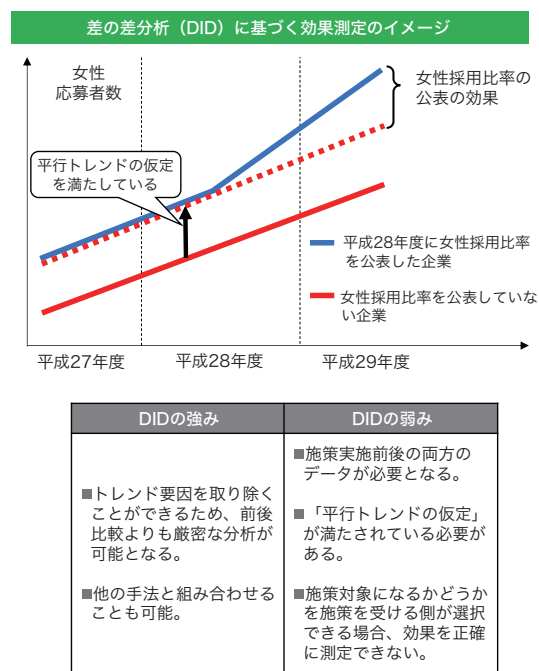


図30 差の差分析 (DID) の概要⁽³⁾

国内の先行研究例：

▶ Nagata H, Sato K, Haseda M, Kobayashi Y, Kondo N. A novel behavioral science-based health checkup program and subsequent metabolic risk reductions in a workplace: Checkup championship. Prev Med. 2022;164:107271. doi:10.1016/j.ypmed.2022.107271

→大企業の従業員を対象に、行動科学に基づく職場健診プログラム「健診戦」への参加がメタボ関連指標の改善に与える影響を、参加者と非参加者の前後差を比較する差の差（DID）分析および逆確率重み付け（Inverse Probability Weighting：IPW）を用いて検証した研究。「健診戦」の参加者は、非参加者と比べて体重、BMI、腹囲、収縮期・拡張期血圧の改善を認めた。特にメタボリックシンドロームのリスク因子を2つ以上有する高リスク者では、LDL コレステロールを含め、より大きな改善が認められた。

(6) 合成コントロール法

・施策非実施地域のデータを合成することにより、施策実施地域が施策を導入しなかった場合の仮想的な状況を推計し、効果を測定する方法である⁽³⁾。

具体例（右図）

- ・類似地域のデータを合成することによって、施策（# 7119）導入地域と類似した合成コントロール（仮想データ）を構築する。それを、施策導入した実際の地域のデータと比較することで、施策の効果を分析することができる。
- ・青いグラフは2017年4月に# 7119を導入した地域であり、オレンジのグラフは合成コントロールである。青いグラフとオレンジのグラフの差分が# 7119の施策効果となる。

国内の先行研究例：

▶ Sun D, Mulder DT, Li Y, et al. The Effect of Nationwide Organized Cancer Screening Programs on Gastric Cancer Mortality: A Synthetic Control Study. Gastroenterology. 2024;166(3):503-514. doi:10.1053/j.gastro.2023.11.286

→ WHO 死亡統計と各国共変量から「合成日本／合成韓国」を構築し、胃がん検

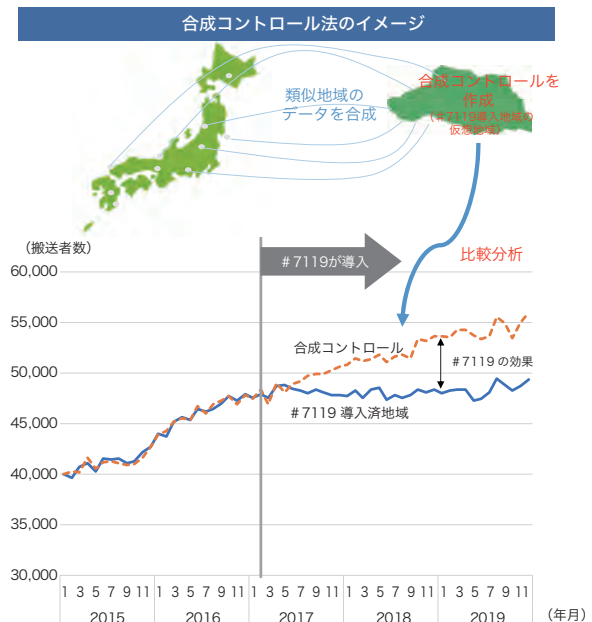


図31 合成コントロール法の概要

診導入が上部消化管疾患死亡率に与えた効果を検証した研究。韓国では、胃がん検診の導入により胃がん死亡率が低下し、導入 15 年後には約 41%減少した。一方、日本では明らかな死亡率低下は認められず、効果は不確実であった。

(7) 回帰分析

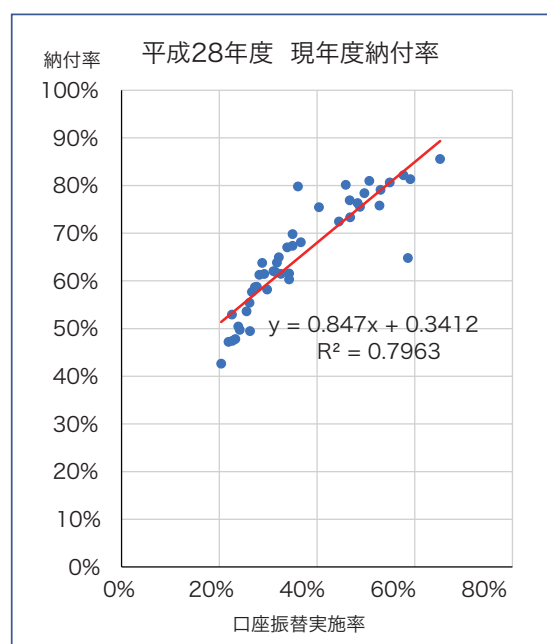
- ・ 回帰分析 (Regression Analysis) は、ある変数 x (説明変数と呼ぶ) が別の変数 y (被説明変数と呼ぶ) にどの程度影響を与えているか、統計学的に明らかにする手法である⁽³⁾。
- ・ 1 つの変数 (説明変数) が 1 つの変数 (被説明変数) に与える影響を評価する単回帰分析と、複数の変数 (説明変数) が 1 つの変数 (被説明変数) に与える影響を評価する重回帰分析に分けられる。
- ・ 例えば、図 32 では、口座振替実施率を x 、納付率を y とした上で、 $y = \beta_0 + \beta_1 x + \epsilon$ という単回帰モデルによって、定数項 β_0 ($=0.3412$) や x の係数 β_1 ($=0.847$) を推定している。(なお、 ϵ は誤差項と呼ばれており、上記のモデルでは説明できない部分を表現している。)
- ・ このモデルにおいて、 β_1 は口座振替実施率が変化した場合に納付率にどの程度影響を与えるかを現している。
- ・ 回帰係数の推定手法としては、最小二乗法が基本的かつ簡便な分析手法であり、上記モデルの誤差項 ϵ の二乗が最も小さくなるように係数を推定する手法である。

国内の先行研究例：

▶ Imamura H, Kogure M, Kita Y, et al. Factors Related to Participation in Health Examinations for Japanese National Health Insurance: NIPPON DATA2010. Journal of Epidemiology. 2018;28(Supplement_III):S53-S58. doi:10.2188/jea.JE20170251

→日本の国民健康保険加入者を対象に NIPPON DATA2010 を用いて、健康診断受診に関連する個人属性・生活習慣・社会経済的要因を重回帰分析で検証した研究。65～74 歳では、高学歴、持ち家所有、低い家計支出が健診受診に関連したが、40～64 歳では社会経済的要因との明らかな関連は認めなかった。

【仮説】 国民年金保険料の収納対策としては、口座振替の推進が有効ではないか。



▲ 単回帰分析の結果。他の年金事務所に比べ口座振替実施率が 1 %高い事務所は、納付率が 0.847%高いことが理論的に予測される。

図32 回帰分析を用いた効果検証の例⁽³⁾

コラム
43

Target Trial Emulation や機械学習などの新たな分析手法

これまでに紹介した方法のほか、近年では新たな分析手法も活用され始めている。例えば、仮想的な無作為化比較試験（RCT）を設定し、観察されたデータから RCT を模倣する形で分析を行い、介入効果を推定する Target Trial Emulation がある。また、機械学習の方法を用いることで、従来は評価が難しかった、複数の属性による効果の異質性（どの集団で効果が大きいか等）の分析が可能になりつつある。これらの手法は、今後のインパクト評価においても活用が期待される。より専門的な内容については、日本医療研究開発機構 E-LIFE ヘルスケアナビ「予防・健康づくりプログラムの統計的評価デザイン～理解・実践のための手引き～」⁽⁴⁾ 参照。

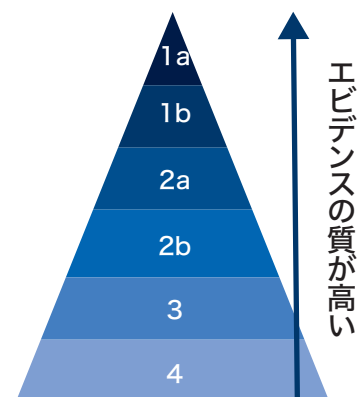
3) エビデンスレベルと評価デザインの選び方

(1) エビデンスレベル

- ・分析により得られた結果（効果の推定値）について、その信頼性の程度を示す概念としてエビデンスレベルがある。一般に、エビデンスレベルは研究デザインの種類に基づいて整理されることが多い。
- ・従来は、最も高いレベルとして、複数の質の高い無作為化比較試験の結果を統合・分析したメタアナリシスや系統的レビュー、その次に無作為化比較試験、さらにその次に準実験研究デザインなどが位置づけられることが多かった⁽⁵⁾。
- ・しかし、メタアナリシスや無作為化比較試験であっても、質の高いエビデンスを提示するためには一定の条件を満たす必要がある。そのため、従来のエビデンスのピラミッドについては見直しの必要性が提案されている⁽⁶⁾（コラム 44 エビデンスのピラミッドの見直しについて）。

（参考）エビデンスの質

レベル	内容	手法の例
1a	無作為化比較試験（RCT）のメタアナリシス、系統的レビュー	複数のRCTの系統的レビュー
1b	少なくとも1つのRCT	RCT（フィールド実験）
2a	少なくとも1つの準実験	回帰不連続デザイン、操作変数法、差の差分分析、傾向スコアマッチング、合成コントロール法
2b	少なくとも1つの非実験的研究	回帰分析
3	比較試験、相関研究、記述的研究	前後比較
4	専門家や事務家の意見	検討委員会による討議、パブリックコメント



（内閣官房行政改革推進本部 行政事業レビューシート作成ガイドブック Ver 1.2より引用、一部改変）

図33 エビデンスレベルと代表的な研究デザイン

(2) 評価デザインの選び方

- ・一方、施策・事業立案の現場では、純然な無作為化比較試験を実施することは実務上きわめて困難な場合が多い。そのため、エビデンスの質の違いを理解した上で、収集可能で施策判断に資する良質なエビデンスを見極め、適切な評価デザインを採用することが重要である⁽⁵⁾。実務上は、傾向スコアマッチングをはじめとする準実験研究デザインが至適となる場合が多い。これらの評価デザインの選択は、専門家の助言を得ながら行うことが望ましい（p60「選択の視点」を参照）。

コラム 44

エビデンスのピラミッドの見直しについて

これまで、無作為化比較試験（RCT）の結果を最も質の高いエビデンスとするヒエラルキー（エビデンスのピラミッド）が提唱されてきた。しかし近年では、RCTであっても研究設計や実施の質によってエビデンスの信頼性は大きく異なることが広く認識されている。そのため、特定の研究デザインや分析手法を用いるだけでエビデンスの質が担保されるわけではないと考えられている。また、複数の RCT を統合した系統的レビューやメタアナリシスについても、対象となる個々の研究の質やばらつき、バイアスの可能性などを考慮した上で慎重に評価すべきであるとされている。このような背景から、従来のエビデンスのピラミッドの見直しが提案されている⁽⁶⁾⁽⁷⁾。



A: 従来のピラミッド

<提案されている新たなピラミッド>

B: 研究デザインを分ける線が波状になり、系統的レビューが切り離される

C: 系統的レビューはエビデンスを見渡すためのレンズであることを表現している

(Murad MH, Asi N, Alsawas M, Alahdab F. New evidence pyramid. Evid Based Med. 2016;21(4):125-127 より引用)

【参考文献】

- (1) Rossi PH, Lipsey MW, Freeman HE, (監訳) 大島 巖, 平岡 公, 森 俊, et al. プログラム評価の理論と方法：システムティックな対人サービス・政策評価の実践ガイド. 日本評論社；2005
- (2) 源 由理子, 大島 巖, 山谷 清志. プログラム評価ハンドブック：社会課題解決に向けた評価方法の基礎・応用. 晃洋書房. 2020
- (3) 総務省. 効果的な政策立案・改善に向けた政策評価のガイドライン ～具体的な事例を中心として～ 2025. Available from : https://www.soumu.go.jp/main_content/000935597.pdf
- (4) 日本医療研究開発機構. E-LIFE ヘルスケアナビ 予防・健康づくりの新たな研究手法の開発 予防・健康づくりプログラムの統計的評価デザイン ～理解・実践のための手引き～. 2025. Available from : https://healthcare-service.amed.go.jp/assets24/pdf/guidelines_evaluation.pdf
- (5) 内閣官房行政改革推進本部事務局. 行政事業レビューシート作成ガイドブック Ver.1.2. Available from : 2025. https://www.gyokaku.go.jp/review/img/R07sakusei-guidebook_ver1-2.pdf
- (6) 2022-2024 年度 国立研究開発法人日本医療研究開発機構（AMED）ヘルスケア社会実装基盤整備事業委託研究「SDGs を意識した予防・健康づくりの多面的経済評価の手法開発」研究班. 橋本 英樹. SDGs を意識した予防・健康づくりの多面的経済性評価；総括的展望と実践への提言（第2報）. 医療経済研究. 2025;37(1)2025.06
- (7) Murad MH, Asi N, Alsawas M, Alahdab F. New evidence pyramid. Evid Based Med. 2016;21(4):125-127

用語 INDEX

あ

アウトカム	5, 16-21, 25-27, 46-49
アウトカム（効果）	16
アウトカム評価	10, 17, 38
行政	11, 13
行政現場の立場	9
意思決定に資するエビデンス	9
インパクト	7, 11, 16, 17
インパクト投資	9, 16
EBPM	4, 5, 9, 10, 11, 49
Evidence Based Policy Making	4, 9, 10
エビデンス	9, 21-25, 35, 48, 49, 73, 74

か

課題設定	5, 13, 14, 15, 19, 25
学識経験者	5, 13, 50
科学的・学術的な妥当性	9, 11, 60
関連要因	14, 15, 25
ギャップ	14, 15, 25
KPI	6, 27, 37
クロスセクター・ベネフィット	15
現状把握	5, 13, 14, 19, 21, 22, 23

さ

サービス提供者	10
社会的インパクト評価	5, 13
社会的インパクト評価の定義	4, 9, 11
社会的インパクト予測評価	6, 33, 34
実施主体	13
実施体制づくり	5, 13
社会的インパクト評価の研究デザイン	7, 50
施策・事業の発展段階	8
説明責任	10, 36, 37, 38
ステークホルダー	21, 22, 37, 46, 48

た

第三者評価	10, 13, 45
大学・研究機関	4, 10
地方自治体	10, 21
中央省庁	4, 10
定量的	11, 15, 35
定性的	11, 15, 35
データベース	10, 24
Theory of Change	18, 19
トレードオフ	8, 26, 60

な

ニーズ	14, 15, 22
-----	------------

は

波及効果	15, 16, 17
バックキャスト法	19
PFS	5, 10, 11, 29, 48, 57-60
PICO	23
比較対照群	7, 50, 51
評価計画	6, 18, 35, 45
評価研究デザイン	8, 35, 36, 39, 48, 50, 60, 61, 64
評価指標	10, 19, 21, 27
評価の5原則	28
フォアキャスト法	20

ま

民間事業者	10, 13, 21
無作為化比較試験 (RCT)	2, 4, 9, 44, 61, 64, 73
見かけ上の効果	7, 9, 10, 16, 17, 48, 50, 52, 53, 54
測定指標	6, 20, 27, 28

ら

ロジックモデル	5, 6, 13, 14, 18-32, 48
---------	-------------------------

**保健・医療・介護（予防）領域における社会的インパクト評価ガイドライン 2026
—エビデンスに基づく政策の実現に向けて—**

2026 年 6 月

発行：一般財団法人医療経済研究・社会保険福祉協会 医療経済研究機構

〒105-0001 東京都港区虎ノ門1-21-19 東急虎ノ門ビル 3F

TEL：03 (3506) 8529

FAX：03 (3506) 8528

