

政策評価の理論と実践

2018.11.07

鳥取大学 地域学部
小野 達也

本日のプラン

- I あらためて評価とは
 - II プログラム評価について
 - III 業績測定について
 - IV 評価の系譜と評価制度
 - V 業績測定のかんどころ
 - VI エビデンスに基づく評価
- 参考文献

I あらためて「評価」とは

...

Evaluation revisited

評価とは

「評価は物事のメリット、値打ち、意義を体系的に明らかにすることである」 (Scriven, 1991:139)

“Evaluation is the systematic determination of merit, worth and significance of things”

評価 = 事実特定 + 価値判断
(Evaluation) (Factual identification) (Value determination)

出所: 佐々木亮(2015)評価士養成講座テキスト

評価の目的(1)

- 形成的評価と総括的評価 (Scriven)
 - ① 形成的評価 (Formative evaluation)

「コックが調理中にスープを味見すること」

 - ・事業や活動の開発・形成
 - ・継続的改善
 - ・立案、実施段階で行う評価
 - ② 総括的評価 (Summative evaluation)

「客が出来上がったスープを味わうこと」

 - ・事業や活動の成果の把握
 - ・プログラム継続・縮小・拡大の判断
 - ・一定期間経た段階で行う評価



相互に補完し合うことで、より効果的なプログラム形成につながる

評価の目的(2)

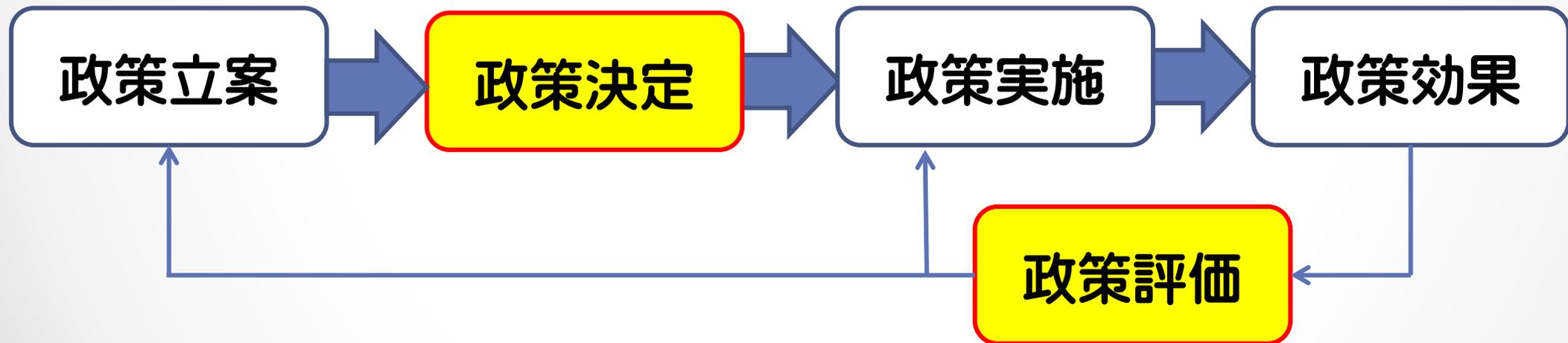
- 評価情報を何のために使うのか (Scriven, 2004)
 - 事業や活動の改善を目的としたもの(形成的評価)
 - 介入のアカウンタビリティの確保を目的としたもの (総括的評価)
 - 科学的な知識・知見を生み出すことを目的としたもの
- その他にもあるが、重要なことは、評価の「目的」をまず明らかにすること。それに合わせて評価方法(誰が、いつ、どのような手法で行うのか)を選択する必要がある。

政策評価の目的

●「政策評価」とは何か？

○政策評価活動は、「情報提供活動」の一種です。

→「政策評価」は、「政策決定」とは異なります。



- 政策実施の改善へのフィードバック ➡ 形成的評価
- 政策立案へのフィードバック ➡ 総括的評価

Ⅱ プログラム評価について

...

Program Evaluation

プログラム (Program) とは

- 「プログラム」とは
 - プログラムとは、ある社会目的を達成するための一連の活動群をさし、それらの活動を実施に導くルール、制度、組織体制や人材を含む取組み全体をさすものである。
 - 特定の組織が主体となって実施する場合もあれば、複数の組織・関係者が協働で実施する取組みもある

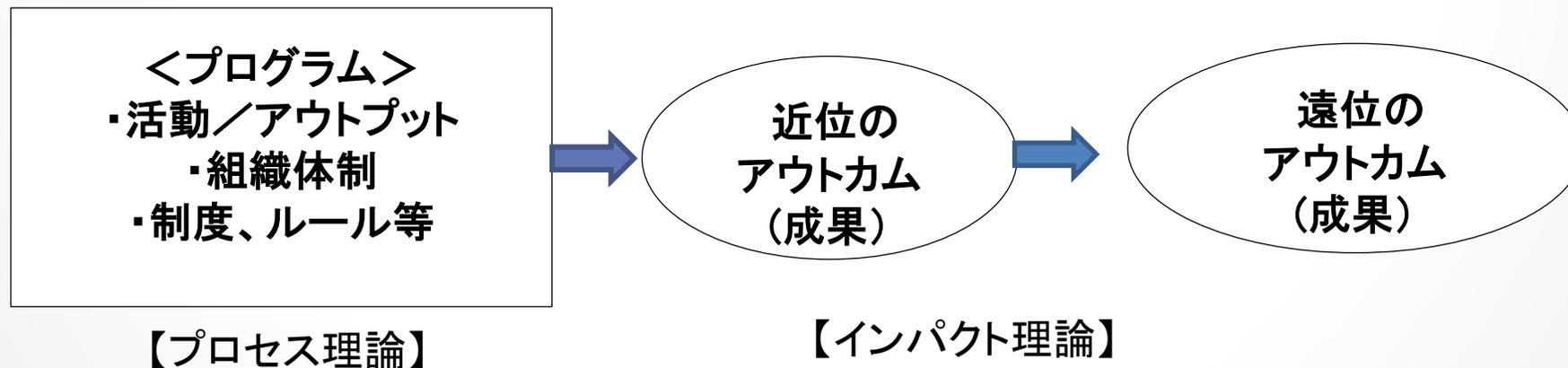
プログラム評価の定義

- プログラム評価とは社会調査の方法を活用し、社会プログラムによる介入の有効性を体系的に調査し、評価を行うこと。その評価は、プログラムを取り巻く政策的・組織的な文脈を考慮した方法で行われるもので、社会状況を改善するための活動の情報源となるものである(Rossi, Lipsey & Freeman, 2004: 16)
- 評価とは、プログラムや政策の改善に寄与するための手段として、明示的または暗示的な基準と比較しながら、プログラムや政策の実施やアウトカムを体系的に査定することである(Weiss, 1998:4)

プログラム理論とは

• プログラム理論 (Program Theory) という考え方

- ・「プログラムの実施と成果との間に存在する媒介メカニズム」(Weiss)
- ・プログラムの作戦や活動内容を決定するにあたり活用できる考え方であり、「めざす成果を得るためには、おそらくこういう活動が必要であろう」という「仮説」である
- ・その「仮説」が有効か、有効にするにはどうしたらよいか、を問う
- ・「インパクト理論」と「プロセス理論」のふたつの理論が包含される



ロジック・モデル

●ロジック・モデルとは何か？

→プログラム理論を踏まえてプログラムを表現する汎用的枠組み(世界標準)である。

●なぜ、ロジック・モデルが必要なのか？

→プログラムの運営方法や実施プロセスを明らかにすることでブラック・ボックス評価を回避できる。(インパクト理論のみならずプロセス理論の明確化)

→介入のメカニズムを明らかにすることで体系的な評価(5階層)が可能になる。

→目的の設定、指標の検討を行うときの枠組みとして活用できる。

→ステークホルダー間で、プログラムのコンセプト、作戦を検討し、共有するためのツールとして活用できる

●ロジック・モデルの構成要素は何か？

→4つの基本要素でプログラムを表現するものである。

「インプット」「アクティビティ」「アウトプット」「アウトカム」

プログラム評価の5階層～体系的に介入を評価する～

(Rossi, Lipsey and Freeman, 2004: 80)

プログラムのコストと効率性の評価

Assessment of Program Cost and Efficiency

プログラムのアウトカム／インパクト評価

Assessment of Program Outcome/Impact

プログラムのプロセスと実施の評価(プロセス評価)

Assessment of Program Process and Implementation

プログラムのデザインとセオリーの評価(プログラム理論評価／セオリー評価)

Assessment of Program Design and Theory

プログラムのニーズ評価

Assessment of Need for the Program

Ⅲ 業績測定について



Performance Measurement

業績測定 Performance Measurementとは

- 実績測定、業績評価、パフォーマンス・メジャーメントとも
- いくつかの定義
 - ＜業績測定のエッセンスというべき定義＞
 - サービスやプログラムの結果（アウトカム）と効率を定期的に測定すること（Hatry, 1999）
 - ＜業績測定という手法の定義＞
 - インプット・作業（活動）・アウトプット・アウトカム・効率（生産性）などの一連のデータを定期的に収集し報告すること（U.S.GAO 1992）
 - プログラムの実績、特に事前に設定した目標や基準の達成度合いについての体系的で継続的なモニタリング（U.S.GAO 2012）
 - ＜業績測定による評価システム・制度の定義＞
 - 業績（performance）の定期的な測定・評価・対策提言・公表」（上野・上野2007）

業績測定とプログラム評価の補完関係

- ①手法としての業績測定の基本は、プログラム評価のプロセス評価・アウトカム評価の階層にほぼ相当する。
- ②評価システム・制度としての業績測定の目的は、個々のプログラムにとどまらず、プログラムの集合や機関全体の継続的なパフォーマンス向上や効率化である。
- ③両者は、目的の異なる評価であり、役割分担によって互いに補完すべき(下表)。
- ④業績測定の結果を踏まえてプログラム評価の対象を選ぶ、プログラム評価の結果を踏まえて業績測定の指標を設定するなど、両者の連携も。

Nielsen&Hunter[2013], Hatry[2013]

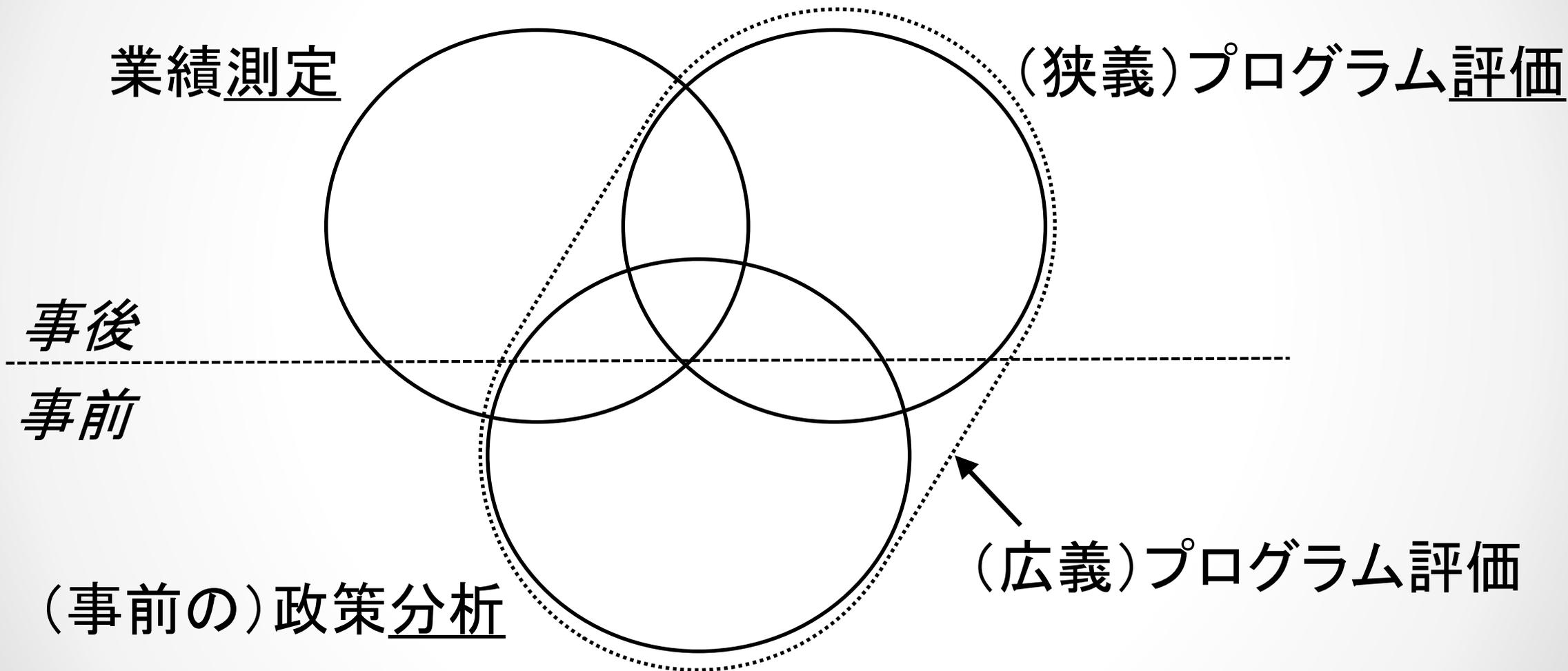
項目	業績測定	プログラム評価
タイミング	定期的・継続的	定期的でない
評価対象(数)	多数(しばしば全数)を対象に、一律	少数を対象に、1本ごと
情報の深さ	シンプルな業績データ	因果関係を追究
評価体制	内部評価が基本	専門的リソースが必要

IV 評価の系譜と評価制度

...

Measurement, Evaluation and Analysis

評価の系譜(拡張版)



系譜と制度

	測定	分析	評価
自治体の評価制度	計画の進行管理、事務事業評価	公共事業評価	
国の政策評価制度	目標管理型政策評価	事業評価(公共事業、ODA、規制)	総合評価、総合性確保評価
国の行政事業レビュー	行政事業レビュー		

V 業績測定のかんどころ

...

Tips for Performance Measurement

業績測定 of 指標に求められるもの

- 妥当性のある指標を設定すること
 - 業績測定において妥当性とはシンプルに「測定すべきものを測定していること」
 - ロジック・モデルに基づいて指標を設定することが推奨されている。
- 目標値は明確に設定すること
 - 根拠となるデータに基づいて、意味が明らかで曖昧さのない数値を
 - 目標達成の難しさが異なれば、達成度合いの評価も異なる。
- 実績値と目標値は的確に比べること
 - 最終的な水準とそこまでの変化のどちらに着目するか、などの的確に選択
- 指標間・プログラム間の比較、集計・集約は丁寧に行うこと
 - 比べられないものは比べないこと(例. フロー指標とストック指標の直接比較は禁物)

①指標の妥当性

- 「妥当性」validity とは、測定すべきものを測定していること。
- 指標が妥当なものでなければ、達成状況を的確に判断することができない。

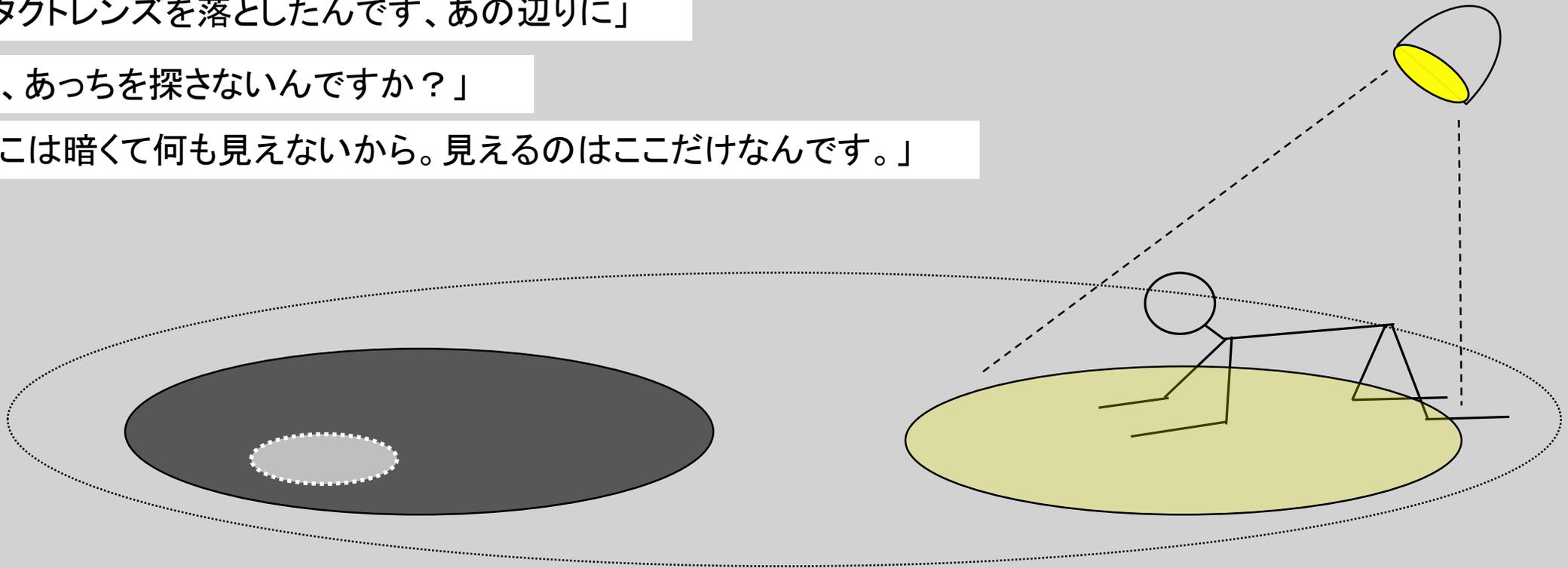
月のない夜、街灯の下で探しものをする男

「何か、お探しですか？」

「コンタクトレンズを落としたんです、あの辺りに」

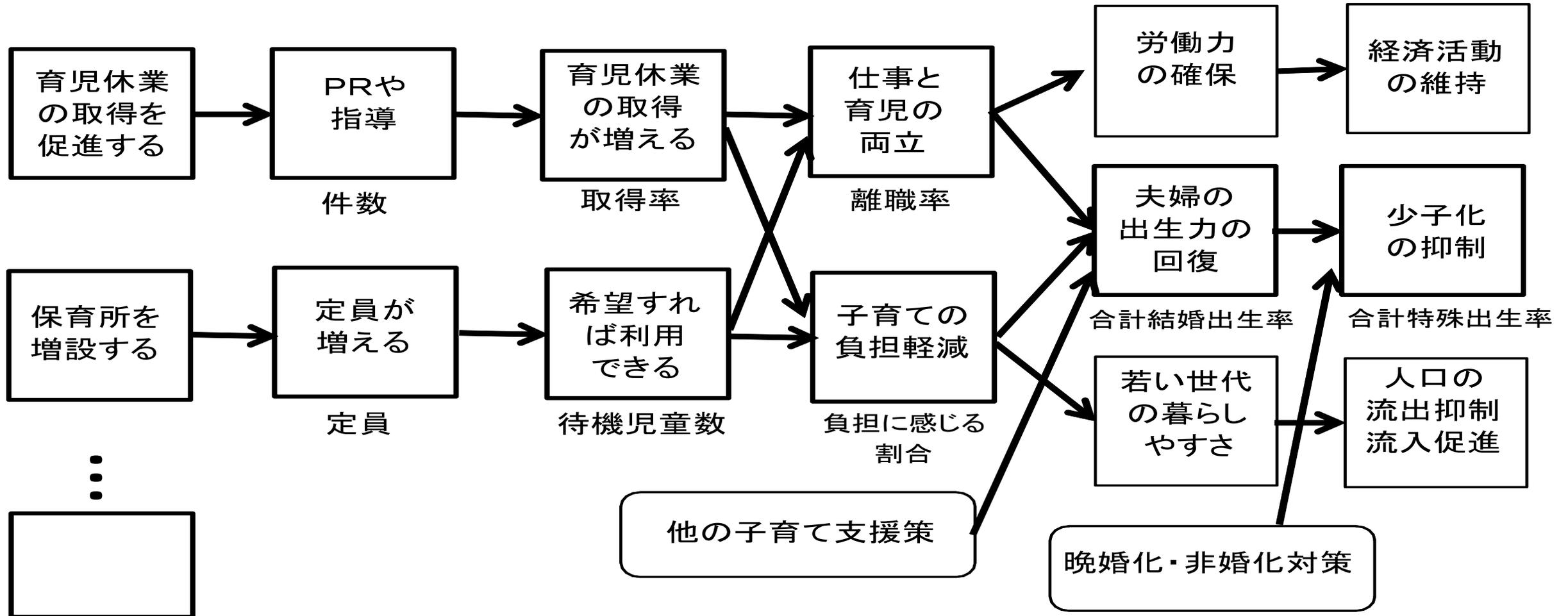
「なぜ、あっちを探さないんですか？」

「あそこは暗くて何も見えないから。見えるのはここだけなんです。」



ロジック・モデル＋指標

子育て支援策のロジックと評価指標（模擬例）



②目標値の性格

例)

①理想的な水準

達成は困難だが目指すべき状態としてあえて掲げる水準

②現実的な水準

一定の行政資源の投入と確度の高い効果の発現によって、達成までの経路を現実的に想定できる水準

③義務的な水準

深刻な問題の解消後の水準や有権者・市民に実現を約束した水準など

③ 目標達成度 — 実績値を目標値と比べる

実績値	目標値
80	100

- 目標達成度 $80 \div 100 \rightarrow 80\%$

開始時	終了時	目標値
50	80	100

- 目標達成度 $30 \div 50 \rightarrow 60\%$

	開始時	1年目	2年目	3年目
実績値	50	70	90	80
目標値	—	60	80	100

- 目標達成度 $240 \div 240 \rightarrow 100\%$

④フローとストックは混ぜるなキケン！

- 2つの政策が毎年度10のアウトカムを生み続けたとして、

ストック指標なら、 $10 \rightarrow 20 \rightarrow 30 \rightarrow 40 \rightarrow \dots$

フロー指標なら、 $10 \rightarrow 10 \rightarrow 10 \rightarrow 10 \rightarrow \dots$

VI エビデンスに基づく評価

...

Evidence-Based Policy Making

エビデンスとは

- 「結論をもたらす根拠となる情報」
- 「データの分析に基づく証拠」
- 「事実を報告する実証的な証拠」
- 「因果関係を示唆する根拠」
- 「プログラムや政策の評価に有用な(はずの), 統計学的手法によって得られる情報」
- 「一般化可能なデータを用いた統計解析やRCTによって明らかになった, 因果関係に関する実証的根拠」

広義エビデンスの階層のイメージ

A メタ・アナリシス、システマティック・レビュー

B RCT

C 準実験・自然実験

D 時系列比較、クロスセクション比較、回帰分析

E 目標値と実績値の比較

F 指標・データの妥当性

G データの信頼性、(非意図的な)誤用・濫用がないこと

H データの意図的な誤用・濫用がないこと

(出所)小野(2018c)

参考文献その1

- ロッシ、リプセイ、フリーマン[2005]『プログラム評価の理論と技法』, 日本評論社
- ワイス[2014]『入門 評価学－政策・プログラム研究の方法』日本評論社
- Newcomer, K. E., Hatry, H. P. and Wholey, J. S. (eds.) [2015]. *Handbook of Practical Program Evaluation, 4th Edition*. Hoboken, NJ: Wiley.
- 古川俊一[2002]「公共部門における評価の理論・類型・制度」『公共政策研究』第2号

参考文献その2

- 小野達也[2013]「政策評価と実績測定－府省の実績測定における計量・計数を巡って－」, 『日本評価研究』13巻2号
- 小野達也[2016]「自治体における業績測定型評価の現状と課題－20年を経過した都道府県の取り組みの点検結果から－」, 『日本評価研究』16巻1号
- 小野達也[2018a]「エビデンスに基づく政策評価－その含意と隘路」, 『統計』69巻7号
- 小野達也[2018b]「ロジックモデルを用いた評価指標の設定－業績測定の実効性の向上に向けて」, 『評価クォーターリー』46号
- 小野達也[2018c－11月刊]「エビデンス・ベーストな業績測定に向けて」『経済系』(関東学院大学経済学部・経営学部紀要)第275集

お疲れ様でした。