

2020年度

政策評価に関する統一研修(中央研修)

2021・1・27

科学的根拠に基づく政策立案・評価：
うざいけど必要なのはなぜか？

橋本英樹

東京大学大学院医学系研究科

5年目となるEvidence-based policy making体制

- 官民データ活用推進基本法施行(2016年12月)
- 統計改革推進会議最終報告(2017・5月)

“限られた資源を有効に活用し、国民により信頼される行政を展開するためには政策部門が、統計等を積極的に利用して、証拠に基づく政策立案(EBPM)を推進する必要がある”

6-7月 模擬セッション

10-11月 行政レビューでの試行

- EBPM推進委員会
(第1回 2017年8月、第2回 2018年4月、第3回2018年8月、第4回2019年9月)
 - 各省「政策立案総括審議官」体制
 - 統計等データの提供等の判断のためのガイドライン、EBPMを推進するための人材の確保・育成
の方針など決定
- 各省でモデル事業対象に政策、施策、事業の各段階での
 - ✓ “エビデンス”の活用
 - ✓ “因果関係”と“ロジックモデル”
 - ✓ “統計分析”と“評価”

その矢先、コロナ禍突入

- 第5回EBPM推進会議（持ち回り）

2020年5月20日

<https://www.kantei.go.jp/jp/singi/it2/ebpm/dai5/gijiyoushi.pdf>

有識者意見として橋本の意見表明（以下全文）

様々な制約のもと、新型コロナウイルス対応に加え行政対応にご尽力いただいていることに、改めて敬意と感謝の意を表させていただきたい。

未曾有の外生ショックを受け、社会・経済・医療はじめとするインフラが大きく揺さぶられるなか、十分な情報・エビデンスがないにもかかわらず、政策的判断を迅速に下さなければならない状況は、行政にとっても国民にとっても先が見えず、軋轢・ストレスの原因となっている。改めてエビデンスに基づくとは、こうしたストレスを回避するためのものだということが再認識される。

といっても、いまは臨床比較試験やら差の差分析やらをやる時ではない。幸い第1波は、諸外国に比べ死者数も低く抑えられているという結果を見る限り、なぜかうまくいっている。しかし、なぜうまくいったのかが説明できていない。いわば神風が吹いたおかげとしか言いようがない状況である。

来る冬、遅くとも来年には本格的な第2波が来る可能性も想定しておく必要があると考えるが、今回と同じように切り抜けられるかは保障がない。今回うまくいったからとのんびりしていれば、次は痛い目を見るだろう。

続き

第1波の収まる気配が見られ余裕が出たときに第2波に備えて体制の見直し・強化を比較的短期間に図る必要がある。いまやるべきことは、その際に必要な情報はなにかを考えて、情報の集積を図ることである。まさに EBPM の根幹となる元データ(量的なものだけでなく質的なものも含めて)を系統的に集積することがいま求められている。

国民が効果的に自粛・規制に付き合ってくれるかどうか、国民がなるほどと思える数値や見通しがあつてのことである。今回はその見通しが無いなか、国民は耐え忍んだ。しかしその我慢は限界に来ている。第2波へ備えるには、政府は方針決定にあたり国民と情報を時事共有し、国民が納得して自ら行動を起こせるように、開かれた意思決定過程を示すことが必要になる(専門家会議がこういいましたので従いました、では次回はやまらないうらう)。

ここでも EBPM の本質である「開かれた政策決定」がカギとなるであろうことを強調し、ひきつづき関係各部署におかれては、官民学の意味疎通・コミュニケーションのあり方を模索されることをお願いしたい。

コロナ禍のもとで試されたEBPM

- 初めてのことでばかりで、そもそもデータもエビデンスもないところで、EBPMのやりようがない？
- 急に「風」が吹いてきて決まった「事業」をEBPMみたいに見せるために後付けで数字を当てはめろと言うのか
- データを示せと言うが、個々の事業設計にあわせて統計取っているものではないし、そんな暇も人出もあるわけない
- 経済活動・社会活動・医療問題がトレードオフになっていて、なにを優先し、なにを評価指標にしているのかさえわからない
- そもそも「統計分析」で答えがでるのか？

改めて、そもそも論として

- エビデンスとは？
- ロジックとは？
- 因果関係とは？
- 「データ」「統計」「分析」とは？
- Evidence-based policy makingは何のため？

を愚直にふりかえってみる

Evidence-based medicine (EBM)

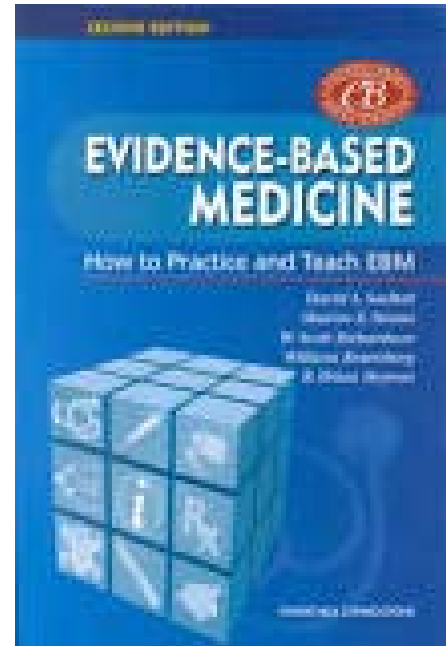
Sackett D. 1997

- “is the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients”

裁量のエビデンスを慎重・明瞭・思慮深い
意思決定のために用いる

- “Good doctors use both individual clinical expertise and the best available external evidence, and neither alone is enough”

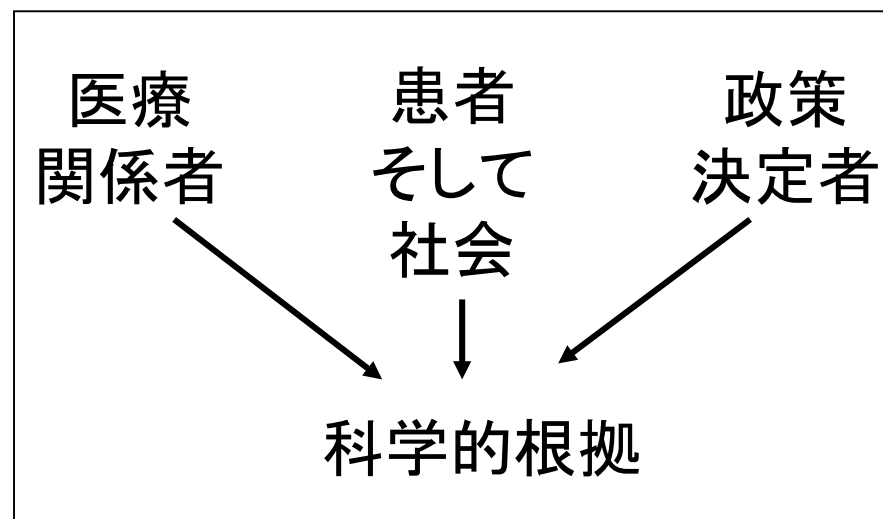
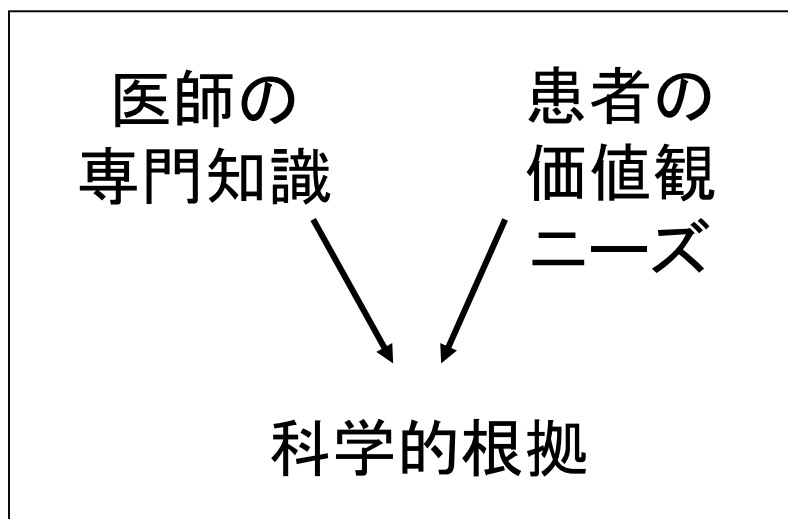
エビデンスと経験のいずれも必要



なぜエビデンスが必要か？ = 開かれた意思決定のため

Evidence Based Medicine (Sacket D) 1997

Evidence Based Healthcare (Gray M) 1997



意思決定の過程を合理的かつ開かれたものにする
ための「共通言語」としての「科学的手続き」

エビデンス

- 科学的手続き＝再現可能で開かれた手続きを経ている
- 意思決定のステークホルダーが、フェアな交渉に参加できることを保証するひとつの材料
- 数字で対処方法・決定がひとつに定まる(押し切れる)、というものではない
- 科学的評価＝誤ることがあることを前提、影響を最小限にとどめ、改善のための示唆を得る
- 既存のエビデンスにただ従うのではなく、実践評価によって修正・加筆されていく情報

EBMをめぐる批判と対応

- “Cook book medicine” 批判
マニュアル対応と、エビデンスによる開かれた
意思決定とは別物
- “情報過多”による実行可能性の問題
情報の系統的なライブラリー化＋情報技術で克服
(情報技術だけ、ではなく戦略的情報集積が重要)
- “数字情報”に頼り、価値・文脈(歴史ほか)・情緒などの実証
データで表せない情報の軽視
Evidence vs. Narrative based medicine論争→補完的關係
あくまで開かれた議論のための“材料”のひとつ
(ただし重要な)

どんなエビデンスを手に入れるのか？

理想的到達点

What we can do best?

理想条件下での効果

無作為化比較試験など実験研究



Efficacy (効能) としての

エビデンス

1980年代

実状

Where we are?

通常条件下での効果

アウトカムリサーチ

観察研究



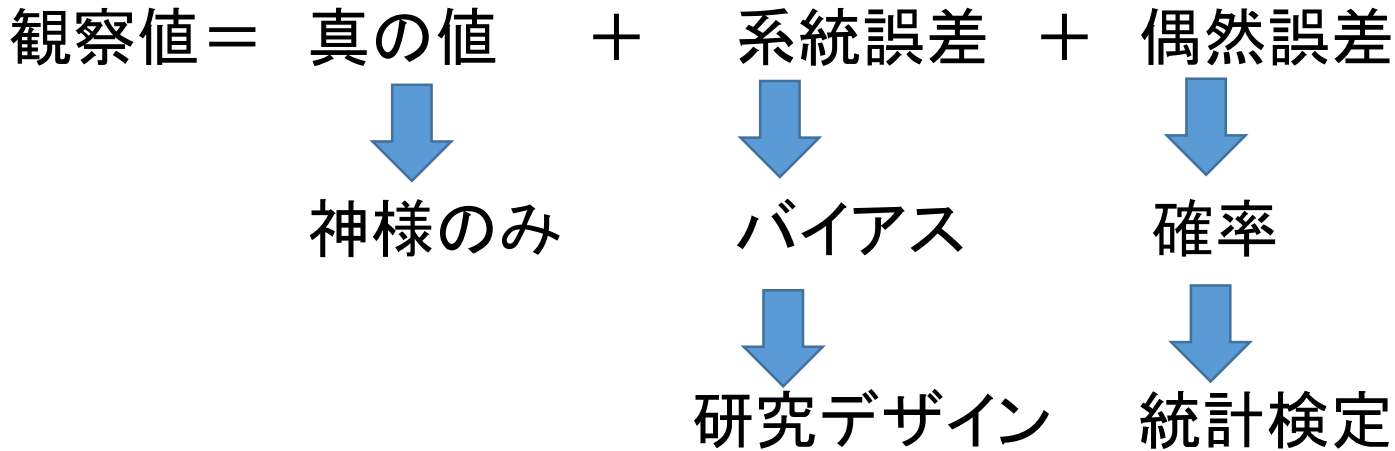
Effectiveness (効果) としての

エビデンス

1990年代

大規模患者登録などを利用したビッグデータ解析 (2010~)

単純認識論のもとでなぜ大規模・無作為化比較試験が 貴ばれるのか？



(もしくは因果推論的統計手法)

無作為化 = 系統誤差の影響をもっとも排除しやすい

大規模 = 統計検出力(“有意差”)を挙げるには標本数が大きいほうが安定

大規模 = バイアスが少ない・一般化できる、というわけではないので注意
(e.g. ビッグデータをAIに放り込めばなんでもわかる、わけではない)

無作為比較試験でないとわからないことは確かにある！

- Cardiac Arrhythmia Suppression Trial (CAST)
 - 1992 心不全の突然死予防のために、当時最新の抗不整脈剤(Class III)の無作為比較試験、予想に反して治療群で死亡率が高いことがわかり中断
- Hormone Replacement Therapy
 - 2002～3 更年期症状の軽減に加え、動脈硬化予防にも役立つといわれていたが、無作為比較臨床試験の結果、むしろ心疾患死亡率が高いことがわかり中断

理論的・基礎実験的に「効果」を期待されていたものが検証の結果「無効・有害」となったものは他にも

ただし無作為化比較試験にも弱点がある

- 時間や費用がかかる
- 無作為割り付けにまつわる倫理的問題
(社会実験がそうそう簡単にできないわけ)
例: Moving to opportunities for fair housing (MTO)
- 理想条件下での話-> real worldにあてはまるか？
 - 臨床治験と違い社会実験ではシンプルな結果は出にくい
 - アウトカムリサーチの台頭(1990年代)

余談;ビッグデータとAIで解決？

- AIの強み

- 膨大なデータのパターンから人では気づけないパターンを抽出することができる

- AIの弱み(というか性質)

- データの質に依存する

- AIに読ませる良質のデータを作成するのに膨大な作業が必要
- 読ませるデータの質が低いと、いくらたくさんあってもゴミ
- データに表現されている「現実」が理想的でない場合、ただ“bad practice”を忠実に再現するだけ
- AIは「なぜ」には答えてくれない。人では説明できないパターンが出てきたとき、説明責任をだれが果たすのか注意が必要

因果推論と論理(ロジック)

推論＝真実であるかどうかを議論する

そのお作法として異なる論理(logic)がある

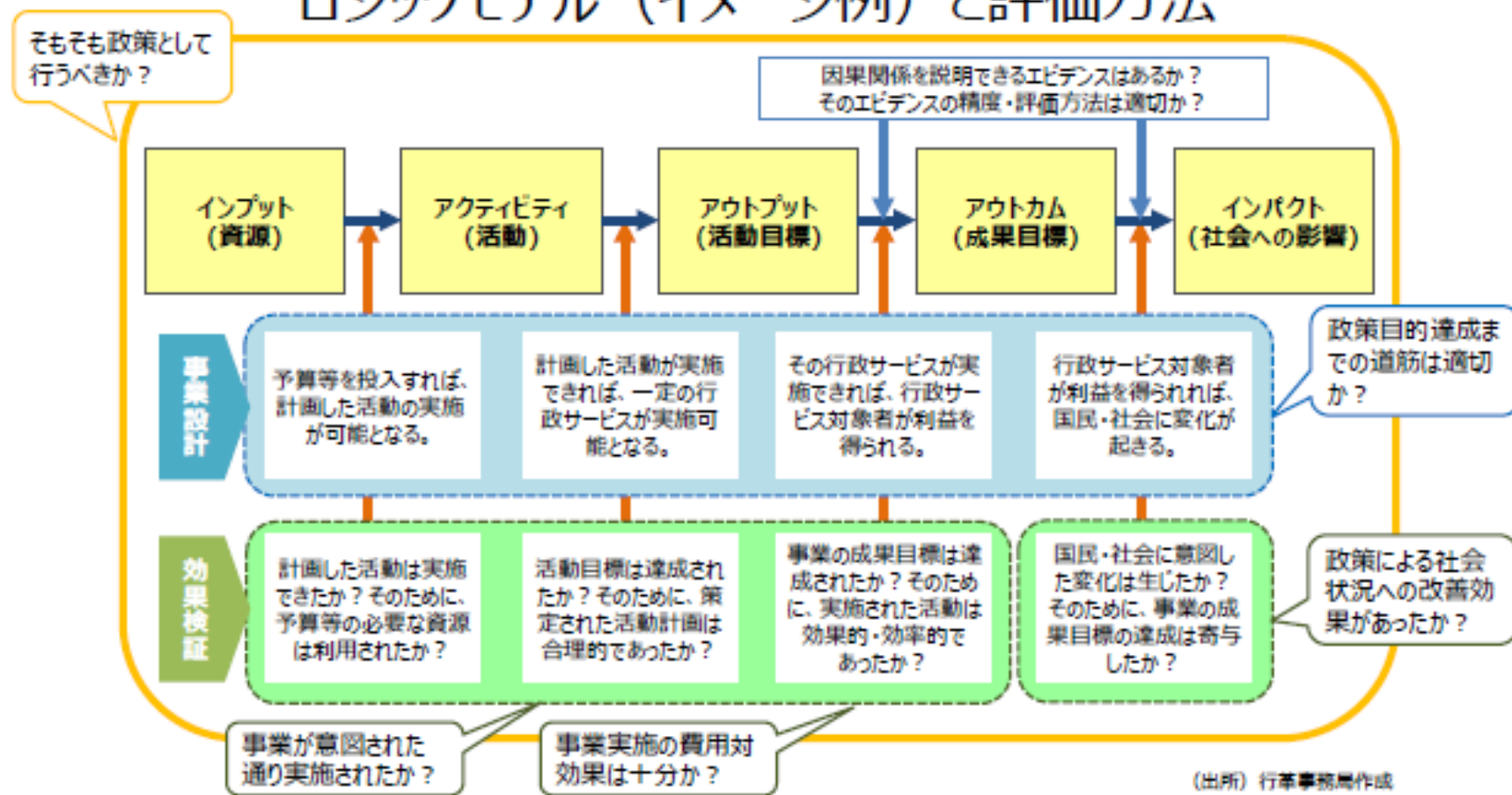
- 演繹(deductive) ⇒ 経済含む社会科学系
一般論(理論)から個別の因果論を導く
- 帰納(inductive) ⇒ 医学・工学など応用科学系
個別・特定の事象から一般的因果論を導く

では政策介入の評価は？

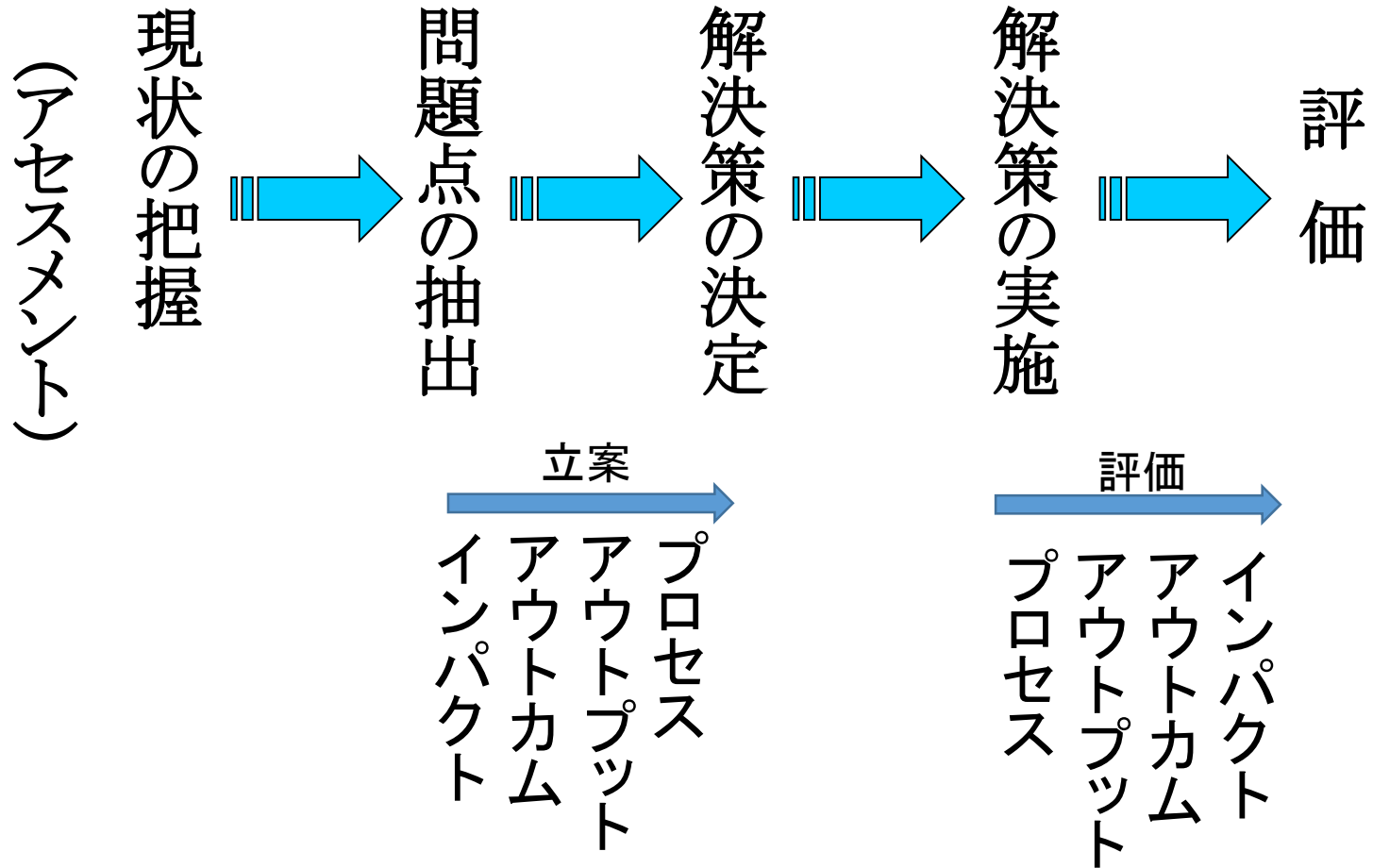
- 基本的にすべて実験(=政策的介入)
- その時代・時期・社会環境のもとでの、ある対象に対する施策の効果について議論
- 他の要因(社会経済状況、自然傾向、突発事故等)の影響下
 - 「科学的因果推論」は一般化できる因果関係を求めるのに対し、EBPMでは、「個別・特定」の施策・事業の効果を、現実社会の様々なノイズのもとで判断するための作業をする
 - 手法論として科学的な分析手法を使うこともあるが、それですべてというわけには行かない
 - 一番重要なのは、「なにを目的としてその政策・施策・事業を起こし、なにを期待していたのか」を説明できるかどうか

ロジックモデルの「ロジック」とは？

ロジックモデル（イメージ例）と評価方法



これとリンクしているのが



つないで輪にしたものがPDCA
もしくはProject Management Cycle

現状分析・問題特定

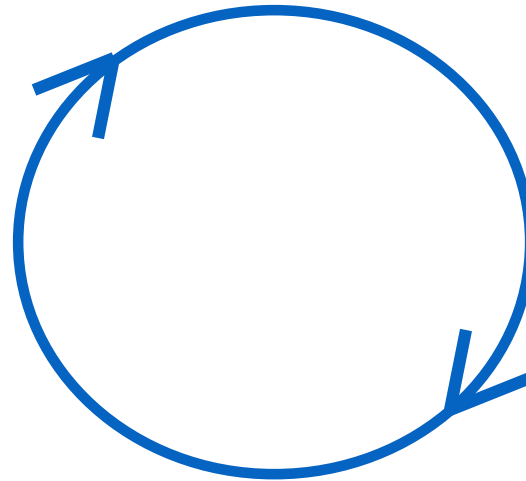
「現状」分析・「目標」設定

改善案提示

対処策作成・
選定

結果評価・
成果判定

実行・
過程評価



プログラム立案・評価の階層 (Rossi, Lipsey, Freeman, 2004)

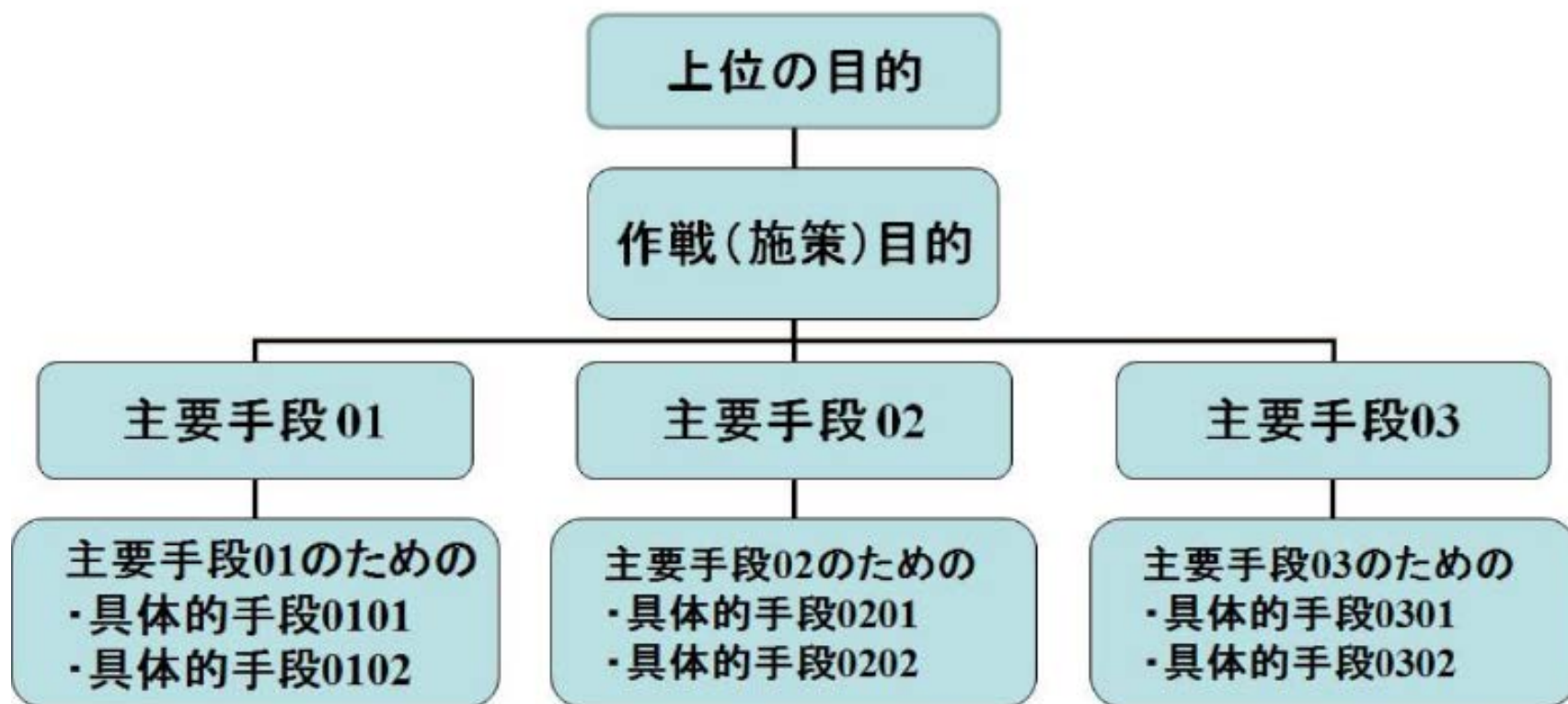
- ニーズアセスメント
- デザインと理論評価
- プロセス評価

記述統計主体

- アウトカムとインパクト評価
- 費用と効率性評価

分析統計要

ロジックモデルというともう一つ (ロジックツリー)

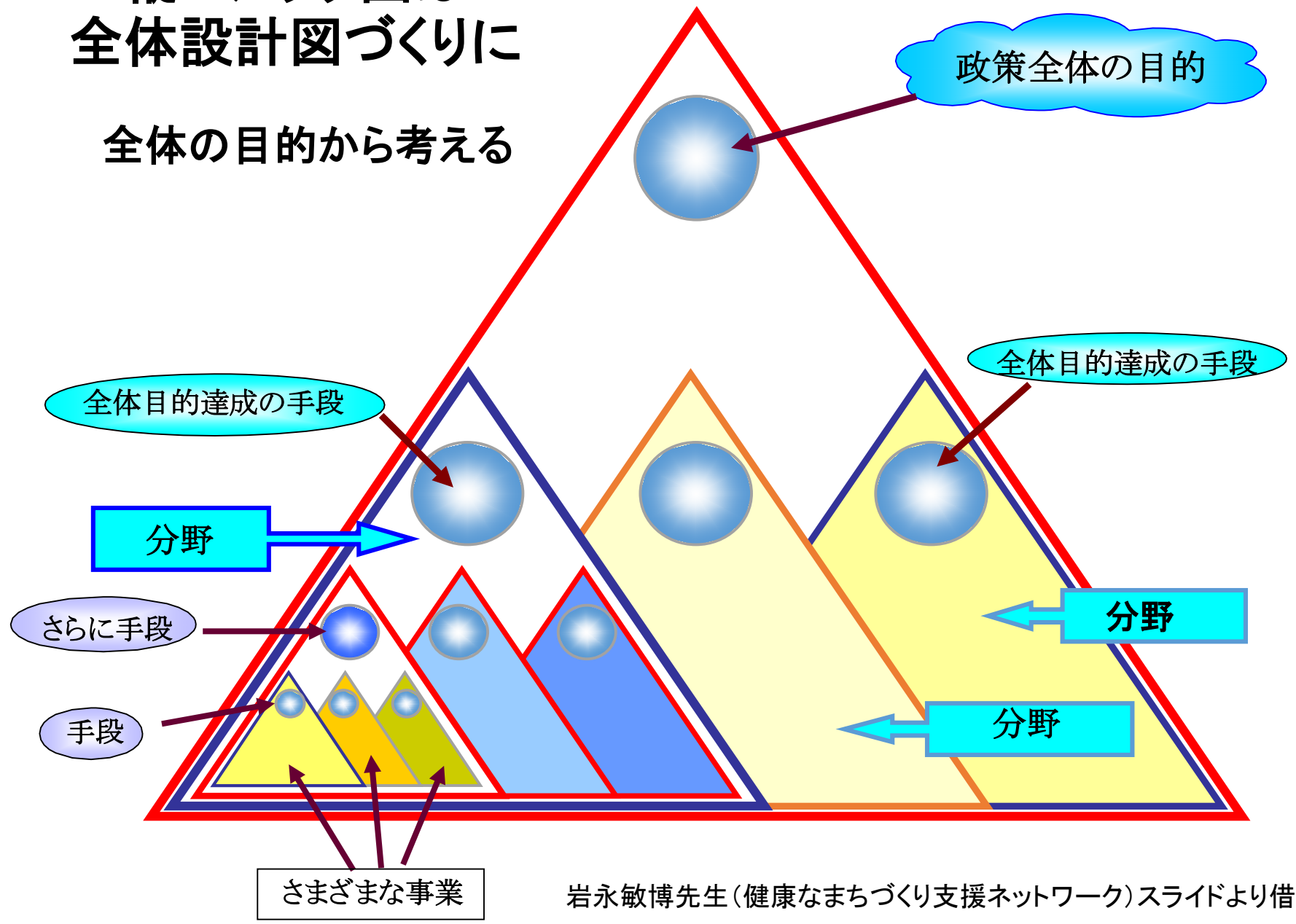


2つのロジックモデル図(横と縦)

- 横ロジック図は因果連鎖とその分析評価を意識
比較的因果関係が明確なものに適用
- 縦ロジック図は目的と手段の複合的連関を意識
上位目標を構成する下位目標群の連関
目標に到達するさまざまな手段の連関

縦ロジック図は 全体設計図づくりに

全体の目的から考える



現状分析と目標設定の2つのアプローチとも呼応

- 現状の「問題」はなにか？この「問題」をどう解決したいか？
 - 問題抽出・分析・解決アプローチ（経済分析型）
 - 「結果」としての問題
 - 「要因」としての問題
 - 両者の因果関係の同定が必要
- 現状はどこか、理想はどこか？理想に到達するにはなにを克服すべきか？
 - 目標指向・戦略的アプローチ（ビジネスマネジメント型）
 - 達成したい価値の同定
 - 動かせるリソース・機会の同定
 - 両者をつなぐ戦略・戦術のマッピング（順序・タイミング）

ふたたびコロナ禍のもとでのEBPM

- 特徴
 - 事前にエビデンスがあるわけではない
 - アウトカム(経済・健康・社会)が分野をまたいで連関
 - 比較的早いスピードで対応が必要(ただし試行錯誤が必要)
- 特徴を踏まえれば
 - 横ロジックよりも縦ロジックがまず重要
 - 戦略的にアウトカム設定し、短いサイクルでquick and dirtyで良いから成否をモニタリング
 - 「統計分析」評価よりも記述統計(場合によっては質的データでよい)でいいから、期待した成果が出ていないときに立ち戻って作戦を練り直せる情報を固める

要約

- ロジックモデルも、アプローチも、因果ロジックもひとつではない
- 「現状」分析、政策目標設定、プロセス評価、アウトカム評価それぞれでもちいるデータの性質、求める「統計値」、分析手法もさまざま
- つまりEBPMのやり方に「これがhow to EBPMです」というものは多分存在しない
- でもEBPMが求められる理由はなにか？
 - 政策に「絶対的正解」や「絶対的失敗」はない
 - 政策は多数多様のステークホルダーを抱えている事業
 - だから説明して納得してもらわなければならない

ゆえに、うざいけど必要とされている