

国際緊急援助の評価：活動の質の向上のために
－緊急援助の評価について－

○正木 朋也 (Tomoya MASAKI)

東京大学大学院 国際保健計画学

ここ数年わが国においても評価に対する関心が高まり、2000年9月には外務省他の支援のもとに日本評価学会が設立され、政府開発援助（ODA）や国内公共事業を対象に、広範な分野における実践的評価研究が始まっている。開発途上国においても国際機関等の協力により評価学会や協会設立の動きがあり、世界的にも援助を施す側と受ける側双方から評価を行う機運にある。この流れは、援助資金の拠出者（納税者）に対する説明責任のみならず、巨額の資金が被援助者に対し適切かつ有益に使われているかを客観的に評価し、将来の効果的な援助に資する重要な役割と目的を担っている。

その際、客観性も重要であるが、評価そのものの科学的妥当性がより強く要求されるべきである。特に、評価対象や背景が大きく異なる緊急状況下において、どの時点でどこまで適切な評価が可能であるのか、また、その後の報告においてその経験が正しく伝達され得るかという点については多くの議論が残されていると思われる。

ここで、多岐に渡る国際緊急援助をひとくくりに論じることは不可能であるが、評価の根底にある共通の枠組みとして、まず「科学的な評価」に関する基本事項とその問題点について話題を提供する。

具体的には、経済開発協力のような長期視点に立つ評価と緊急援助評価との違いおよび留意点を示す。例として、災害発生後の緊急事態における経時的動向（自然経過）

を踏まえた、介入効果判定の陥穽について指摘する。その際、現実には実施が不可能と考えられるランダム化比較の方法や比較対照の必要性を強調するが、これはあくまでも介入の真の影響を客観的に評価するために必要な科学的論理枠組みの再認識を目的とし、現実の対応については当日の議論を通じて糸口を見出したい。

ここではさらに「証拠に基づく」評価、言うなれば Evidence Based Evaluation (EBE) の実践的側面を念頭におき、評価結果が、評価専門家や研究者のみならず、現場の担当者や一般市民にもわかりやすく共有できるような仕組みの重要性についても述べたい。具体的には、報告書作成時における、構造化抄録作成の重要性を強調し提案する。さらに、データの収集方法と解析方法の開示、否定的なデータの取り扱いや収集したデータを具体的な形にする作業の進め方に問題がないか注意するといった、評価者自身の内省 (reflexivity) 的立場の重要性についても言及したい。

ここで述べる基本的な考え方や視点は、定量データを扱う場合のみならず、質的データを扱う際にも有用であり、これらを含めた総合評価を行う際の重要な示唆を与えられるものと考えられる。

以上の情勢、基本事項および各視点を共有した上で、他発表者らの緊急援助における評価の実際に照らし、そのあり方についての議論が深まることを期待したい。

**国際緊急援助の評価
活動の質の向上のために**

—緊急援助の評価について—

第17回日本国際保健医療学会大会
2002.8.2 神戸大学

話題提供

- 背景
- 「評価」とは?
- 科学的評価の枠組み(レビュー)
- 評価結果のフィードバック(経験の共有)
- 報告書と構造化抄録の重要性
- ガイドライン策定へ
- 補足事項等

2002.8.2 緊急援助の評価について 2

はじめに

- 日本評価学会設立と国際開発評価学会
- 評価すること、されること
- 評価を拒む62の理由
 - UNESCO/Internal Oversight Service: 62 (Good) Reasons for Avoiding Evaluation in the United Nations System.

2002.8.2 緊急援助の評価について 3

評価を拒む62の理由

1. 我々のプロジェクトは例外だ
2. コストがかかり過ぎる
3. 時間がない
4. プロジェクトの活動は極めて限られている
5. 業務計画書がない
6. 今まで一度もやったことがない
7. 行政や組織はそれを望んでいない
8. 資金がない
9. 自分の責任ではない
10. 評価する必要はない

2002.8.2 緊急援助の評価について 4

評価を拒む62の理由

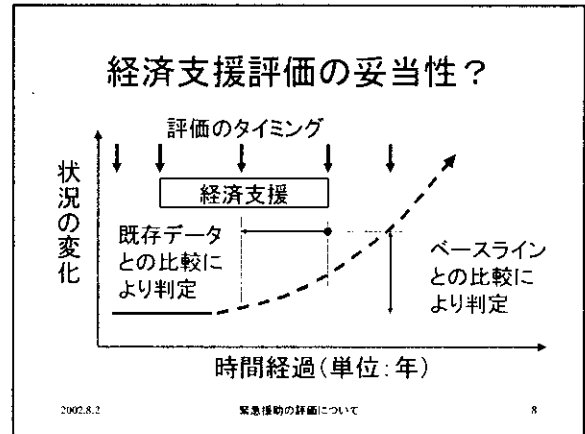
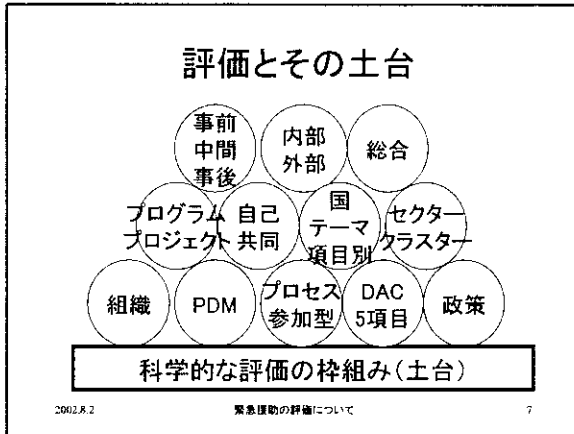
11. 論理が難解過ぎる
12. もっと現実をみよう
13. 我々の知ったことではない
14. 既に機能している、なぜ変えるのか?
15. まだ準備ができていない
16. 予算に含まれていない
- 17, 18, 19. 担当者が変わった、決まっていない
27. 「それは畏だ!」... 56. 雨期だから!...
62. 他の援助機関が同時に我々を評価したいと言うかもしれない

2002.8.2 緊急援助の評価について 5

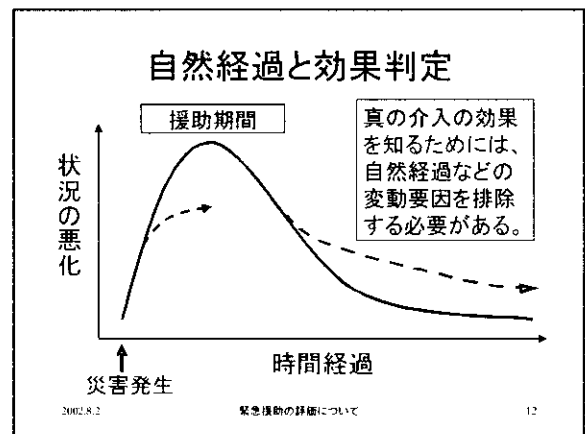
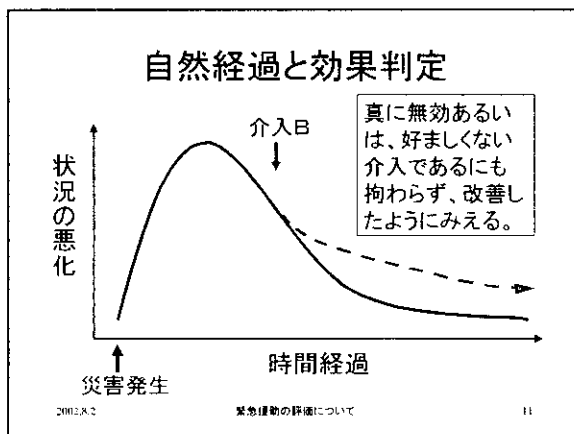
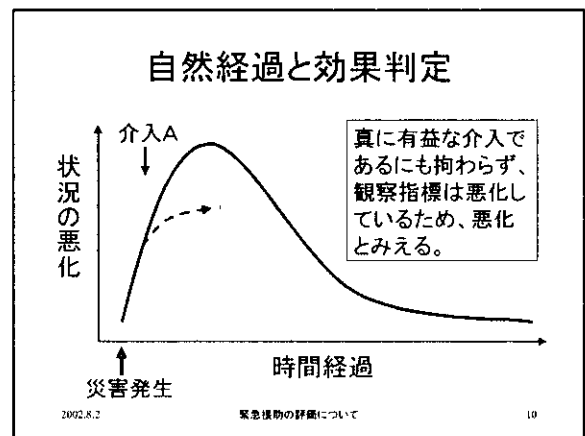
Types of evaluations

- Cluster evaluation(クラスター評価)
- Country program evaluation / assistance evaluation(国別プログラム評価/援助評価)
- Ex-ante evaluation(事前評価)
- Ex-post evaluation(事後評価)
- External evaluation(外部評価)
- Formative evaluation(形成的評価)
- Independent evaluation(独立評価)
- Internal evaluation(内部評価)
- Joint Evaluation(共同評価)
- Meta-evaluation(メタ評価)
- Mid-term evaluation(中間評価)
- Participatory evaluation(参加型評価)
- Process evaluation(プロセス評価)
- Program evaluation(プログラム評価)
- Project evaluation(プロジェクト評価)
- Review(レビュー)
- Risk analysis(リスク分析)
- Sector program evaluation(セクター・プログラム評価)
- Self-evaluation(自己評価)
- Summative evaluation(総括的評価)
- Thematic Evaluation(テーマ評価)

OECD/DAC (18-Oct-2001). Glossary of Evaluation and Results Based Management Terms. 2002.8.2 緊急援助の評価について 6



- ### 科学的評価の枠組み
- ・ 介入(援助)と結果の因果関係を評価する
 - 介入なきところに因果なし(質的評価においても同様)。
 - 有識者の評価(経験と勘) ≠ 科学的評価(証拠に基づいた判断)
 - ・ 自然経過等、他の要因による影響(バイアス、交絡因子)を排除し評価する
 - 介入による影響を(抽出し)判定する。
 - 「援助 = 良い介入」と考えるべきであるが、科学的な比較の枠組みがなければ、真の影響について、受益者、援助者、あるいは、評価者の誰も評価できない。
- 2002.8.2 緊急援助の評価について 9



緊急援助とその評価

- 評価目的の明確化
 - いつ評価するか?
 - どのような視点(立場)で評価するか?
 - 何をどのような手法で評価(判断)するか?
 - 真の影響を見るための唯一の方法はランダム化比較のデザインである。→ 緊急援助時の評価目的にも合致して実施可能か?
 - 目的をもった情報蓄積 → 事後解析、評価

2002.8.2 緊急援助の評価について 13

緊急援助とその評価

- 評価目的の明確化
 - 派遣時期と評価時期
 - 援助形態別の整理
 - 被災種別の整理
 - プロセス重視?
 - アウトプット～アウトカム?
 - さらに、インパクトまで…???
- これらは事前に明らかにしておく必要がある
 - ガイドライン策定

2002.8.2 緊急援助の評価について 14

評価結果のフィードバック

- 報告書の構成
 - ランダム化比較試験論文投稿のための CONSORT 声明 CONSORT (Consolidated Standards of Reporting Trials) statement
 - システマティック・レビュー/メタアナリシス論文投稿のための QUOROM 声明 QUOROM (Quality Of Reporting Of Meta-analyses) statement

2002.8.2 緊急援助の評価について 15

構造化抄録

- IMMRD (Introduction, Material and Method, Result and Discussion)
- Haynes RB, et al. *Ann. Intern. Med.* 1990;113(1):69-76.
 - 1)Objective 2)Design 3)Setting
 - 4)Patients or Other Participants
 - 5)Intervention(s)
 - 6)Main Outcome Measure(s) 7)Results
 - 8)Conclusion(s)

2002.8.2 緊急援助の評価について 16

構造化抄録

診療ガイドラインの作成の手順 ver.4.3(2001.11.7)

- 1次研究
 - 1)目的 2)研究デザイン 3)研究施設 4)対象患者
 - 5)介入 6)主要評価項目とそれに用いた統計学的手法
 - 7)結果 8)結論
- システマティック・レビューないしメタアナリシスの論文
 - 1)目的 2)データソース 3)研究の選択
 - 4)データ抽出と質の評価 5)主な結果 6)結論

2002.8.2 緊急援助の評価について 17

質と量のバランス

重要な評価指標については、評価の目的に合わせ、可能な限り計測する姿勢

2002.8.2 緊急援助の評価について 18

評価者の姿勢

- 内省的姿勢
- 科学性と倫理性のバランス
- データの収集方法と解析方法の開示
- 否定的データの取り扱い方
- 解析結果の解釈と評価結果の客観性
- 結果、報告書の公表

2002.8.2 緊急援助の評価について 19

まとめ

- 評価目的の明確化
- 科学的評価の枠組み(レビュー)
- 報告書作成と構造化抄録の重要性
- 継続的な情報の蓄積と評価体制
 - 証拠に基づいた判断とフィードバック
- ガイドライン整備の必要性
- 評価者の科学的・倫理的、内省的姿勢

2002.8.2 緊急援助の評価について 20

参考資料等

1. UNESCO/Internal Oversight Service: 62 (Good) Reasons for Avoiding Evaluation in the United Nations System. http://www.unesco.org/ios/eng/evaluation/tools/eutil_11e.htm
2. 津谷喜一郎, 小島千枝, 中山健夫, 訳. CONSORT声明: ランダム化並行群間比較試験報告の質向上のための改訂版勧告. JAMA<日本語版> 2002年6月号: 118-24. (チェックリスト: http://www.sphere.ad.jp/conf/CONSORT_Statement/menu.html)
3. 診療ガイドラインの作成の手順 ver. 4.3 2001.11.7. <http://www.iph.go.jp/gigl-4.3rev.htm>
4. Haynes RB, Mulrow CD, Huth EJ, Altman DG and Gardner MJ. More Informative Abstracts Revisited. *Ann. Intern. Med.* 1990;113(1):69-76.
5. 青木 仕. 構造化抄録の基礎知識. 中嶋宏(監修), 津谷喜一郎, 他(編). EBMのための情報戦略. 中外医学社, 2000; 82-93
6. 青木 仕. Structured Abstracts の概要と我が国における活用の可能性. *医学図書館* 2000;47(1):52-60

2002.8.2 緊急援助の評価について 21

国際緊急援助の評価 活動の質の向上のために — 緊急援助の評価について —

第17回日本国際保健医療学会大会
2002.8.1-3 神戸大学

正木朋也
東京大学大学院国際保健計画学
Department of Health Policy and Planning,
The University of Tokyo

Annex 6

リスク社会の指標と測定
日本犯罪社会学会第 29 回大会
東京, 2002.10.26 (土)
静岡県立大学国際関係学部
津富 宏

リスク社会の指標と測定

津富 宏（静岡県立大学）

1 リスクアセスメントの始まり

犯罪者処遇における「リスクアセスメント」は、犯罪を犯す見込み（確率）をリスクと呼ぶ予測研究に端を発する。その嚆矢は、1920年代に始まるグリュック夫妻の一連の予測研究で、第2次大戦以降、保護観察官による利用を目途とした予測尺度の開発が始まり本格化した。チェックした数の合計を求めるタイプの予測尺度が開発されたが、1970年代に行われた実証研究は、当時の予測尺度の有用性を否定した。例えば、予測得点によって caseload を操作したところ、intensive supervision が再犯の防止に有効であるどころか逆効果であることが見出され、また、予測尺度の妥当性・信頼性も非常に低く実用に耐えないという知見も定着した。にもかかわらず、予測尺度は、再犯確率に応じて監視の程度を設定すべきであるという要請に屈服した行政当局によって引続き利用された。

2 カナダのリスクアセスメント

1980年代後半から、カナダのオンタリオ州を中心とする研究者が連邦矯正局と連携し、ブレークスルーを実現した。彼らは、多領域の予測因子を盛り込むことにより、精度を上げ、客観尺度に基づく

予測の精度が、臨床的な判断に基づく予測の精度を上回ることを見出した。彼らのリスクアセスメントは多変量解析に基づいて領域ごとにチェックする項目数を設定したチェックリストによる。彼らのもっとも大きな功績は、リスク概念を犯罪確率から予測要因に拡張し、変容しえない要因 static risk と、社会的学習理論から導かれた変容しうる要因 dynamic risk を概念化したことである。こうして、予測と処遇は、原因論を通じて結びつけられることになり、処遇実験による dynamic risk の確定が始まった。処遇実験を通じて、施設内処遇は社会内処遇に比べて dynamic risk の変容量が劣るため再犯率が高いことが見出されたため、連邦矯正局は施設収容の減少に向かって10年間にわたりキャンペーンを行い、収容人員を押さえ込んできた。リスクの把握が合理的な刑事政策へとつながる好例である。

3 Risk factor prevention paradigm におけるリスクアセスメント

1980年代後半から行われてきたワシントン大学の社会開発研究所の実践をもとに、1990年代後半から risk factor prevention paradigm が唱えられた、こ

これは公衆衛生学を非行予防研究に持ち込み、地域ごとに同定された risk factor を改善しようというパラダイムである。残念ながら、このパラダイムは、a-theoretical であろうとする疫学のスタンスを継承したため risk factor を分化して捉えるには至っておらず、また、学校やコミュニティを単位とする介入を原則としたため ecological fallacy も疑われる。このパラダイムにおけるリスクアセスメントは、ある地域について、先行研究においてリスク要因として知られている項目について値を測定し、その値がその地域を含むエリアの平均値と比べて、有意に高ければ介入の対象とするというものである。つまり、多変量解析に基づく予測尺度を欠くため、恣意的に設定されたベンチマークとの比較を行うにすぎない。従属変数と独立変数を結びつける係数の大きさを明確にしないまま、介入を行うという論理的に破綻した構造である。

4 リスク概念の発展

リスク概念は犯罪を統制するための計量化の道具として、統制技術の進歩に伴って発展しつつある。これをまとめると以下ようになる。

予測研究： 予測研究においては、リスクは、犯罪確率の推定値、すなわち、従属変数の推定値として定義される。この

場合、リスクを推定できたとしても、減少させる方法を伴わなければ有用でない。そこで、犯罪確率の分散のうち、統制可能な部分とそうでない部分が区別されるようになる。これが、アンドリュースの dynamic risk である。

予測研究から実験研究へ： dynamic risk の効果の大きさは、実験研究によって確定されるが、これは、犯罪確率を予測する要因の、犯罪確率に対する寄与の大きさである。つまり、dynamic risk の大きさは、統制側の選択肢の有効性の大きさを意味し、予測研究における「リスク」と比べて格段に有用な概念である。

実験研究からメタ実験研究へ： 実験研究のメタ・アナリシス（メタ実験研究）が行われるにつれ、処遇効果の推定値の大きさの分散（つまり、標準誤差）への着目がされ、同じ期待値であればより分散の小さい処遇を選択するという議論がなされるようになった。これは、投資においてハイリスク・ハイリターンというときのリスクである（期待値の高い投資は期待値の分散が大きい）。処遇効果の標準誤差については、「リスク」の語は当てられていないが、統制という意図的行為（リスク・テッキング）の不確かさを表す概念である。最近行われるようになってきた、費用便益分析も、意図的行為の「副作用」に着目するもので、このような進

展を見ていると、犯罪者処遇における「リスク」概念も、他の分野における「リスク」概念に近づいてきたように思われる。

リスク社会の指標と測定

静岡県立大学
津 富 宏

リスクアセスメントの目的

- リスクアセスメントの目的は、
犯罪行動の統制
- 犯罪統制の不確実性を減少させ、
効果をあげる道具

マネジメントの道具としての リスクアセスメント

- 第1世代: 予測研究
 - リスクの確定
 - 予測因子の確定 → 予測尺度の開発
- 第2世代: 予測研究から実験研究へ
 - Manageable risk と unmanageable risk の分離
 - 単なる予測因子と処遇対象因子の分離
- 第3世代: 実験研究からメタ実験研究へ
 - 統制に伴うリスクの算定へ

リスクアセスメントの基本式

- $y = f(x)$ 回帰式なら $y = \sum \beta_i x_i$
 - y: 犯罪 x: 予測要因
- 「f」は、通常一般化線形モデルを用いる。関数形についての研究は実用レベルでは不十分
 - 回帰分析 → ロジット分析 → 生存時間分析
 - * 生存時間分析においては、本人の「死」をリスクと定義するのでニュアンスが異なる

リスクアセスメント: 第一世代

- $y = f(x)$ y: 犯罪 x: 予測要因
 - 犯罪の分散を、予測可能な分散と誤差分散に分離
 - リスクは再犯確率を意味する
- グリュック夫妻に始まり、Wisconsin Scaleなどを生み出す
- 1970年代の研究によれば
 - 予測得点によって監視程度を変えても犯罪は減らない
 - 予測妥当性も小さく、数年間たつと失われる
- しかし、実務において使われ続ける
- 予測しているだけで、なんら犯罪を統制していない

リスクアセスメント: 第二世代

- $y = f(x_1, x_2)$ 回帰式であれば $y = \sum \beta_{1i} x_{1i} + \sum \beta_{2j} x_{2j}$
 - y: 犯罪 x_1 : 操作不可能な要因 x_2 : 操作可能な要因
 - 予測可能な分散を、統制不可能な分散と可能な分散に分離
 - 予測研究から因果研究へ(操作可能性への着目)
 - リスクはリスク要因を意味する
- カナダのアンドリュースらのグループが連邦矯正局とともに予測尺度を開発 多領域化による予測精度の向上
- static risk と dynamic risk (人格障害をきむ) の区別
 - 後者をターゲットにした処遇実験による x_2 の確定
 - 統制の効果の推定値 β_2 への着目
- 社会内処遇の有効性をもとに施設内処遇を避けるキャンペーンを展開
- 個別犯罪者の統制技術であったリスクアセスメントが、イデオロギー論争を離れた、刑事政策の判断基準を提供

犯罪学におけるリスク概念の矮小化

- 犯罪原因論=risk taking behavior
 - Cohen: 中流階級の価値への反発
 - Greenberg: 「従順な労働者」的価値への反発
 - Hagan: 資本主義の本質としての非行
- 合理的選択理論におけるrisk
 - Risk discounting
- Routine activity theory におけるrisk
 - Risk of victimization
- 「p」(再犯)と「1-p」(非再犯)は同値なのに、なぜ、確率を「リスク」と言い換えるのか(研究者自身が価値中立でないことに無自覚)

Risk communication

- 事例1 risk manager としての立場の主張
 - 認知行動療法家の一部は、relapse prevention approachの有効性を主張 risk はmanageし続けないとコントロールできないというもの。自らをrisk managerの位置におき、professional interestに奉仕するだけの論理。
 - セルフ・ヘルプグループが再犯リスクの減少に有効であるとの主張
- 事例2 risk fixationによる支持の動員
 - 性犯罪者は再犯リスクが高く、矯正できないという言説 cf. メーガン法

Risk communication

- 事例3 risk fixationに対する対抗
 - カナダ連邦矯正局は、riskという用語の使用をやめ、reintegration potentialと言い換え
 - Riskが可変であり、manageableであるという connotation
- 明快なエビデンスがあるのは、事例3の主張のみ。
- 政治的立場にかかわらず、再犯「リスク」をあえて正しく伝えない社会 stakeholderがリスク概念を濫用
- それぞれの主張に伴う「リスク」(副作用)を改めて測定し、情報公開する必要性

リスクアセスメントの測定単位

- リスクアセスメントは必要か
 - 癌の検診を個別に行うよりも、社会全体で禁煙運動をしたほうが癌の罹患率は下がる
- これは、測定単位(レベル)の問題
 - 犯罪者処遇におけるリスクは、(第二世代までは)個人レベルで想定されてきた
 - 個人レベルのリスクアセスメントは、システム自体がもたらす副作用(過剰収容、刑務所収容に伴う家族のコスト、地域社会のコストなど)の見落としにつながっている

結論: 第三世代のリスクアセスメントへ

- 犯罪学における「リスク」=「確率」概念の問い直し → より豊かな(たぶん、システム的な)「リスク」概念へ
- 「リスク」=「確率」概念に伴う Risk communicationの濫用 → 「リスク」(=副作用)の測定
- 個人レベルのリスク定義の問題性 → 「リスク」のシステムレベルでの定義

結論: 第三世代のリスクアセスメントへ

- 「リスク」定義自体の変更
 - 統制の確実性(統制効果の標準誤差)の測定
 - 統制の副作用の測定
- 真の意味で、take risk
- はじめて、システムレベルでの、選択肢としての統制の選択に寄与

Annex 7.

Social Experiment と科学的評価の基本枠組み
日本評価学会第3回全国大会
東京, 2002.12.7 (土)
東京大学大学院医学系研究科
国際保健計画学 正木朋也

Social Experiment と科学的評価の基本枠組み

Social experiment and basic framework of the scientific evaluation

○ 正木朋也
Tomoya MASAKI

東京大学大学院国際保健計画学
Department of Health Policy and Planning, The University of Tokyo

<要約>

事業評価や政策決定において科学的評価の方法論は強力な判断基盤を提供する。科学的枠組みに基づく評価は、その透明性と説明責任を担保するのみならず、事業や施策の施行者自身がエビデンス（科学的方法論により得られた確実性の高い証拠）に基づいた最善の対応を行っていることを確信するうえでも重要である。近年、諸外国において数多くの社会実験の報告がなされているが、わが国においては未だにエビデンスに基づく事業や政策の評価が実施されているとは言い難い現状である。本稿では、社会実験の有用性と実施可能性に関する議論を深めることを目的とし、科学的評価の基本枠組みについて、いくつかの例をもとに一般的に行われている評価の問題点を指摘する。また、科学的評価の実践的側面として、エビデンスを「つくり、つたえ、つかう」ための具体的方策等について、先行事例を参考に検討を加えた。

<キーワード> 介入、因果関係、エビデンス、構造化抄録、ガイドライン

はじめに

社会実験の目的は、介入（intervention）と結果の因果関係を明らかにすることである。分野領域を問わず、介入の影響を把握することは評価の基本である。即ち、制御不能な未知のバイアスや交絡因子を排除し、介入による真の影響のみを正しく読みとって判断することが肝要である。その方法論として、自然科学分野で一般的に用いられる手法「ランダム化比較試験（RCT: Randomized Controlled Trial）」があり、またその手法から得られた結果を質の高いエビデンスとして統合評価（メタアナリシス）するといった、いわゆる「証拠に基づいた評価（EBE: Evidence Based Evaluation）」の考え方が生まれる。この考え方は、1990年代初頭よりEBM（Evidence Based Medicine / Health Care）として保健医療分野において普及し、わが国においても既に保健医療介入の評価とその実践的な応用の基盤として根付いている。臨床研究や疫学調査の結果等を蓄積し相互に参照することにより、

現在もなお数々のリスクが特定され、健康科学上の発展に実質的に貢献している。

翻って、政府開発援助や国内公共投資の評価をみると、わが国においては殆ど応用例や報告がみあたらない。倫理面、社会的、経済的な側面を考えれば、解決すべき問題が多々あることは容易に推察できるが、これらの問題は元来保健医療分野においても同様であった。今回別途報告されるように、諸外国では、既に数多くの社会実験が実施されている。得られたエビデンスをもとに、意義ある介入のみを実施し、無効あるいは害のある介入を是正排除し、事業や政策に反映させていくことは今後益々重要となる。

1. 科学的評価の基本枠組み

ある対象に対し何らかの事業や施策を施す場合、そこで得られた結果と施した事業や政策との因果関係を明確にすべきである。これら事業や施策は対象集団に対する介入と見なすことができ、観察された結果がその介入による影響かどうかを判定しなければならない。介入の無いところに因果は生じない。因果関係を明らかにするためには科学的な比較検討の枠組みが必要となる。

経済支援評価の妥当性？

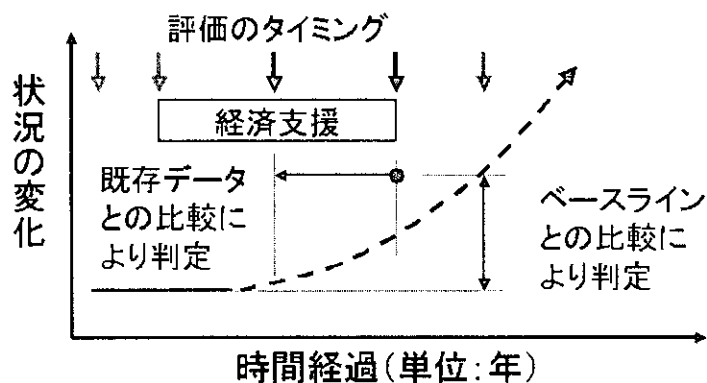


図1. 経済支援評価枠組みへの疑問

図1は、経済支援の評価について模式的に示したものである。開発介入（この場合は経済支援）の効果を判定するために一般的に行われるのがベースラインとの比較である。実際には、様々な理由によりベースラインを把握できない場合があり、得られる範囲の過去の値をベースラインの変わりとした評価も多い。この経時的変化を比較することは一見合理的に思えるが、果たしてこれで開発介入の効果を判定できるのであろうか？ 開発介入

* 科学的評価の立場では、あらゆる事業や施策は「介入(intervention)」とみなす。

が現状を改善したのか、例えば、グローバルな社会的・経済的發展により観測した指標が改善したように見えているのか？ 逆に、気候変動や政変等による負の要因があったにも拘わらず、開発介入が功を奏し観測指標が改善した結果であるかもしれない。あるいは、開発介入と指標の改善は全く無関係であるという可能性もある。自然経過や他の因子による変動の可能性を完全に排除しない限り、このような単純な前後比較から介入そのものの効果判定はできないことが明白である。

この経済支援を例とした模式図では、暗黙の了解として、介入以外の全ての状況が不変ないしは等しい (*ceteris paribus*) ことを前提とすることに問題がある。即ち、未知既知を問わず制御不能あるいは予測不能な要因が観測指標に影響している可能性は捨てきれず、上述のような単純な前後比較の枠組みだけでは、科学的な根拠をもって経済支援が奏功したことを証明することはできないのである。

次いで、自然変動の影響が効果判定に矛盾をもたらす例を模式的に示す(図2)。図は災害発生時の救援活動を例にしたもので、縦軸に状況の悪化、横軸に時間経過をとり、実線の推移は本来我々が知ることのできない「自然経過」を示し、破線は介入の結果、観測値として我々が認知可能な指標の動きを示している。この「自然経過」は反事実的状況 (counterfactual) と考えても良い。具体的には、介入を行わなかった場合の下痢症や感染症の発現頻度などを思い浮かべれば理解が容易である。この例では、災害発生直後より被災地の衛生環境が悪化するためこれら疾患が急増するが、ある期間後のピークを越えた時点で次第に減少の方向に転じている。これは模式的に想定した「自然経過」であって、実際には援助者側も被援助者側あるいは第三者の何れもが知ることができない経過であり、現実には必ずしも模式図のような単純な推移をすとも限らない(いわば神のみぞ知る自然変動である)。

さて、災害発生後まもなく緊急援助隊が到着し救援活動を行う。実際に各疾病の発現頻度を減じてはいるのだが、ピーク前の状況悪化の勢いに押され、その救援による効果がうち消されてしまう。結果的に観測可能な指標で前後比較を行うと過去の経験から最適と考えられる救援活動を行っているにも拘わらず、事態が悪化しているように見える。次に、ピークを越えたところで、緊急援助隊が任務期間を終え撤退する状況を想定する。その際、撤退時の値とその後の予後観測値を比べれば、指標の改善が見られるため、緊急援助活動の効果が持続的に奏功しているかのように見える。

しかしながらここで、本来知ることのできない「自然経過」を踏まえ(神様になったつもりで)上記の観察指標についての評価を行えば、結果が逆転することに気づく。即ち、ピーク前の救援活動に関しては、観察される指標(例えば疾病発生頻度)が前後比較により悪化しているように見えるが、実際には状況の悪化を抑制しており、救援活動が「真に意義ある活動である」ことがわかる。また、ピーク後については、確かに観測指標は徐々に状況改善しているが、自然経過に比べると緩解を遅らせている。即ち、援助隊撤退後は、(何らかの理由で)自然経過よりも「真に事態が悪化している」ことがわかる。

自然経過と効果判定

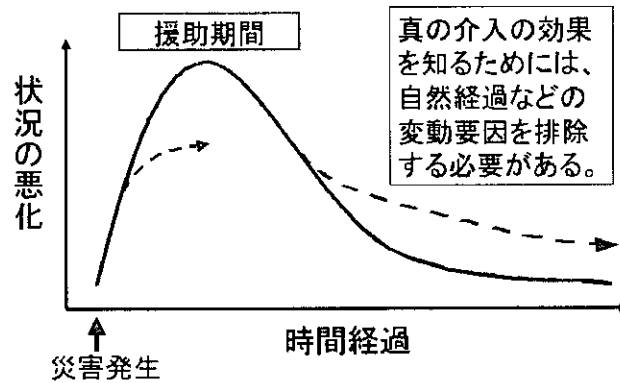


図2. 自然変動がもたらす効果判定の矛盾

上述の例は、あくまでも模式的に事例を示したものであるが、重要な点はこれらの相反する状況が生じていたとしても、一般に行われている単純な前後比較だけでは、誰にも真の介入による影響を知ることができないことである。実際に評価者が知りたい真の介入効果は、文字通り「神のみぞ知る」状況にあり、これでは、いつになっても適切な援助支援活動を実施したことを立証できない。真に最適な活動を行っているにも拘わらず、そのことが正しく評価されない状況が生じ、あるいは、(あってはならないことであるが)良かれと思って施している援助や事業が逆に対象の状況を悪化させているような場合にも、その事実を把握できない事態も生じ、援助側被援助側双方にとって不本意な状況が生じ得る。

基本的な評価枠組み上の問題点は、個々の評価においてどのように認知され対応されているであろうか？ そもそも評価者は介入と結果の因果関係をどのような方法で知り評価に反映し得るのであるか？ このような問題は日常生活の中でも矛盾をもって体験される。例えば、図2の災害発生の様式図を、自分が風邪をひいた場合に置き換えてみるとこの問題の本質がより直感的に実感されるかもしれない。ピークまでの急性期に頓服した風邪ぐすりが「真に有効」なものであったとしても、症状が悪化したと実感(錯覚)した本人としては無効と感じ、更には好ましくないと判定するかもしれない。一方、緩解期において「真に無効あるいは有害」な治療が施されたとしても、結果的に症状が改善していると認識(錯覚)すれば、その治療が有効であると感じてしまうことになる。診察している医師にも患者自身にも真の状況が分からず、災害援助の場合と同様に評価に矛盾を生じることがわかる。即ち、この例から得られる教訓は、真の介入効果について、当事者を含め誰も正しく評価判定できていないことである。

これらの事象を想定すれば、今後可能な場合には、積極的に科学的な評価枠組みを取り入れる必要があることが理解される。何をよりどころに評価判定を行うかは、本来、科学

的評価の枠組を踏まえ、事前に十分議論しておく必要がある。

状況を正しく判定する科学的手法としてランダム化比較の方法があるが、各領域における現実的な実施可能性を考えれば、必ずしもその適用が強調されるべきではない。実際には、実施可能性を優先し、利用するエビデンスのレベル**を下げることで対応せざるを得ないものと考えられる。評価指標が何であるにせよ、日常的に実施している多くの評価が、実は科学的な評価に値しない枠組みの中で行われていること - 評価者自身が意図せず誤った評価判定をする危険性があること - を真摯に受け止めこれを改善する努力を払うべきである。

2. 評価経験共有のための基盤整備

いかなる評価においても、経験の蓄積と利用は重要な課題であり、単に過去事例を評価報告書として記録・保管するだけでは不十分である。科学的な枠組みにより得られたエビデンスは無駄なく蓄積し、正しく評価した経験を集めて共有・伝達する基盤が必要となる。EBMで言うところの「つくる、つたえる、つかう」というエビデンスの流れの「つたえる」がこれに当たる。既に、保健医療分野では世界的なネットワークとして1992年から実施のコクラン共同計画(CC: Cochrane Collaboration)があり、また、教育・政治・社会学全般の領域においては2000年に発足したキャンベル共同計画(C2: Campbell Collaboration)がある。このような自然科学の方法論に則した評価基盤の整備は、情報の電子化により今後益々加速するものと思われる。インターネットの普及に伴いさらに拍車がかかるものと思われるが、その際重要なことは、単に得られた知見を広く公開することではなく、批判的な吟味を受けたうえでなおかつ強いエビデンスを示す結果に、利用者が的確にたどり着ける仕組みを構築し提供することである。

これまでのところ、わが国における国内外の公共投資や各所で実施される事業等の評価には(仮に実施可能な場合であっても)実験科学的手法は適用されておらず、結果として説得力のあるエビデンスを見出すことなく事業評価や政策決定がなされているように見受けられる。評価判断に対し批判的あるいは建設的な意見が出されることはあっても、科学的な論拠に基づく議論ができないため、具体的かつ発展的な解決策に到達しづらい状況も想像される。またそのような状況下では、選りすぐったエビデンスを適切に蓄積し再利用する仕組みも醸成されにくいものと思われる。

このような現状認識のもとで、今後は内省的姿勢にて、時間のかかる基盤整備も念頭におき、先行する他の分野や諸外国の事例に学び、より積極的に科学的な評価枠組みを取り入れるべきであろう。例えば、手始めに、評価報告書およびその抄録を構造化し、第三者

** エビデンスのレベル: EBMを参考とすれば、一事例研究 < 複数事例研究 < ケース・コントロール研究 < コホート研究 < ランダム化比較研究 < ダブル・ブラインド・ランダム化比較研究 < メタアナリシス、の順にエビデンスが高くなる。因みに、エビデンスのレベルという意味からは有識者の意見や考えなどは、これらの何れよりも低く見なされる。

による効率的な再評価を容易にする仕組みを作ることや、そのためのガイドラインの整備に着手することを提案する。これら整備により、実施面での質の向上が期待される。

3. 他の評価研究関連分野との関係等

本稿では、わが国の評価研究における現状を踏まえ、自然科学的な方法論や考え方をやや強調過ぎる形となっているが、これは本来意図するところではない。これらの考え方は、全ての評価の根底に位置する基本事項であり、その考え方は、既存の各種評価手法と競合したり齟齬を来すものではない。また、EBE が主として扱う定量的データは、対象集団の価値観や社会的、文化的、歴史的な文脈を踏まえたうえで解釈されて初めて意味を持つものである。科学的枠組みに基づく評価を行う場合には、当然こうした定性的側面についての十分な配慮とバランスの良い評価判断が要求される。

さらに、科学的評価を適時に、かつ実践的に政策等に反映させるための枠組みとしてリスク分析がある。想定されるリスクを最低限に押さえ、社会実験等により得られた判断要素をより適切に現実の社会に適用させることが目的となる。いわば科学的評価とリスク分析は車の両輪に例えることができ、双方のアプローチが科学的かつ相補的に機能することによって、より適切で効果的な事業方針や政策決定が可能となる。これらの枠組みはEBMと疫学研究に基づいたリスク分析に基づく公衆衛生学的なアプローチとの関係にも通じ、本年設立された予防医学リスクマネジメント学会他、関連領域の動向にも着目する必要がある。

おわりに

評価研究におけるわが国の現状を鑑み、科学的評価とは何かという基本に立ち返って現行の評価枠組みの問題点を明らかにした。また、評価の透明性と説明責任を担保するエビデンスを蓄積・伝達するための基盤整備の必要性を述べ、今後の方策を提案した。今後はさらに、全ての評価研究活動の背景に根ざす評価理念をも踏まえ、社会実験が有効となる領域と今後の評価研究の方向性についての議論を深めたい。

参考文献

1. わかりやすい EBM 講座, 厚生省健康政策局研究開発振興課医療技術情報推進室(監修). 厚生科学研究所 2000.5.23.
2. Anonymous. Try it and See. *The Economist*, 2 March 2002: 71-2 (原文 http://www.economist.com/science/displayStory.cfm?Story_id=1010802, 邦訳 <http://www.f.u-tokyo.ac.jp/~pecout/event/campbell.html>).
3. 診療ガイドラインの作成の手順 ver. 4.3 2001.11.7. (<http://www.iph.go.jp/glg1-4.3rev.htm>).
4. 青木 仕. 構造化抄録の基礎知識, 中嶋宏(監修),津谷喜一郎,他(編). EBM のための情報戦略.中外医学社, 2000; 82-93.
5. 中山建夫. EBM を指向した「診療ガイドライン」と医学データベースに利用される「構

Social Experiment と科学的評価の枠組み
日本評価学会第3回全国大会 2002.12.7-8(東京)

造化抄録」作成の方法論の開発とそれらの受容性に関する研究. 平成13年度 総括・分担研究報告書. 厚生科学研究費補助金 21世紀型医療開拓推進研究事業. 2002年4月.