ISSN 1346-6151

日本評価研究

Japanese Journal of Evaluation Studies

Vol. 20, No. 2, July 2020

特集：「エビデンスに基づく政策立案（EBPM）」の現状と課題

巻頭言　「エビデンスに基づく政策立案（EBPM）」の現状と課題
佐々木 亮　正木 朋也

エビデンスに基づく政策立案（EBPM）と研究論文の質改善のための報告ガイドライン　正木 朋也　津谷 喜一郎

エビデンスに基づく政策立案（EBPM）の推進に向けて
—医療の経験からの示唆—　田辺 智子

日本におけるエビデンスに基づく政策形成（EBPM）の現状と課題
—Evidence-Basedが先行する分野から何を学び何を乗り越える必要があるのか—　小林 庸平

国の行政機関におけるEBPMの取組実例の研究—現状と課題—　砂山 裕

地方自治体におけるEBPMの組織への定着に向けた課題の考察　西畑 宏治

教育分野におけるエビデンス仲介機関の特徴　森 俊郎・岡崎 善弘

国際開発分野におけるEBPMの現状　佐々木 亮

研究論文

地方創生アンケートから評価する市町村の合併と連携　村上 裕一

実践・調査報告

参加型・質的評価手法MSC（Most Significant Change）の国内3事例を考察する—学習・改善目的中心の評価—　田中 博

第20回全国大会開催報告

第21回全国大会のご案内

日本評価学会
Japan Evaluation Society

『日本評価研究』編集委員会
Editorial Board

編集委員長
Editor-in-chief
西野 桂子（関西学院大学）
Keiko NISHINO

常任編集委員
Standing Editors
大島 巌（日本社会事業大学）
Iwao OSHIMA
牟田 博光（東京工業大学）
Hiromitsu MUTA
山谷 清志（同志社大学）
Kiyoshi YAMAYA

編集委員
Editors
岩渕 公二（NPO政策21）
Koji IWABUCHI
岡本 義朗（島根県立大学）
Yoshiaki OKAMOTO
小野 達也（鳥取大学）
Tatsuya ONO
窪田 好男（京都府立大学）
Yoshio KUBOTA
佐々木 亮（国際開発センター）
Ryo SASAKI
田中 弥生（会計検査院）
Yayoi TANAKA
南島 和久（新潟大学）
Kazuhisa NAJIMA
西出 順郎（明治大学）
Junro NISHIDE
林　　薫（文教大学）
Kaoru HAYASHI
源 由理子（明治大学）
Yuriko MINAMOTO

〒108-0075　東京都港区港南1-6-41　芝浦クリスタル品川12階
一般財団法人国際開発センター内
特定非営利活動法人日本評価学会編集委員会事務局
E-mail: journal@evaluationjp.org

日 本 評 価 研 究

第20巻　第２号　2020年７月

目　次

特集：「エビデンスに基づく政策立案（EBPM）」の現状と課題
佐々木 亮　正木 朋也
巻頭言　「エビデンスに基づく政策立案（EBPM）」の現状と課題 …… 1
正木 朋也　津谷 喜一郎
エビデンスに基づく政策立案（EBPM）と研究論文の質改善のための報告ガイドライン … 3
田辺 智子
エビデンスに基づく政策立案（EBPM）の推進に向けて
—医療の経験からの示唆— …… 19
小林 庸平
日本におけるエビデンスに基づく政策形成（EBPM）の現状と課題
—Evidence-Basedが先行する分野から何を学び何を乗り越える必要があるのか— …… 33
砂山 裕
国の行政機関におけるEBPMの取組実例の研究—現状と課題— …… 49
西畑 宏治
地方自治体におけるEBPMの組織への定着に向けた課題の考察 …… 65
森 俊郎・岡崎 善弘
教育分野におけるエビデンス仲介機関の特徴 …… 77
佐々木 亮
国際開発分野におけるEBPMの現状 …… 89

研究論文
村上 裕一
地方創生アンケートから評価する市町村の合併と連携 …… 105

実践・調査報告
田中 博
参加型・質的評価手法MSC（Most Significant Change）の …… 121
国内3事例を考察する—学習・改善目的中心の評価—

第20回全国大会開催報告
開催の報告とお礼…… 137
プログラム詳細…… 139
シンポジウム報告…… 142
共通論題セッション報告…… 143
自由論題セッション報告…… 151

第21回全国大会のご案内 …… 157

日本評価研究刊行規定…… 158
日本評価研究投稿規定…… 160
日本評価研究執筆要領…… 162
日本評価研究査読要領…… 165

Publication Policy of the Japanese Journal of Evaluation Studies …… 167
Information for Contributors (For English Papers) …… 169
Writing Manual of the Japanese Journal of Evaluation Studies (For English Papers) …… 171
Referee-Reading Guideline…… 173

【巻頭言】

「エビデンスに基づく政策立案（EBPM）」の現状と課題

佐々木　亮・正木　朋也

1．特集の目的

この特集は、およそ5年ごとに実施してきた「エビデンスに基づく実践（EBP：Evidence-Based Practice）」に関する学会誌特集号の4回目にあたる。日本評価学会に社会実験分科会（Social Experimentation Study Group）が設立されたのが2001年であり、その後、2006年、2010年、2016年に特集号が組まれてきた。今回の2020年の特集では、各分野における最新の普及の現状と課題とともに日本ではここ数年、政府主導で急速に普及進展しつつあるEBPM（Evidence-Based Policy Making）の動向について報告して議論することを目的としている。

2000年に日本評価学会が設立された当初は、医療分野を除いて世界的にもEBPに関する認識は高くなく、ましてや日本国内では、一部の研究者を除いてほぼ全く認識されておらず普及もしていなかった。その後、コクラン共同計画、キャンベル共同計画、貧困アクションラボ（J-PAL）、3ie（International Initiative of Impact Evaluation）などの国際的な団体の活動により、EBPが強力に推進され、国際的な議論の場での普及が進んだ。

そして、2019年度のノーベル経済学賞が、マサチューセッツ工科大学（MIT）の貧困アクションラボ（J-PAL）の研究者3名に授与された。公式の受賞理由は、「世界の貧困軽減に対する実験的アプローチに対して」であった。

国内に目を向けると、ほぼ唯一、国内での横並びを許されずに常に国際的な動向に追いついていくことを強く要請されるODA分野において、JICAを中心としてEBPの取組がなされてきた。一方、日本のその他の行政分野ではEBPの普及が遅々として進んでいなかったが、2016年から動きが見え始めた。官民データ活用推進基本法（2016年）の成立後、いわゆる「骨太の方針（経済財政運営と改革の基本方針2017）」で「EBPMの推進」が謳われたのち、現場におけるEBPMの取組がやっと始まった段階である。

その後EBPMについての現場認識が上述の国際潮流と整合しない状況も垣間見え、日本評価学会においても何度かこの問題について共通セッションを設けて議論を重ねてきた。こうした状況を踏まえて、本特集号では、国際動向にも整合し始めたEBPの動きとともに急展開中の日本のEBPMの現状と課題についても議論し、国内各分野におけるEBPおよびEBPMの在り方に関しても議論することとした。EBPMも踏まえた国内各分野の現状を扱うのは、EBPの特集号を開始して以来、初の試みであり、編者としてもたいへんうれしく思っている。ただし、国際的にも、そして特に日本国内の議論において、EBPMで扱う「エビデンス」の意味が拡大しており、本来EBPで厳密な意味で用いられてきた「エビデンス」を再確認して、その理解と普及が図られねばならないと言える。

執筆者は、各分野の最前線で活躍する研究者および実務者で構成されている。この特集号が今後のEBPMの普及に大きな貢献となることを期待している。

日本評価学会『日本評価研究』第20巻第2号、2020年、pp.1-2

2. 各論文の概要

本特集号は合計7本の論文で構成されている。

正木朋也・津谷喜一郎の「エビデンスに基づく政策立案（EBPM）と研究論文の質改善のための報告ガイドライン」では、保健医療領域で経験した研究報告の質の改善に関わる問題と課題解決への取組について概観し、EBPMにつながる社会科学などの他領域への適用可能性とその課題について考察している。

田辺智子の「エビデンスに基づく政策立案（EBPM）の推進に向けて－医療の経験からの示唆－」では、医療におけるエビデンス活用の経験を概観し、EBPMへの示唆を抽出しており、エビデンスは社会全体で蓄積し共有すべきであることをはじめとする7点の示唆が示されている。

小林庸平の「日本におけるエビデンスに基づく政策形成（EBPM）の現状と課題－Evidence-Basedが先行する分野から何を学び何を乗り越える必要があるのか－」では、日本のEBPMの現状を整理すると共にその評価を行い、問いの設定の重要性や、エビデンスの範囲、つくる・つたえる・つかうのウエイト、エビデンスに対する需要創出の重要性などを指摘している。

砂山裕の「国の行政機関におけるEBPMの取組実例の研究－現状と課題－」では、EBPMのリーディングケースの提示を目指し、国の政策を対象として、総務省、関係府省および学識経験者による政策効果の把握・分析手法の実証的共同研究が行われた結果、「EBPM的要素」として日本の行政運営上有用と考えられる項目がいくつか見出されつつあることを指摘している。

西畑宏治の「地方自治体におけるEBPMの組織への定着に向けた課題の考察」では、EBPMの地方自治体組織への定着に向けて、EBPMが要請する行動様式を地方自治体の行政運営に適用しようとした場合、地方自治体のどこに課題が生じ、それを解決するためにどのようなことを検討する必要があるのか、また、どのような解決策が考えられるのかを考察している。その際、組織への定着に着目した動きを早くから進めている広島県の取組も取り上げて、参考にしている。

森俊郎・岡崎善弘の「教育分野におけるエビデンス仲介機関の特徴」では、3つのエビデンス仲介機関（WWC、EEF、CUREE）を対象として、レビューの対象となる研究を選別する基準・プロセスや提供しているエビデンスの種類・数について調べており、日本の教育分野におけるエビデンス仲介機関の在り方について考察している。

最後の佐々木亮の「国際開発分野におけるEBPMの現状」では、新興の3つの非営利の研究機関である貧困アクションラボ（J-PAL）、IPA、3ieについて概観し、さらに、伝統的な援助機関である世界銀行、アメリカ国際開発庁（USAID）、イギリス国際開発局（DFID）、日本の国際協力機構（JICA）におけるインパクト評価の現状について概観して、国際開発分野における今後のEBPMの課題について考察している。

本稿が、日本におけるエビデンスに基づく政策立案（EBPM）の普及の一里塚となることを切に希望する次第である。

【実践・調査報告】

エビデンスに基づく政策立案（EBPM）と研究論文の質改善のための報告ガイドライン

正木　朋也
国際協力機構／北里大学
masakit-tky@umin.ac.jp

津谷　喜一郎
東京有明医療大学／東京大学
tsutanik@gmail.com

要　約

日本では2010年代後半以降、公的セクターでの「エビデンスに基づく政策立案」（evidence-based policy making: EBPM）の方針が出された。医学領域では1996年にランダム化比較試験（randomized controlled trial: RCT）論文の質向上のためのCONSORT声明が開発された。報告ガイドラインの嚆矢となる。その第3版の全25項目からなるCONSORT 2010は16カ国語に翻訳され世界で広く使われている。また、観察研究（断面研究、後向き研究、前向き研究）やシステマティック・レビューなど基本的な研究デザインについて、さらにその拡張版としてクラスターランダム化や非薬物系などの種々の報告ガイドラインが開発され、その総数は400件を超す。これから研究計画を作成しようとする者にとり利用価値の高い、種々の報告ガイドラインをカバーするEQUATOR Networkが2008年に設立され広く用いられている。EBPMに報告ガイドラインを使う事はそれなりの実現可能性をもつ。

キーワード

EBPM、報告ガイドライン、CONSORT、STROBE、EQUATOR Network

1．はじめに

私たちは本誌に過去2回「エビデンスに基づく医療」（evidence-based medicine: EBM）についてその展開を含めて報告してきた。ここでは2006年の論文（正木・津谷 2006）を第1報、2010年の論文（正木・津谷2010）を第2報と呼ぶこととする。それ以後の10年間でEBMの関連領域で起きたことが2つある。

1つは、日本の公的セクターでのEBPM（evidence-based policy making、エビデンスに基づく政策立案）の動きである。2017年5月、首相官邸が招集した「統計改革推進会議」の最終とりまとめ[1]が公表された。翌2018年10月には、「EBPM（エビデンスに基づく政策立案）に関する有識者との意見交換会報告（議論の整理と課題等）」[2]が公表され、全領域で公的なファンドを用いた研究が要請される時代となった。

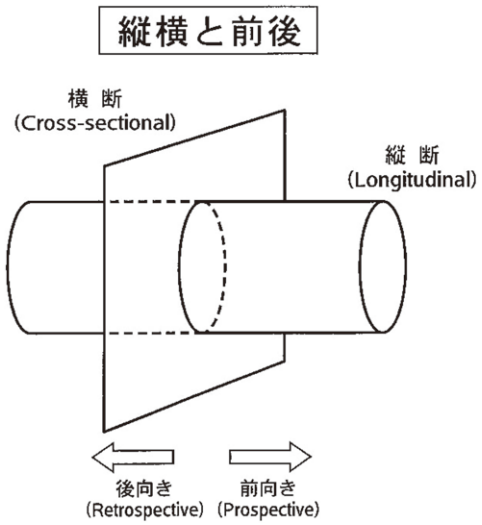
著者の一人、津谷は『いろいろな分野のエビデンス：温泉から国際援助までの多岐にわたるRCTやシステマティック・レビュー』を刊行した（津谷 2015）。その全12章のうち、国際援助と教育の章を除き、他の10章はすべて医学領域であった（第1報と第2報では「保健医療領域」の用語を用いた。本稿ではより分かりやすい「医学」を広義のものとして用いることとする[3]）。またそ

れらの多くは非政府系の活動であった。政府系、すなわち公的セクターでのEBPMは、研究者と納税者の双方の立場から質の高い研究が望まれる。

もう1つは、論文などの質を高めるための「報告ガイドライン」(reporting guidelines: RG）の多様な領域での数多くの発行とその使用である。ランダム化比較試験（randomized controlled trial: RCT）に対する1996年のCONSORT（以降CONSORT声明をこう略す。他のRGについても同様）第1版は、私たちの第1報で簡単に触れられた。また2001年の第2版は、第2報で若干紹介された。だが2010年3月に大きく改訂された第3版は、第2報が同時期の2010年3月発行であったことから紹介されていなかった。このCONSORT第3版は“CONSORT 2010”として知られ、その後多くの拡張版（Extension）の作成が加速した。

一方、第2報で述べたように、1）コホート研究（cohort study)、2）ケース・コントロール研究（case-control study)、の時間軸をもつ2つの縦断研究（longitudinal study）と、時間軸を持たない、3）横断研究（cross sectional study）の3つをカバーするSTROBE（von Elmら2007, 上岡 2008)、またその「解説と詳細」(explanation and elaboration: E&E, Vandenbrouckeら2007、日本語訳 福原ら2009）が2007年に発行された。この3つの研究デザインの関係を図1に示す（中嶋ら 2000)。

図1　時間軸を含めた研究デザインの整理



本稿では、第1に、日本でのEBPMの動向を念頭におき、まずバイアスの少ない、すなわちより正確（accurate）な結果を得られるRCTのためのCONSORTについてその歴史的発展と拡張版などを含める。第2に、観察研究のSTROBEの3つの研究デザインのなかでケース・コントロール（case-control）研究の部分のチェックリストを紹介する。第3に、多様な機能を持つEQUATOR NetworkとICMJEの動きについて述べ、第4に、医学の領域で開発されたこれらの手法やシステムのEBPMへの技術移転の実現性（feasibility）について論ずる。

2. CONSORT

(1) CONSORTの誕生

1996年のCONSORT（Beggら 1996;日本語訳: 津谷ら 1997）が注目されたのはその画期性とともに、周到にそのRGが有効かどうかの検証が行われたことが影響している。1996年にCONSORTが発表される前に公表されたRCT論文の質と、発表後に公表されたRCT論文の質を、世界的に著名な4つの総合医学雑誌を対象にして比較したのである。当時、すぐにCONSORTを論文投稿の条件とした3雑誌（*BMJ, Lancet, JAMA*）は、以後、統計学的に有意に論文の質が向上した。一方でその時点ではCONSORTを投稿の条件にしていなかった*NEJM*では質の向上がなかったことを報告した（Moherら 2001)。

RCT論文の質評価のチェックリストの開発は1960年代から始まる。そのうち最も引用回数の多いのは予備的解析ではChalmers（1981）であった（津谷・佐久間 1993)。1992年に始まるコクラン共同計画（The Cochrane Collaboration: CC）の初期[4]には、RCT論文の質評価に基づく選択に様々なチェックリストが使われていた。1996年の“Jadad Scale”は、ランダム化（randomization)、ブラインディング（blinding)、除外・脱落（withdrawal and dropouts）の3項目から成り立つ簡便なものであった（Jadadら 1996)。当時のコクランハンドブック（ver. 4.2.6）にも引用された。だがその後、単純すぎる、ブラインディングに重きがなされ

すぎている（Clarkら 1999）－などの理由で使われなくなる。

コクランレビューのアプリであるReview Manager（RevMan）は最初の版が1993年にリリースされた（Chandlerら 2013）。その2008年のversion 5（RevMan.5）に導入されたのが“Risk of Bias”toolのグラフである。そこでは7つの項目で各RCT論文の質を評価しその結果を棒グラフで表示している（Higginsら2011）。

（2）CONSORT 2010の内容

RGは進化していくものである。1996年の第1版のチェックリストは項目にナンバリングがなかった。2001年の第2版は内容の改訂とともにナンバリングがなされ全22項目となった（Moherら 2001;津谷ら 2002）。2010年の現行の第3版ではチェックリストの項目数が25項目へと増加した（Schulzら 2010；日本語訳: 津谷ら 2010）。表2に示すこの3項目の追加には、臨床試験 を取り巻く環境の大きな変化が反映されている。

第23項目の「臨床試験登録」は、最初の試験対象者をリクルートする前に試験内容を項目毎にレジストリ（Registry）[5]に登録しそのレジストリの名称と登録番号を記入することである。Primary RegistryやData Providerは、日本を含めて現在世界17か国・地域に存在する。それらの試験情報はWHO-ICRPで統合されて2006年からすべて検索することができる。なお、WHOが必須とするdata setは当初の20項目から2017年に24項目に増加し2019年をゴールとするものである（津谷 2017a、2017b）。第24項目は、臨床試験に参加した個人のデータシェアリング・プラン（individual data sharing plan: IDSP）も含まれる（津谷 2018a、2018b）。

これの臨床試験登録の主たる目的は、結果が悪いと発表しない、または本来の1次アウトカム指標がよくない場合、2次アウトカム指標とされたものからよいものをあたかも1次アウトカム指標であるかのように報告する「選択的報告」（selective reporting）を含む「パブリケーションバイアス」を避けるためのものである（Macleodら 2014; 奥村ら 2014）。つまり結果がよくとも悪くともそれをプロトコール通りに解析し公開すべきだという考えである。それには、多くの患者の代表として試験に参加した患者から得られた結果は患者に戻されるべき公共財（public goods）であるという考え方がある[6]。

第24項目の「プロトコール入手可能性」は、プロトコールが遵守されず、上記のように本来とは異なった解析がなされる問題が顕在化したことに対応する。上記の選択バイアス防止の一手段である。

第25項目の「資金提供者」（funder）は、研究資金の提供主体が試験結果に影響する、つまり企業資金の場合、企業に有利な結果が報告されやすい、または不利な結果が報告されにくい、というエビデンスに基づく（Melander 2003）。なお、資金提供者とスポンサー（sponsor）は別のものであることに留意されたい（津谷・澤田 2007）。スポンサーの定義は3つの要素からなり分かりづらい[7]。

CONSORTのもう1つの重要な要素はフローチャートであり、図2に示す。このフローチャートは、先のJadad Scaleの第3項目の「除外・脱落」に関する事項を、より認知しやすい図形にしたものと理解できる。

なお、CONSORT声明は3つの版とも「解説と詳細」（explanation & elaboration）が発行され free access となっている。第2版のみ日本語訳が存在する（Altmanら 2001; 津谷・栗原 2002）

（3）CONSORTの拡張版

表1に本稿執筆時点でのCONSORTのExtensionを、年代順に整理して示した。

ここで、SPIRIT（standard protocol items: recommendations for interventional trials）を紹介しよう（Chanら 2013、日本語訳 折笠ら 2017）。臨床試験が終わり、原稿を書く段になり、CONSORTのチェックリストに記載すべき情報が欠損していることが分かることがある。多くはプロトコールが不十分なためである。そこでプロトコール段階のチェックリストとして開発されたものである。大変有用なツールである。RCTの実施の前にprotocolを作成するのであるから「前向き」の「拡張版」ともいえ、後述するEQUATOR NetworkではCONSORTのカテゴリーに入っている。

表1　CONSORTのRCT領域に限ったExtensionのいろいろ

	試験デザイン	介入内容	データ
2001	・クラスター(第1版)(日)		
2002		・鍼(第1版)(日)	
2004			・害(日)
2006		・ハーブ	
2008	・実践的(日)	・非薬物介入(第1版)(日)	・抄録(日)
2010		・鍼(第2版)(日)	
2012	・クラスター(第2版) ・非劣性・同等性		
2013	・SPIRIT(プロトコール用)		・患者報告アウトカム(PRO)(日)
2014		・不妊治療	
2015	・N-of-1	・歯科矯正	
2016	・パイロット・実現可能性		
2017	・個人内	・中薬複方 ・非薬物介入(第2版)(日)	
2018	・段階的割り込み型クラスター	・社会的・心理的介入	
2019	・多群並行群間 ・クロスオーバー ・N-of-1 TCM(中医学)		

(注)(日)日本語訳あり
(出所)津谷ら(2019)の整理にEQUATOR Networkをもとに更新(2019-10-12)

一部のRGでは、その理解を促すための「解説と詳細」も作成され多くはfree accessが可能である。これらRGやE&Eの一部は日本語にも翻訳されており、雑誌での公表や書籍として出版されている（中山・津谷 2008; 2016）。各国語訳されたRGやE&Eは、同じくEQUATOR Networkのwebsiteからpdfを入手できる。日本語訳についてはJPT Onlineが、それらをよくカバーしていると思われる。

(4) CONSORTチェックリストとフローチャートの使い方

投稿者にとっての実際の使い方としては、論文を作成し投稿する者はCONSORTのチェックリストのフォーム（表2）に各項目の記載された原稿の「報告頁」を記載したものと、フローチャート（図2）にそれぞれの患者数を記載したものの2つを、投稿先へ原稿と共にemailなどで送る。これのチェックリストやフローチャートはpdfとWord版が各websiteにあり、これらをdownloadしてそこに記入すればよい。

それを編集委員会のスタッフが事務的にチェックし、ついで編集長や査読者（reviewer）にもそれらが原稿とともに送られて、各項目が実際に記載されているのかをチェックする。

フローチャートは論文そのものに掲載される。一方、チェックリストは多くはOpen access雑誌では、supplementary informationとして公開されることとなる。

例としてonline journalで、free accessとなっている*PLoS ONE*誌に掲載され、鍼の領域で2017年度の代田賞を獲得した、会陰部の夜間のかゆみに対するローラー鍼の二重盲検RCTの例をあげる（Iimuraら 2016）。ここから、チェックリストは、原稿の報告ページ番号を記載するが、電子出版の論文においてはページ番号が変わり、項目名で示されていることが分かる。またプロトコールや倫理委員会の審査記録などはこのsupplementary informationに収載され公開されている。

(5) 非薬物介入とクラスター RCT

EBPMでの利用が期待されるCONSORT ExtensionのRGで社会科学の領域での使用が期待されるものが2つある。

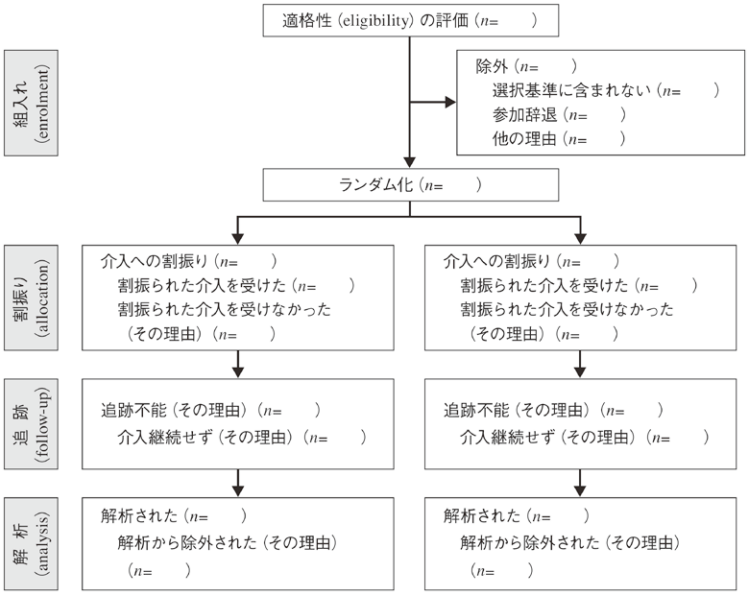
1つは、CONSORT非薬物介入（CONSORT NPT（non-pharmacologic treatment））である（初版Boutronら 2008a、2008b、日本語訳 上岡・津谷 2016、第2版 Boutronら2017、日本語訳 津谷ら 2019）。非薬物介入の具体的なものとして、手術、手技（たとえば血管形成術などの技術的処置）、デバイス（ペースメーカーなど）、リハビリテーション、理学療法、行動療法、心理療法などを

表2　CONSORT 2010チェックリスト

章／トピック (Section/Topic)	項目番号 (Item No)	チェックリスト項目 (Checklist Item)	報告頁 (Reported on page No)
タイトル・抄録 (Title and Abstract)			
	1a	タイトルにランダム化比較試験であることを記載。	
	1b	試験デザイン（trial design），方法（method），結果（result），結論（conclusion）の構造化抄録（詳細は「雑誌および会議録でのランダム化試験の抄録に対する CONSORT 声明」[21, 31]を参照）。	
はじめに (Introduction)			
背景・目的 (Background and Objective)	2a	科学的背景と論拠（rationale）の説明。	
	2b	特定の目的または仮説（hypothesis）。	
方法 (Method)			
試験デザイン (Trial Design)	3a	試験デザインの記述（並行群間，要因分析など），割付け比を含む。	
	3b	試験開始後の方法上の重要な変更（適格基準 eligibility criteria など）とその理由。	
参加者 (Participant)	4a	参加者の適格基準（eligibility criteria）。	
	4b	データが収集されたセッティング（setting）と場所。	
介入 (Intervention)	5	再現可能となるような詳細な各群の介入。実際にいつどのように実施されたかを含む。	
アウトカム (Outcome)	6a	事前に特定され明確に定義された主要・副次的アウトカム評価項目。いつどのように評価されたかを含む。	
	6b	試験開始後のアウトカムの変更とその理由。	
症例数 (Sample size)	7a	どのように目標症例数が決められたか。	
	7b	あてはまる場合には，中間解析と中止基準の説明。	
ランダム化 (Randomization)			
順番の作成 (Sequence generation)	8a	割振り（allocation）順番を作成（generate）した方法。	
	8b	割振りのタイプ：制限の詳細（ブロック化，ブロックサイズなど）。	
割振りの隠蔵機構 (Allocation concealment mechanism)	9	ランダム割振り順番の実施に用いられた機構（番号付き容器など），各群の割付けが終了するまで割振り順番が隠蔵されていたかどうかの記述。	
実施 (Implementation)	10	誰が割振り順番を作成したか，誰が参加者を組入れ（enrollment）たか，誰が参加者を各群に割付けた（assign）か。	
ブラインディング (Blinding)	11a	ブラインド化されていた場合，介入に割付け後，誰がどのようにブラインドかされていたか（参加者，介入実施者，アウトカムの評価者など）。	
	11b	関連する場合，介入の類似性の記述。	
統計学的手法 (Statistical method)	12a	主要・副次的アウトカムの群間比較に用いられた統計学的手法。	
	12b	サブグループ解析や調整解析のような追加的解析の手法。	
結果 (Results)			
参加者の流れ (Participant flow)（フローチャートを強く推奨）	13a	各群について，ランダム割付けされた人数，意図された治療を受けた人数，主要アウトカムの解析に用いられた人数の記述。	
	13b	各群について，追跡不能例とランダム化後の除外例を理由とともに記述。	
募集 (Recruitment)	14a	参加者の募集期間と追跡期間を特定する日付。	
	14b	試験が終了または中止した理由。	
ベースライン・データ (Baseline data)	15	各群のベースラインにおける人口統計学的（demographic），臨床的な特性を示す表。	
解析された人数 (Number analyzed)	16	各群について，各解析における参加者数（分母），解析が元の割付け群によるものであるか。	
アウトカムと推定 (Outcome and estimation)	17a	主要・副次的アウトカムのそれぞれについて，各群の結果，介入のエフェクト・サイズの推定とその精度（95%信頼区間など）。	
	17b	2項アウトカムについては，絶対エフェクト・サイズと相対エフェクト・サイズの両方を記載することが推奨される。	
補助的解析 (Ancillary analysis)	18	サブグループ解析や調整解析を含む，実施した他の解析の結果。事前に特定された解析と探索的解析を区別する。	
害 (Harm)	19	各群のすべての重要な害（harm）または意図しない効果（詳細は「ランダム化試験における害のよりよい報告：CONSORT 声明の拡張」[28]を参照）。	
考察 (Discussion)			
限界 (Limitation)	20	試験の限界，可能性のあるバイアスや精度低下の原因，関連する場合は解析の多重性の原因を記載。	
一般化可能 (Generalisability)	21	試験結果の一般化可能性（外的妥当性，適用性）。	
解釈 (Interpretation)	22	結果の解釈，有益性と有害性のバランス，他の関連するエビデンス。	
その他の情報 (Other information)			
登録 (Registration)	23	登録番号と試験登録名。	
プロトコール (Protocol)	24	可能であれば，完全なプロトコールの入手方法。	
資金提供者 (Funding)	25	資金提供者と他の支援者（薬剤の供給者など），資金提供者の役割。	

（注）＊標準2010年の改定で23、24、25の3項目が追加された。
（出所）津谷ら（2010）

図2　2群間並行ランダム化比較試験の各段階の過程を示すフローチャート
（組み入れ、介入への割振り、追跡、データ解析）



（出所）津谷ら（2010）

含み、幅広い領域をカバーする。

このCONSORT NPT Extensionは主に精神科医と外科医によって作られたものである。“Treatment”は「治療」だけではなく、「処置」や刑務所での「矯正」の訳がある。このRGは社会科学領域でもそのまま適用可能なRGであろう。EBPMの現場で行われるインパクト評価の計画と報告書作成にも有益である[8]。そこでは非薬物介入の複雑さ、施設とケア提供者の専門性の影響、ブラインド化の困難さも論じられている。

EBPMを推進しようとしている方は、RCTの論文原稿作成時ではなく、まずプロトコール作成時に先のSPIRITとこのCONSORT NPTを使用することが勧められる。

2つめは、クラスター RCTへの拡張版である（Campbellら 2004、2012）。ここには、同時に抄録の書き方のExtension（Hopewellら 2008）も含まれ、その全てのチェックリストが記載されている。10番目の実施（Implementation）の項目は10a、10bおよび10cに分割されている。

クラスター RCTはランダム割付け（random allocation またはrandomization）を、集団（cluster）を単位として行うものである。2009年頃より北米と西欧諸国を中心として世界的にクラスターRCTの利用が増えつつある（増島ら 2013）。なお、クラスターランダム化と他のランダム化の方法などの基本事項については他書を参照されたい（佐藤 2007）。

3. STROBEのケース・コントロール研究

「疫学研究における観察研究報告の質改善」（STROBE: strengthening of reporting of observational studies in epidemiological studies）のRG である。STROBEは、観察研究がその主体ともいえるEBPMでの応用が大いに期待される。観察研究は疫学研究から介入研究である臨床試験を除いたものである。

疫学（epidemiology）はその漢字表記から日本では医学領域であると誤解されがちである。だが、epiは周辺・辺縁の、-demosは人間・集団の、-logosは学問であり、人間を集団としてみる観点からあらゆる因果関係を明らかにしようとする

表3　STROBE: ケース・コントロール研究の報告において記載すべき項目のチェックリスト

	No	推奨	報告頁
タイトル・抄録［title and abstract］	1	（a）タイトルまたは抄録のなかで，試験デザインを一般に用いられる用語で明示する。 （b）抄録では，研究で行われたことと明らかにされたことについて，十分な情報を含み，かつバランスのよい要約を記載する。	
はじめに［introduction］			
背景［background］/論拠［rationale］	2	研究の科学的な背景と論拠を説明する。	
目的［objective］	3	特定の仮説を含む目的を明記する。	
方法［methods］			
研究デザイン［study design］	4	研究デザインの重要な要素を論文のはじめの［early］部分で示す。	
セッティング［setting］	5	セッティング，実施場所のほか，基準となる日付については，登録，曝露［exposure］，追跡，データ収集の期間を含めて明記する。	
参加者［participant］	6	（a）適格基準，参加者の母集団，ケース［case］の確定方法とコントロール［control］の選択方法を示す。ケースとコントロールの選択における論拠を示す。 （b）マッチング研究［matched study］の場合，マッチングの基準，ケースあたりのコントロールの人数を記載する。	
変数［variable］	7	すべてのアウトカム，曝露，予測因子［predictor］，潜在的交絡因子［potential confounder］，潜在的な効果修飾因子［effect modifier］を明確に定義する。該当する場合は，診断方法を示す。	
データ源［data source］/測定方法［measurement］	8*	関連する各因子に対して，データ源，測定・評価方法の詳細を示す。二つ以上の群がある場合は，測定方法の比較可能性［comparability］を明記する。	
バイアス［bias］	9	潜在的なバイアス源に対応するためにとられた措置があればすべて示す。	
研究サイズ［study size］	10	研究サイズ［訳者注：観察対象者数］がどのように算出されたかを説明する。	
量的変数［quantitative variable］	11	量的変数の分析方法を説明する。該当する場合は，どのグルーピング［grouping］がなぜ選ばれたかを記載する。	
統計・分析方法［statistical method］	12	（a）交絡因子の調整に用いた方法を含め，すべての統計学的方法を示す。 （b）サブグループと相互作用［interaction］の検討に用いたすべての方法を示す。 （c）欠損データ［missing data］をどのように扱ったかを説明する。 （d）該当する場合は，ケースとコントロールのマッチングをどのように行ったかを説明する。 （e）あらゆる感度分析［sensitivity analysis］の方法を示す。	
結果［result］			
参加者［participant］	13*	（a）研究の各段階における人数を示す（例：潜在的な適格［eligible］者数，適格性が調査された数，適格と確認された数，研究に組入れられた数，フォローアップを完了した数，分析された数）。 （b）各段階での非参加者の理由を示す。 （c）フローチャートによる記載を考慮する。	
記述的データ［descriptive data］	14*	（a）参加者の特徴（例：人口統計学的，臨床的，社会学的特徴）と曝露や潜在的交絡因子の情報を示す。 （b）それぞれの変数について，データが欠損した参加者数を記載する。	
アウトカムデータ［outcome data］	15*	各曝露カテゴリー内の数，または曝露の集約尺度を示す。	
おもな結果［main result］	16	（a）調整前［unadjust］の推定値と，該当する場合は交絡因子での調整後の推定値，そしてそれらの精度（例：95%信頼区間）を記述する。どの交絡因子が，なぜ調整されたかを明確にする。 （b）連続変数［continuous variable］がカテゴリー化されているときは，カテゴリー境界［category bordary］を報告する。 （c）意味のある［relevant］場合は，相対リスク［relative risk］を，意味をもつ期間の絶対リスク［absolute risk］に換算することを考慮する。	
他の解析［other analysis］	17	その他に行われたすべての分析（例：サブグループと相互作用の解析や感度分析）の結果を報告する。	
考察［discussion］			
鍵となる結果［key result］	18	研究目的に関しての鍵となる結果を要約する。	
限界［limitation］	19	潜在的なバイアスや精度の問題を考慮して，研究の限界を議論する。潜在的バイアスの方向性と大きさを議論する。	
解釈［interpretation］	20	目的，限界，解析の多重性［multiplicity］，同様の研究で得られた結果やその他の関連するエビデンスを考慮し，慎重で総合的な結果の解釈を記載する。	
一般化可能性［generalisability］	21	研究結果の一般化可能性（外的妥当性［external validity］）を議論する。	
その他の情報［other information］			
研究の財源［funding］	22	研究の資金源，本研究における資金提供者［funder］の役割を示す。該当する場合には，現在の研究の元となる研究［original study］についても同様に示す。	

*ケースとコントロールに分けて記述する。

（出所）上岡ら（2014、p.s116 表2）

学問であり、社会科学系の領域においても相補的に活用できる学問体系といえる。

表4に、日本で、すでに電子媒体などに蓄積されている過去のデータが使える「後向き」（retrospective）デザインであるケース・コントロール研究チェックリストを示した。なお、時間軸を含まない横断研究では原則的に因果関係の推論は不可能である。

4. EQUATOR NetworkとICMJE

EQUATOR Networkはこれまで述べてきたRGの動きを背景として2006年に設立されたネットワークである。以下の4つの目的を持つ。1）良質な研究報告（論文）の重要性についての関心の向上、2）各種研究デザインに対するよいRGの開発、流通（dissemination）、実施、の支援、3）医学論文の質のモニタリング、4）論文の質にインパクトを与える事項に関する－である（Simera 2009）。

臨床研究を計画し、また実施した後に投稿しようとする研究者、さらに編集者などのユーザーにとってはEQUATOR のwebsiteが大いに役立つ。2019年10月中旬時点でそのwebsiteは表4のように全領域を11項目に分類している。

そこには全419件のRGが存在する。対応する個々のRGにアクセス可能とし、そのExtension版など関連するガイドラインがリストされ、その多くはfree access の雑誌などにリンクされている。

このEQUATORと近い関係にあるのが国際医学雑誌編集者委員会（international committee for medical journal editors: ICMJE）である。書誌情報の記載法の標準化を行った1978年のVancouver Groupをその母体とする。

1979年から発行されてきた“Uniform Requirements for Manuscripts Submitted to Biomedical Journals: Writing and Editing for Medical Journal”（Uniform Requirementsと略されてきた、医学生物学雑誌の一般投稿規定）は数年に一度の改訂をして、世界の多くの医学雑誌の投稿規定のモデルとなり、医学雑誌の論文の質向上に貢献してきた。

ICMJE Recommendationsにfollow（従う）と宣言している雑誌は5,000誌以上ありリストが公開されている[9]。

これまで述べてきたCONSORTはその1997年の改訂で取り込まれた（Huthら 2004）。さらに利益相反に関わる記述の導入（Fontanarosa 2010）などがなされてきた。また医学メディアを取り巻く環境の急速な変化を反映して2010年代となり改訂が多くなった[10]。

2013年8月の改訂でその名称は“Recommendations for the Conduct, Reporting, Editing and Publication of Scholarly Work in Medical Journals”（ICMJE

表4　試験の種類毎の報告ガイドライン

研究方法など	ガイドラインの略称	Extension など
ランダム化試験	CONSORT	20 の拡張・関連版 (表 2 参照)
観察研究	STROBE	11 の拡張・関連版
システマティック・レビュー	PRISMA	9 つの拡張・関連版
研究計画書	SPIRIT	PRISMA-P
診断／予後研究	STARD	TRIPOD
症例報告	CARE	3 つの拡張・関連版
臨床実践ガイドライン	AGREE	RIGHT
質的研究	SRQR	COREQ
動物前臨床研究	ARRIVE	
質改善研究	SQUIRE	
経済評価	CHEERS	

（出所）EQUATOR Network. Webページを元に著者作成（閲覧 2019-10-12）

Recommendations）と変更された。“Requirements”から”Recommendations”とやや弱くなった形である。

RGは先に述べたように1997年とCONSORT発行の翌年に取り入れられたが、その後種々のRGが開発されたのに対応して、ICMJE Recommendationsの最新の2019年12月版にはこれまで述べてきたCONSORT、STROBE、さらにPRISMA（システマティック・レビュー用）、STARD（診断の正確性）が含まれるとともに、EQUATOR Networkが“resource”として記載されている。

ICMJEは、英米の有力雑誌5誌と世界各地の雑誌9誌の計14誌[11]と、米国国立医学図書館（National Library of Medicine: NLM）とWorld Association of Medical Editors（WAME）で構成される小さい組織である。アジア地域としては現在、韓国の旗艦雑誌（national flag journal）といえる*Journal of Korean Medical Science*（*JKMS*）が入っており、日本からは入っていない。このICMJEは規約がない組織であるが世界の医学雑誌をリードする存在でその影響力は大きい。

いわばEQUATOR NetworkがRGの開発機関とすると、ICMJEはそれらをそのRecommendationsに取り入れることで世界の雑誌に影響を与える執行機関で、表裏をなしているといえる。

5. 報告ガイドラインのEBPMへの技術移転

（1）形式がすぎて創造性を制限するのではないかとの批判

まず、RCTそれ自体の不理解と実施上の議論があることは押さえておく必要がある（Cartwright 2010; Deaton 2018）。そのうえで、しばしばRGのようなチェックリストは、無機的・機械的で味気なく、研究者の創造性を制限するものだと批判される。看護の領域では「いまだ定式化されていないものの、看護学の新しい知識の創出につながるような、画期的な研究が行われにくくなる一因となるのではないか」という考えも述べられている（萱間 2015）。著者らも最初にそれがRGの普及の阻害要因となるのでないかと考えた。

だが研究者は好きで研究を行うだけではない。特にRCTなどの臨床試験は人を対象にした人体実験である。そう考えると、個々の試験参加者に対する個別的倫理（individual ethics）の面からも、社会への集合的倫理（collective ethics）の面からも、この種のRGに沿った項目に沿って報告することは大切である（Ioannidis 2014）。また先に述べたICMJE Requirementsの項目として入り、RGに沿って記載されていな限り原稿が採択（accept）されないという制度的システムができたことも、医学領域において一定の成果を収めた理由になろう。

これはEBPMの領域においても同様である。試験参加者はそれなりの犠牲とリスクを伴うものであるとの認識が必要である。

（2）ELSIの問題

ELSIはEthics, Law and Social Implicationの略である。ここでは2つを取り上げる。

1つは、先に述べたクラスター RCTである。集団（cluster）からのインフォームド・コンセント（informed consent: IC）をいかに受けるべきか、また、介入の有無により試験参加者集団の利益と不利益の均衡（equipoise）をいかにとるべきかの問題がある。クラスター RCTを行う場合の倫理面からの質向上のためのガイドラインとしてオタワ声明があり（Weijerら 2012）、その後の議論とともに更新の提案がなされている（van der Graafら 2015）。

2つは、利害関係者の調停と相互に利益相反の生じ兼ねない関係者間の問題である（Chalmers 2014）。医療における意思決定は、受益者である患者のために、公衆衛生上の政策と保険財務政策上の「折り合い」をつけ「落としどころ」をみつけて制度・政策として適用することがある。種々の利害関係者（stakeholder）、すなわち1）患者（とその家族や支援団体）、2）医療技術の開発者または製造販売企業、3）医療機関関係者（医師、病院経営者）、4）保健政策担当者（厚生労働省ほか保険財政・制度にかかわる関係各機関）、それら相互の利益や負担を全体として調停することが必要となる。

この種の問題はELSIとして研究対象になって

きた。医学領域での事例が、EBPMの領域での問題解決に役立つことはあろう。

(3) ビッグデータやリアル・ワールド・データから得られる解とエビデンス

グローバルな巨大情報産業群（Google, Apple, Facebook, Amazon, Microsoft: GAFAM）などが開発・適用している人工知能（AI）を用いた即時提案型の情報技術に基づくビジネスモデルの応用も一考には値する。しかし、ここで扱われる「即時対応に用いる情報やAIが導く場当たり的な最適解」は、個々の場やケースにおいてのみ有効なものであるので、EBMに始まった「社会的に価値あるエビデンスを蓄積精査し公開しながら未来につなげようとする立場やアプローチ」とは本質的に異なるであろう。

なお、Oxfordの疫学統計学者Richard Petoが共著者に入りNEJMに掲載された“The magic of randomization versus the myth of real-world evidence”はこの領域に関心ある者には必読の論である（Collins 2020）。そこではまずデータの量が多くなろうとも観察研究の弱点そのものは変わらないことが述べられている。

(4) システマティック・レビューから1次研究への揺り戻し

2009年に「システマティック・レビューとメタアナリシスの質改善」（Preferred Reporting Items for Systematic Reviews and Meta-Analyses: PRISMA）のRGが公表された。またシステマティック・レビューのプロトコール作成のためのPRISMA-Pが2016年に公表された。

ところが2019年の夏、ショッキングな報告がなされた（Niforatosら 2019）。それは、米国医師学会誌内科学（*JAMA Internal Medicine*）にInvitation letterとして掲載された2ページの論考である。その内容はSRとMAの数（SRMA）とRCTの数およびその比（SRMA/RCT比）を1995年から2017年までの22年間の経過を推移として追ったものであった。

医療分野毎の傾向をみたグラフでは、2017年時点で既にその比が1を超えている（つまりSRがRCTより多い）分野や診療領域が散見される状況があることが示された。これは同じRCTが幾度も別のSRに組み込まれて評価されている状況を示している。

その原因として、非ランダム化比較試験を含むSRやMAの増加、また中国の研究者らの投稿がPubMed内で検索し得るMAの3分の1あったことなどが挙げられている。SRの数の増加は筆者らの調査においてその動向を指摘している（唐ら 2014）[12]。また検索対象がPubMedだけであることや手法上の限界があり、どの程度のSRMA/RCT比が妥当であるかなどの議論はできないとしている。

先のビッグデータやRWDへの懸念とともに、どうやら1990年代に始まるシステマティック・レビューの成長は少なくとも医学領域ではその量と質から陰りを見せ、1次研究、特に質の高いRCTが再び重要性を取り戻したようだ。EBPMにおいても常にエビデンスの質に留意する必要があろう。

(5)「産官学」での意思決定のための文書作成の異同

これまで述べてきたようにRGの方法は「産官学」の「学」（academia）、特に医学の領域で発展してきた。他の領域へのその技術移転の実現可能性（feasibility）はどうだろうか？「官」の公的機関によるEBPMではそのプロセスの中で、雑誌に掲載される際の査読（review）がとられることはあるだろうか？なかには外部の査読付きの雑誌に投稿することもあろう。本誌は査読付きの雑誌である。

だが、行政機構の中では「査読」の用語はあまり一般的ではないかもしれない。しかし政策実施のための資料はその性格から、決して一人で作成されるものではなく執筆者の所属する組織内、また関係する他の組織の、複数の目を経て加筆修正されて、最終版は作成される。そこでの質管理にはRGが同様に利用される可能性が多いであろう。

医学以外の「学」の分野ではどうだろうか？社会科学の領域ではインパクト評価の実施数が増えつつある（Cameronら 2015）。また社会科学の重要な部門の1つである経済学においてすでに

RGが取り上げられている。

「公」ではどうであろうか？世界を見れば、2017年に世界銀行は、経済学研究は殆ど、研究デザインにおける重要な項目の詳細を報告していないために、選択（selection）、損耗（attrition）、実行（performance）、検出（detection）、報告（reporting）、サンプルサイズ（sample size）に関わる6種類のバイアスが生じている可能性があるとしている。そしてその対策としてCONSORTを利用することを推奨している（Ebleら 2017）。

政府間機関（intergovernmental organization）である国連や国連システムの1つである世界銀行はいわば公的機関である。日本の都道府県の議会での発言中の「エビデンス」の頻度はかなりばらつきがある[13]。一方でいくつかの地方自治体はすでにEBPMのための政策文書を作成している[14]。

6. おわりに

バイアスを小さくしエビデンスの強い方法を開発して適応するという枠組みは、社会改善（social betterment）の本質である。医学領域と他の領域に違いはない。EBMの発展の中で2010年代になり急速に進化したRGのうち2つを紹介した。またその開発・使用を支援するEQUATOR-Networkを紹介した。RGのEBPMへの技術移転の実現可能性はそれなりに存在することを論じた。そこでRGは単なる報告書作成時のマニュアルとしてではなく、その本質を理解し「賢く」計画段階からチェックリストも念頭にデザインを考え、実施可能なところは積極的に適用してゆくことが重要である。

謝辞

本稿作成に協力いただいた聖路加国際大学看護学研究科八重ゆかり氏に謝意を表する。

注記

1 「統計改革推進会議 最終取りまとめ」2017.5

2 総務省EBPMに関する有識者との意見交換会事務局、「EBPM（エビデンスに基づく政策立案）に関する有識者との意見交換会報告（議論の整理と課題等）」、2018.10

3 日本医学会は、基礎医学、社会医学、臨床医学の3つの部門をもつ。本稿では、歯学、薬学、看護、また多様なco-medicalの領域を含むものとして、他領域の関係者にも分かりやすい「医学」を用いる。

4 そのwebsiteには2019.10時点で8,107件のコクランレビューのサマリーがfree accessとなっている（2015.1.31より単にCochrane（コクラン）に名称変更しロゴマークも変わった。本稿ではこれまでどおり「コクラン共同計画」の日本語訳を用いた）。

5 臨床試験の登録システムのこと。国際的に合意された項目のWHO data setは現在14項目（ver.1.3.1）. UMIN-CTRやjRCTなど日本のものも含めて各国のRegistryの登録項目はそれよりも一般に多い。なお本稿では以降CONSORTの項目番号とWHO ICTROの項目番号の2種が記載され混乱しやすいが、一度、現物をご覧になると容易に区別できるであろう。

6 津谷喜一郎、臨床試験登録の意義と役割について、UMIN臨床試験登録システム シンポジウム（2005.2.2）

7 WHO GCP（1995）とICH GCP（1996）の定義の原文は、“The individual, organization, group or other legal entity which takes responsibility for initiating, managing and/or financing a study”である。ここで、3つの要素を“and/or”で関係づける表現は、そのまま訳すと日本語としてわかりにくい。そこで、以下がよいと思われる。「ある試験の企画、管理、資金のすべて、もしくはいずれかに責任を持つ、個人、組織、グループ、または他の法的主体」津谷喜一郎 編集ガイドラインとCOI. in: 第7回日本医学雑誌編集者会議（JAMJE）第7回シンポジウム「不正論文と編集者の責務」、slide 24

8 国際開発分野のインパクト評価のエビデンス情報の集積と提供を行うInternational Initiative for Impact Evaluation（3ie）は2009年より国際開発分野としてC2に合流した。

9 だが、雑誌リストの前書きにこの長いリストは完全なwhite listではなく、悪徳雑誌（predatory journal）も入っている可能性があるとしており、日本医学会も同様の注意喚起を行っている。

10 ICMJE Recommendationsの改訂履歴（Archives）はWeb上で確認可能である。

11 14の雑誌と2つの機関がある。5大雑誌は、通常、米国のNew England Journal of Medicine、JAMA（Journal of American Medical Associationの略だがJAMAが正式名称）、Annal of Internal Medicine －の3つと、イギリスのLancetとBMJ（こちらもBritish Medical Journalの略でBMJが正式名称）－の2つである。いずれにしろ、EBPMのために日本の各領域における指導的雑誌の質向上が期待され、その投稿規定にCONSORTなどの基本的RGの使用が含まれることが望まれる。

12 医学領域のPubMedにindexされる狭義のシステマティック・レビューの総数は2013年末で約75,000件と推計された（唐ら 2014）。世界の多くの研究方法論者（methodologist）が待っていたsystematic reviewのPublication Type（PT）が昨2019年にようやく作成された。PubMedでそのPTを用いてsystematic reviewを1950.1.1から2019.12.31までとして検索すると約124,000件、2019年の1年間で13,000件である。急速に大量のsystematic reviewが「つくら」れていることは間違いない。

13 2019年1月末時点の各都道府県の議会での「エビデンス」の用語の頻度分析「地方議会における EBPM（証拠に基づく政策立案）の現状と課題（堀江 2019）」の報告がある。

14「エビデンスに基づく政策展開の推進」調査研究チーム 最終報告書（平成31年3月）」（北海道調査研究チーム 2019）。

参考文献

家子直幸・小林庸平・松岡夏子・西尾真治（2016）「エビデンスで変わる政策形成～イギリスにおける「エビデンスに基づく政策」の動向、ランダム化比較試験による実証、及び日本への示唆～」、『三菱UFJリサーチ&コンサルティング政策研究レポート』、https://www.murc.jp/wp-content/uploads/2016/02/seiken_160212.pdf（閲覧日 2019-9-30）

奥村泰之・吉田和樹・清水沙友里（2014）「臨床疫学研究における報告の質向上に向けて：研究者の倫理.」、『精神科』、24（5）：551-557

上岡洋晴・津谷喜一郎（訳）（2008）「疫学における観察研究の強化（SROBE声明）：観察研究の報告に関するガイドライン」、中山健夫・津谷喜一郎編著（2016）『臨床研究と疫学研究のための国際ルール集Part2』、ライフサイエンス出版、202-209

上岡洋晴・中山健夫・津谷喜一郎（2014）「EQUATOR Networkから得られる，質の高い研究報告のための国際ルール－③ STROBE: 疫学における観察研究報告のためのルール」、『薬理と治療』、42（suppl.2）：s109-117

上岡洋晴・津谷喜一郎（訳）（2016）「CONSORT グループの方法とプロセス: 非薬物療法評価試験のための拡張版」、中山健夫・津谷喜一郎編著『臨床研究と疫学研究のための国際ルール集Part 2』、ライフサイエンス出版、70-9

萱間真美（2015）「論文の査読ガイドラインの構築に向けて－誰のためのガイドラインか」、『看護研究』、48（7）：642-7

佐藤俊哉（2007）「人間栄養学講座 連載 ランダム化臨床試験をするまえに【第4回】ランダム化の方法」、『栄養学雑誌』、65（5）：255-60

総務省（2018.10）『EBPM（エビデンスに基づく政策立案）に関する有識者との意見交換会報告（議論の整理と課題等）』、EBPMに関する有識者との意見交換会事務局 http://www.soumu.go.jp/main_content/000579366.pdf（閲覧日 2019-9-30）

統計改革推進会議（2017）『統計改革推進会議：最終とりまとめ』（平成29年5月19日統計改革推進会議決定）、https://www.kantei.go.jp/jp/singi/toukeikaikaku/pdf/saishu_honbun.pdf（閲覧日 2019-9-30）

中嶋宏（監修）・津谷喜一郎・坂巻弘之・山崎茂明（2000）『EBMのための情報戦略－エビデンスをつくる、つたえる、つかう』、中外医学社

中山健夫・津谷喜一郎（2008）『臨床研究と疫学研究のための国際ルール集』、ライフサイエンス出版

中山健夫・津谷喜一郎（2016）『臨床研究と疫学研究のための国際ルール集〈part2〉』、ライフサイエンス出版

津谷喜一郎・佐久間昭（1992）「臨床試験論文チェックリストのチェック－情報医学的観点から－」、（1993）『臨床薬理』、24（1）：49-50

津谷喜一郎・澤田弘（2007）「臨床試験のsponsorはスポンサーではない」、『臨床薬理』、38 Supple：S268

津谷喜一郎・元雄良治・中山健夫（2010）「CONSORT 2010声明：ランダム化並行群間比較試験報告のための最新版ガイドライン」、『薬理と治療』、38（11）：

939-947
津谷喜一郎（2010）「CONSORT 2010 訳者解説」、『薬理と治療』、38（11）：948-9
津谷喜一郎（2015）『いろいろな分野のエビデンス－温泉から国際援助までの多岐にわたるRCTやシステマティックレビュー』、ライフサイエンス出版
津谷喜一郎・中澤幾子・佐藤元（2017a）「WHO 臨床試験登録データセット（Version 1.2.1, Version 1.3）」、『薬理と治療』、45（12）：1905-10
津谷喜一郎（2017b）「WHO 臨床試験登録データセット解説」、『薬理と治療』、45（12）：1911-6
津谷喜一郎・中澤幾子・佐藤元.（2018a）「WHO 臨床試験登録データセット（Version 1.3）の背景となる4つの資料の日本語訳）」、『薬理と治療』、46（5）：669-682
津谷喜一郎・中澤幾子・佐藤元（2018b）「WHO 臨床試験登録データセット（Version 1.3）の背景となる4つの資料の解説」、『薬理と治療』、46（6）：965-712
津谷喜一郎・上岡洋晴・折笠秀樹・佐藤元（2019）「2017 CONSORT 非薬物介入版の紹介と解説－2017 CONSORT NPT Extension－」、『薬理と治療』、47（6）：865-84
津富宏訳・津谷喜一郎監訳（2008）「CONSORT声明: クラスターランダム化試験への拡張」、中山健夫・津谷喜一郎 編著（2008）『臨床研究と疫学研究のための国際ルール集』、ライフサイエンス出版、105-17
唐文涛・小島原典子・河合富士美・津谷喜一郎（2014）「診療ガイドラインとシステマティック・レビュー」、『薬理と治療』、42（3）：189-197
福原俊一・山口拓洋・山崎新・林野泰明・竹上未紗（2009）『観察的疫学研究報告の質改善（STROBE）のための声明：解説と詳細』、（2009年5月28日　日本語翻訳検討グループ翻訳）
正木朋也・津谷喜一郎（2006）「エビデンスに基づく医療（EBM）の系譜と方向性: 保健医療評価に果たすコクラン共同計画の役割と未来」、『日本評価研究』、6（1）：3-20
正木朋也・津谷喜一郎（2010）「エビデンスに基づく医療（EBM）の展開から学ぶもの－EBM普及に伴い経験した課題と解決策－」、『日本評価研究』、10（1）：3-16
増島祥・正木朋也・唐文涛・津谷喜一郎（2013）「日本と世界におけるクラスターランダム化比較試験の現状と課題」、『第34回日本臨床薬理学会学術総会』、東京
Altman, D. G., Schulz, K. F., Moher, D., Egger, M., Davidoff, F., Elbourne, D., Group, f. t. C.（2001）. The Revised CONSORT Statement for Reporting Randomized Trials: Explanation and Elaboration. *Annals of Internal Medicine*, 134（8）, 663-694.［津谷喜一郎、栗原千絵子（訳）（2002）ランダム化比較試験報告のためのCONSORT声明改訂版－解説と詳細、『臨床評価』、30（1）: 163-208］
Begg, C., Cho M., Eastwood, S., et al（1996）. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA* 276（8）: 637-9.［日本語訳：津谷喜一郎・小島千枝訳（1997）「無作為化比較試験の報告の質を改善する方法CONSORT声明」、『*JAMA*〈日本語版〉1997年7月号』、74-7］
Boutron I., Moher D., Altman D.G., et.al, CONSORT Group.（2008a）. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. *Ann Intern Med*, 148: 295-309.
Boutron I., Moher D., Altman D.G., et.al, CONSORT Group.（2008b）. Methods and processes of the CONSORT Group: example of an extension for trials assessing nonpharmacologic treatments. *Ann Intern Med*, 148:W60.
Boutron I.B., Altman D.G., Moher D., et.al for the CONSORT NPT Group.（2017）. CONSORT Statement for randomized trials of nonpharmacologic treatments: a 2017 update and a CONSORT Extension for nonpharmacological trial abstracts. *Ann Intern Med*, 167: 40.
Cameron, D.B., Mishra, A., Brown, A.N.（2015）. The growth of impact evaluation for international development: how much have we learned? *Journal of Development Effectiveness* 8: 1-21.
Campbell M.K., Elbourne D.R., Altman D.G.（2004）. CONSORT statement: extension to cluster randomised trials. *BMJ*; 328:702-8.
Campbell, M.K., Piaggio, G., Elbourne, D.R., et.al & Group, C.（2012）. Consort 2010 statement: extension to cluster randomised trials. *BMJ*, 345, e5661. doi:10.1136/bmj. e5661
Cartwright, N., & Munro, E.（2010）. The limitations of

randomized controlled trials in predicting effectiveness. *Journal of Evaluation in Clinical Practice*, 16(2), 260-266.

Chalmers, I., Bracken, M. B., Djulbegovic, B., et.al.（2014）. How to increase value and reduce waste when research priorities are set. *Lancet*, 383（9912）, 156-165.

Chan, A.-W., Tetzlaff, J. M., Altman, D. G., Laupacis, A., Gøtzsche, P. C., Krleža-Jerić, K., Moher, D.（2013）. SPIRIT 2013 Statement: Defining Standard Protocol Items for Clinical Trials. *Annals of Internal Medicine*, 158(3), 200.［日本語訳: 折笠秀樹・津谷喜一郎・上岡洋晴（2017）「SPIRIT 2013 声明：臨床試験のための標準的なプロトコール項目の規定」、『薬理と治療』245(12): 1895-1904］

Chan, A. W., Song, F., Vickers, A., et.al.（2014）. Increasing value and reducing waste: addressing inaccessible research. *Lancet*, 383（9913）, 257-266. doi:10.1016/S0140-6736(13)62296-5

Chandler, J., Hopewell, S.（2013）. Cochrane methods - twenty years experience in developing systematic review methods. *Systematic Review*, 2(1), 76.

Clark, H. D., Wells G. A., Huët C., et al.（1999）. Assessing the quality of randomized trials: reliability of the Jadad scale. *Controlled Clinical Trials*, 20(5): 448-52.

Collins R, Browman L, Landray M, et. al(2020). The magic of randomization versus the myth of real-world evidence. *New England Journal of Medicine*. 382：674-678.

Deaton, A., & Cartwright, N.（2018）. Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine*, 210, 2-21.

Eble A., Boone P., Elbourne D.（2017）. On minimizing the risk of bias in randomized controlled trials in economics. *The World Bank Economic Review*, 31(3): 687-707.

Fontanarosa, P. B., Flanagin, A., & DeAngelis, C. D.（2010）. Implementation of the ICMJE form for reporting potential conflicts of interest. *JAMA*, 304(13), 1496.

Higgins, J. P. T., Altman, D. G., Gøtzsche, P. C., et. al.（2011）. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*, 343: d5928.

Huth J., E., Case, K.（2004）. The URM: Twenty-five years old. *Science Editor*, 27(1), 17-21.

Iimura K., Watanabe N., Masunaga K., et al.（2016）. Effects of a gentle, self-administered stimulation of perineal skin for nocturia in elderly women: A randomized, placebo-controlled, double-blind crossover trial. *PLOS ONE*, 11(3), e0151726.

International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing and Publication of Scholarly Work in Medical Journals（ICMJE Recommendations, formerly the Uniform Requirements for Manuscripts）. http://www.icmje.org/recommendations/archives/（閲覧日 2019-9-30）

Ioannidis, J. P., Greenland, S., Hlatky, M. A., et. al.（2014）. Increasing value and reducing waste in research design, conduct, and analysis. *Lancet*, 383(9912), 166-175.

Jadad A. R., Moore R. A., Carroll D., et .al.（1996）. Assessing the quality of reports of randomized clinical trials: is blinding necessary?. *Control Clin Trials*. 17(1), 1-12.

Macleod, M. R., Michie, S., Roberts, I., et.al.（2014）. Biomedical research: increasing value, reducing waste. *Lancet*, 383(9912), 101-104.

Melander H., Ahlqvist-Rastad J., Meije G., et.al.（2003）. Evidence b(i)ased medicine－selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. *BMJ*, 326, 1171-5.

Moher D., Schulz K. F., Altman D. for the CONSORT Group.（2001）. The CONSORT statement: revised recommendations for improving the quality of parallel-group randomized trials. *JAMA*, 285(15), 1987-1991［津谷喜一郎・小島千枝・中山健夫訳（2002）「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定版勧告」、『JAMA〈日本語版〉2002年6月号』、74－7］

Moher D., Jones A., Lepage I., for the CONSORT Group.（2001）. Use of the CONSORT statement and quality reports of randomized trials: a comparative before-and-after evaluation. *JAMA*, 285, 1992-5.

Moher, D., Hopewell, S., Schulz, K. F., et.al.（2010）. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ*, 340, c869.

Niforatos J. D., Weaver M., Johansen M. E.（Published online July 29, 2019）. Assessment of publication trends of systematic reviews and randomized clinical trials, 1995 to 2017. *JAMA Intern Med*.

Schulz, K. F., Altman, D. G., Moher, D., & Group, C.（2010）. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ*, 340, c332. doi:10.1136/bmj. c332［日本語訳: 津谷喜一郎・元雄良治・中山健夫訳（2010）「CONSORT声明: ランダム化並行群間比較試験報告のための最新版ガイドライン」、『薬理と治療』、38（11）: 939-947］

Simera, I., Altman, D. G.（2009）. Writing a research article that is "fit for purpose": EQUATOR Network and reporting guidelines. *Evidence-based medicine*, 14（5）, 132-4.

van der Graaf, R., Koffijberg, H., Grobbee, D. E., et al.（2015）. The ethics of cluster-randomized trials requires further evaluation: a refinement of the Ottawa Statement. *J Clin Epidemiol*, 68（9）, 1108-1114.

Vandenbroucke J. P, von Elm E., Altman D. G., et. al. STROBE Initiative（2007）. Strengthening the Reporting of Observational Studies in Epidemiology（STROBE）: Explanation and Elaboration. *Epidemiology*,18, 805-835.

von Elm E., Altman D. G., Egger Mattjoas, et. al.（2007）. Strengthening the Reporting of Observational Studies in Epidemiology（STROBE）. *PLoS Medicine*, 4（10）, 1623-1627.

Weijer C., Grimshaw J.M., Eccles M.P., et al.（2012）. The Ottawa statement on the ethical design and conduct of cluster randomized trials. *PLoS Med*, 9, e1001346.

（2020.5.25 受理）

Reporting Guidelines for Evidence-Based Policymaking (EBPM) and Improving Research Paper Quality

Tomoya MASAKI
Japan International Cooperation Agency（JICA）;
Kitasato University
masakit-tky@umin.ac.jp

Kiichiro TSUTANI
Tokyo Ariake University of Medical
and Health Sciences
tsutanik@gmail.com

Abstract

In Japan, since the late 2010s, “evidence-based policy-making（EBPM）” has been introduced into the public sector. In the medical field, the CONSORT statement was developed in 1996 to improve the quality of Randomized Control Trial（RCT）articles. It was the pioneer of reporting guidelines. The third edition of CONSORT 2010, consisting of all 25 items, has been translated into 16 languages and is widely used throughout the world. In addition, over 400 reporting guidelines have been developed for basic study designs such as observational studies（cross-sectional, retrospective, and prospective studies）and systematic reviews, as well as expanded versions such as cluster randomization and nonpharmacologic treatment（NPT）. We also introduced the EQUATOR Network, which covers many reporting guidelines that will be useful for people starting to create research plans. Together, these findings suggest the use of reporting guidelines is a feasible option for EBPM.

Keywords

EBPM, reporting guidelines, CONSORT, EQUATOR Network, STROBE

【総説】

エビデンスに基づく政策立案（EBPM）の推進に向けて
ー医療の経験からの示唆ー

田辺　智子
国立国会図書館
stanabe0509@gmail.com

要　約

EBPMの起源は「エビデンスに基づく医療（EBM）」にあるとされる。医療では政策立案と異なる理想的な状況下でエビデンスの創出や利用が可能だと思われがちであるが、実は医療においてもエビデンスのみに基づいて意思決定ができるわけではなく、またエビデンスと現実の診療にギャップがあることが課題とされてきた。本稿では、医療におけるエビデンス活用の経験を概観し、EBPM推進に当たって参考となる知見を抽出した。その結果、①EBMではエビデンスとは医療行為の有効性評価だと理解されていること、②エビデンスは社会全体で蓄積し共有すべきであること、③エビデンスの伝達が重要であること、④エビデンスは意思決定に用いられる情報の一つに過ぎないこと、⑤エビデンスは必ずしも容易に普及するわけではないこと、⑥厳密な有効性評価と業績測定を適切に併用すべきこと、⑦政府にはEBPMのインフラ整備が求められること、の七点の示唆が得られた。

キーワード

エビデンスに基づく政策立案（EBPM）、エビデンスに基づく医療（EBM）、エビデンス、
ランダム化比較試験、医療の質

1．本稿の背景と目的

「エビデンスに基づく政策立案」（Evidence-Based Policy Making：EBPM）が関心を集めている。日本評価学会では10年以上前からエビデンスに基づく実践について学会誌や大会で繰り返し取り上げ、議論を行ってきた。当初は日本ではなじみの薄かったこの概念も、ここ数年で急速に人口に膾炙し、今や国でも自治体でもEBPMは大きな潮流となっている。その一方、エビデンスやEBPMについての理解は人によって大きく異なり、議論や目指す方向が噛み合わないまま、EBPMという名の下に多様な取組が併存している感も否めない。

こうした中、EBPMのポイントを捉え直し、理解を深める上では、医療における経験が参考になる。よく知られているように、EBPMの起源は「エビデンスに基づく医療」（Evidence-Based Medicine）にある。しかし、EBMがどのようなものなのか、その具体像は意外に知られていないように思われる。多くの人は、医療がエビデンスに基づくのは当然のことであり、EBMでは政策立案とは異なる理想的な状況でエビデンスが産出され、日常医療にエビデンスが浸透してい

日本評価学会『日本評価研究』第20巻第2号、2020年、pp.19-31

ると思っているのではないだろうか。しかし実際は、医療といえども常に厳密な有効性評価に基づくわけではなく、エビデンスを基に機械的に意思決定ができるわけでもない上、エビデンスの裏付けがある医療が思うように普及しないことが課題と認識されてきた。そもそもEBMという概念が登場したのは1990年代と比較的新しく、その確立・普及の過程では多くの誤解も生じた。

こうした医療の経験から、EBPMが学べることは多い。本稿の目的は、医療分野におけるEBMやエビデンス普及の取組について概観し、EBPMへの示唆を抽出することである。第2章ではEBMの基礎概念、第3章では医療におけるエビデンスの普及、第4章では特に日本の医療における状況をまとめた上で、第5章でEBPMをめぐる海外及び日本の現状を整理し、第6章でEBMから得られるEBPMへの示唆をまとめる。

2. EBMの概要

(1) EBMとは何か

「エビデンスに基づく医療」(Evidence-Based Medicine: EBM) とは、患者への医療行為の意思決定において、その時点で得られる最善のエビデンス（科学的根拠）を利用しようとする医療のあり方である。後述するように、その際はエビデンスに加え、患者の病状・環境や意向も考慮に入れられる。EBMは1991年に提唱された後[1]、それまでの経験主義的な医療に変わる新しいパラダイムとして急速に普及し、医療の世界を一新したと言われている。

では、「エビデンス」とは何か。EBMにおけるエビデンスとは、治療・予防等の医療行為の有効性についての信頼性の高い研究結果だと理解されている。ここでポイントとなるのは、「有効性」についての研究だという点である。医療に関する研究には様々なものがあり、たとえばある病気の発症原因やメカニズムに関する基礎研究、罹患率や予測因子等を明らかにする疫学研究もある。しかし、それらをすべて「エビデンス」と呼ぶわけではなく、あくまで治療や予防の「有効性」に関する研究がエビデンスと呼ばれている。しかも、実験室でマウスを使って行うような基礎研究ではなく、人間、すなわち実際の患者を被験者として行う「臨床研究」と呼ばれるものがエビデンスとなる。

(2) エビデンスの強弱

臨床研究の方法には様々なものがあり、いずれを採用するかによって結果の信頼性は大きく異なる。よく知られているように、有効性を最も厳密に評価できるのは、対象者を無作為に介入群と非介入群に分けて比較を行うランダム化比較試験（Randomized Controlled Trial: RCT）である。ただし、研究対象によっては、このような実験が常に可能なわけではない。たとえば、外科手術では医師や患者による条件の差が大きく、RCTが難しいとされる（星 2013）。このため、観察研究と呼ばれる方法も多く用いられる。観察研究には、集団を継続して観察することで疾病と治療法等の関係を分析するコホート研究と、ある集団を対象に過去に遡って調査することで同様の関係性を分析する症例対照研究（ケースコントロール研究）がある。これらのほか、症例報告や、多数の症例を集めた症例集積等もあるが、信頼性は低いとされる。

さらにEBMでは、単一の臨床研究より、複数の臨床研究を統合することによって、より信頼性の高いエビデンスが得られると考える。RCTは、対象集団について厳密に因果関係を検証できる方法であるが（これを「内的妥当性」という）、その結果が別の集団にも当てはまるか（「外的妥当性」という）については必ずしも保証されない。このため、異なる集団を対象とした複数のRCTを統合することで、より一般化可能な知見が得られると考えられている。EBMでは、既存の研究を体系的に収集して質を批判的に検証し、信頼性の高いもののみを残して結果を統合した「系統的レビュー」(Systematic Review) というメタ分析が、最上位のエビデンスとされる。

(3) EBMの背景

このように、EBMが人間を対象とした臨床研究、特にRCTを重視する背景には、かつては有効

だと信じられ実施されていた医療行為の中に、RCTを実施して初めて効果がない、あるいはかえって有害だと判明した事例があったことがある。たとえば、かつては未熟児に酸素を投与する処置が一般的に行われていたが、臨床研究の結果、酸素を投与すると、投与しない場合に比べ失明が有意に増加することが判明し、実施されなくなった（トーガーソン・トーガーソン 2010、pp.5-6）[2]。

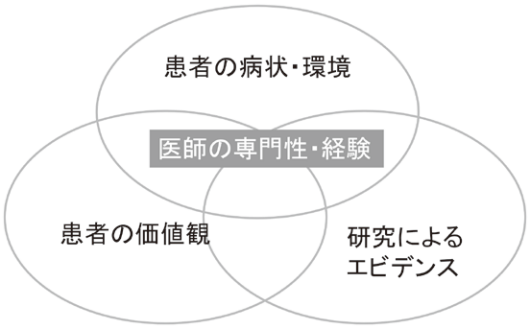
なお、RCTという手法自体は古くからあったものであり、特に新規の医薬品の承認を得るためには治験と呼ばれる臨床研究による有効性の証明が求められてきた。EBMは、そうした臨床研究を、より一般的に医療行為全体に適用しようとするものといえる。

（4）EBMの意思決定

EBMをめぐっては、当初、医療者の経験や専門性よりもエビデンスを重視するものといった誤解も見られた。また、EBMというと、エビデンスを基に機械的に意思決定を行うマニュアル医療のようなイメージを持たれる場合もある。しかし今日では、EBMにおける意思決定は図1のように理解されている。そこでは、「研究によるエビデンス」に加え、「患者の病状・環境」、「患者の価値観」という各要素を統合し、医療者が専門性･経験を活かして意思決定を行うとされる（中山 2014、p.373）。つまり、エビデンスは意思決定に用いられる情報の一つに過ぎない。

たとえば、ある治療法に明確なエビデンスが存在したとしても、それを目の前の患者に適用できるかについては専門的な判断が必要となる。若い男性を対象としたRCTの結果を女性や高齢男性に適用してよいかという判断や、研究上のコントロールのため合併症のない患者のみを対象とする場合が多いRCTを、現実の医療現場ではむしろ一般的な、糖尿病でかつ高血圧といった複数疾患を持つ患者に適用してよいかといった判断である。また、仮に適用可能なエビデンスがあったとしても、患者の意向、通院の可否等の社会的背景やコストも意思決定に影響する。そういった点を考え合わせて患者にとって最善の選択を行う上では、依然として医師の経験と技能が必要と考えられている。

図1　EBMにおける意思決定



（出所）中山健夫（2014）「診療ガイドラインの真の役割を説く」、『新医療』、41（3）を参考に筆者作成

（5）EBMの三局面

EBMを社会全体で推進することを考えた場合、エビデンスを「つくる」「つたえる」「つかう」の三つの局面が想定できる（中山 2010、p.259）。最初の「つくる」局面は、臨床研究によって信頼性の高いエビデンスを産出する段階である。三番目の「つかう」局面は、病院等の医療現場で医師がEBMを実践する段階である。その間をつなぐ「つたえる」は、産出されたエビデンスを現場の医師に仲介する役割ということになる。

医学は日進月歩であり、膨大な研究が日々生み出されるため、多忙な医師が自らの専門領域の最新知識を常に入手し続けるのは容易なことではない。このため、医師に代わって最新の研究を収集し、わかりやすい形で医療関係者に伝える役割が必要とされる。

このために重要となるのが、診療ガイドラインである。診療ガイドラインとは、診療上の重要度の高い医療行為について、患者と医療者の意思決定を支援するため、系統的レビュー等をもとに最善と考えられる推奨を示す文書であり（福井・山口 2014、p.3）、学会等によって疾病・医療分野ごとに作成される。診療ガイドラインはあくまで指針であって医師が行う医療行為を拘束するものではないが、個々の医師の信念や技能によって診断・治療が過度にばらつくのを防ぐことができる。

かつての診療ガイドラインは、権威ある医師

の経験や信念に基づいて作成される傾向があったが、近年はEBMの手順に則り、系統的レビューから推奨を導く方法に世界的に変わってきている。近年ではさらに、患者や専門外の医師等、様々な関係者の視点を反映して推奨を決めることが重視されている（上田・五十嵐・中山 2018）。こうした診療ガイドラインの作成方法を体系的にまとめたものがGRADE（The Grading of Recommendations Assessment, Development and Evaluation）システムであり、国際的なスタンダードとなっている。また、診療ガイドラインでの推奨は医薬品や医療機器の売上げに直結するため、作成に当たっては利益相反を避けた透明性・中立性の高い手順と方法が求められる[3]。

もう一つ、エビデンスを「つたえる」上で重要な役割を持つのが、コクラン[4]に代表される二次情報データベースである。コクランは、世界中で行われた臨床研究を収集し、系統的レビューを実施してコクラン・ライブラリーというデータベースで提供する国際組織である。商業的な資金提供は受けず、多くの国で中立的なNPOとして活動を行っており、130か国以上から研究者・専門家が参加している。

3. 医療におけるエビデンスの普及

（1）EBMと医療の質

医療がエビデンスに基づいて実施されるのは当然と考えられるかもしれないが、実際には病院等の医療現場で、研究が示すエビデンスどおりの診療が常に実施されているわけではない。前述のように、EBMの意思決定ではエビデンス以外の要素も考慮に入れられるため、ある治療がたとえ有効であっても患者の状況によっては採用されないケースがあり得る。しかし、その患者の状況を総合的に勘案すれば明らかに実施すべき治療法でも、採用されない場合も存在する。

この問題は「エビデンス－診療ギャップ」と呼ばれ、問題視されてきた。たとえば、1991年のEBM登場から10年余り経った2003年時点での米国の研究によれば、エビデンスに基づく推奨どおりの医療を受けていた人の割合は全体の55%に留まっていた（McGlynn, et al. 2003）。このため、2000年代以降、各国でエビデンス－診療ギャップを埋め、医療の質を高めるための取組が継続的に行われてきた。

「医療の質」というのは、実はEBMと密接にかかわる概念である。一般には「医療の質」は様々な文脈で使われる言葉であり、スタッフの対応の良し悪しや、「神の手」と呼ばれるような高度な手術技能と結び付ける向きもあるだろう。しかし、医学の専門的な見地からは、医療の質とは「個人や集団を対象に行われる医療が望ましい健康状態をもたらす可能性の高さ、その時々の専門知識に合致している度合い」（福井 2017、p.16）と定義されている。最新の専門知識に合致し、健康状態を改善する可能性が高い医療とは、すなわちエビデンスのある医療にほかならない。つまり、医療の世界においては、EBMに則った医療こそが、質の高い医療だと考えられている[5]。

（2）QIによる医療の質評価

医療の質を客観的に評価するために用いられるのが「医療の質指標」（Quality Indicator: QI）である。医療の質は、ストラクチャー（構造）、プロセス（過程）、アウトカム（結果）の三つの側面を対象に評価することができる[6]。ストラクチャーとは医療提供に用いられる施設やスタッフ等の資源、プロセスとは手術や投薬等の医療内容、アウトカムとは生存率や生活の質（Quality of Living: QOL）といった医療行為の結果としての患者の健康状態を意味する。

これらのうち、EBMに直接対応するのはプロセスの指標である。先進国では医療のストラクチャーはほぼ充足しており、またアウトカムは患者条件の影響を受け医療機関同士の比較が難しいことから、医療の質を評価する上ではプロセス指標を中核とする場合が多い。

（3）医療の質の向上策

以上をまとめると、質の高い医療を提供するためには、まずEBMで有効性の高い医療行為を明らかにした上で、それを実際の医療現場で提供する必要がある。後者の実現度合いを見るの

がQIであり、QIを測定しその改善を図ることが、医療現場で実施できる医療の質の向上策ということになる。

QI改善のためには、まず医療を直接提供する医師に対し、診療ガイドライン配布や生涯教育等で働きかけることが考えられる。しかし、研究によれば、単に知識を提供するだけでは行動変化までは生じにくい。医師が新たな医療行為を採用するためには、その内容に納得して肯定的な意見を持ち、さらに時間・人員・設備等の制約を乗り越えて、自らの行動を変える必要があるためである（東 2007、p.543；鄭・近藤2010、p.932）。

このため、QIを改善する上では、医師個人の努力を超えた組織的・体系的な取組が必要だと考えられている。その手法としては、日本の自動車業界等、産業界で発展した品質改善手法が広く用いられている。医療の質の測定結果を定期的に医師にフィードバックすることで行動変容を促すクリニカル・オーディット、改善のためのチームを組織し継続的に質を改善するPDCAサイクル等、様々な手法がある。QI改善の具体例としては、糖尿病患者の血糖管理の改善、救急外来待ち時間の短縮、院内感染の新規発生率の減少等がある（小松ほか 2016）。個別の医療機関の改善事例としては、米ハーバード大学附属のマサチューセッツ総合病院において、院内感染予防の基本となる医療スタッフの手指消毒の徹底にPDCAサイクルを活用し、2003年に8%だった遵守率を2009年には90%以上に改善させた例等が報告されている（小松 2012b、p.215）。

さらに、政策的に医療の質を向上させるためには、医療機関へのQI公表の義務付けや、QIの測定結果を基に医療提供者に経済的インセンティブを与える「実績に基づく支払」（Pay for Performance: P4P）等の取組があり、諸外国で普及が進んでいる（田辺 2019）。

4. 日本におけるEBMとエビデンス普及

（1）EBMの導入

わが国では、EBMについては1990年代前半のEBM創成期から一部研究者が紹介を始め、1997年にEBMに関する最初のセミナーが開催されている（津谷 2010、p.313）。1998年には、厚生省（当時）内に置かれた「医療技術評価推進検討会」で、EBMを推進する一方策としての診療ガイドラインのあり方が検討された。これを受け、厚生労働省は、厚生労働科学研究費補助金を用いて学会等による診療ガイドライン作成の支援を行った。この結果、多くの疾病や医療分野で診療ガイドラインが整備された（吉田 2018）。

また、厚生省では、診療ガイドラインや質の高い医療情報を医療提供者等に届けるためのEBMデータベースの整備も行った。公益財団法人日本医療機能評価機構が2013年から提供している「Minds」というサイトがそれに当たる[7]。

ただし、当初はEBMに則った診療ガイドライン作成方法が確立しておらず、初期の診療ガイドラインはエビデンスというよりは専門家の合意に基づくものだったとされる。その後、Mindsが2007年に「診療ガイドライン作成の手引き」を公表し、数次の改定を経て現在ではGRADEシステムに完全準拠したものとなった。これを受け、国内でも質の高い診療ガイドラインが少しずつ作成されるようになったとされる（南郷 2019、p.21）。

（2）臨床研究の不足

診療ガイドラインを整備する過程で明らかになったのは、日本人を対象とした質の高いエビデンスが極めて乏しいという現実であった（中山 2010、p.260）。このため、診療ガイドライン作成で引用された臨床研究の大半は、諸外国からのものとなった。この背景として、日本では伝統的に基礎医学を重視する傾向が強く、臨床研究が弱かったことがあると指摘されている（福井 2011、p.1676）。

2000年代以降は、経済成長戦略として基礎研究を実用化につなげる研究支援が重視されたこともあり、厚生労働省と文部科学省は共同で治験・臨床研究の活性化を進めてきた（厚生科学審議会臨床研究部会2019）。

(3) エビデンスの普及と医療の質評価

日本の医療については、平均的に見たパフォーマンスは高いものの、病院間の医療技術の格差が大きいと指摘されている（桐野・松田 2015、p.782）。病院ごとに治療方法、手術の術式、在院日数等が大きく異なり、手術後の合併症や死亡率等のアウトカムでもばらつきが大きいとされる（よしかわ 2017、pp.18-22、40-42）。このことは、病院によってはエビデンスに基づく標準治療とのかい離があること、すなわちエビデンス－診療ギャップが存在することを示唆している。

エビデンス－診療ギャップを埋めるには、まずどれだけのギャップがあるかを測定することが出発点となり、そのために用いられるのがQIによる医療の質評価である。しかし、わが国では医療の質評価の取組は諸外国に比べ遅れを取っており、非常に熱心に取り組む病院もある一方で、無関心な病院もあり温度差があるとされる（桐野・松田 2015）。

単一病院による先進的な取組事例としては、聖路加国際病院がある。同院では2005年から医療の質評価を実施しており、PDCAサイクルの導入によって、患者の転倒・転落防止、中心静脈カテーテル挿入時の合併症減少等の改善を実現している（福井 2015）。

複数病院による先進事例としては、国立病院機構が2005年頃から系列病院の質改善に取り組んでいる。一例として、急性脳梗塞で入院後4日以内にリハビリを開始した割合が、2010年には病院により20 ～ 90%であったが、2013年には大部分の病院で70%以上に改善したと報告されている（伏見 2015）。同機構では詳細なQI測定方法のマニュアルを公表しており、他の病院・病院群でも活用されている。こうした複数病院による医療の質評価を支援するため、厚生労働省は2010年度から「医療の質の評価・公表等推進事業」を行っている。

5. EBPMの現状

EBPMとは、以上で見た医療のEBMを、政策立案に応用したものと理解できる。そのモデルは、先進事例とされる英国と米国の取組に求められる[8]。

(1) 英国におけるEBPM

英国のEBPMは、1997年からの労働党ブレア政権で導入された。行政改革の方針を示した1999年の白書『政府の近代化（Modernising Government）』では、政策立案でエビデンスを活用することがうたわれた。また評価に関する各種ガイドブックが整備され、その中では有効性評価の方法としてはRCTが最も強力であるとされた。政府が実施したRCTとしては、失業者や低賃金労働者のための就業継続プログラムの評価が成功例とされる（家子ほか 2016、p.11）。また、行動経済学を応用して政策効果の向上を目指すナッジ・ユニット[9]という組織が多くの政策効果検証を行い、コスト削減を実現する成功を収めた。

英国の取組で特徴的なのは、エビデンス活用を促進する官民共同組織、エビデンス・センター（What Works Centre: WWC）の整備である。エビデンスの創出・伝達・適用の支援を行う組織であり、医療、教育、犯罪防止、地域経済活性化、福祉等、政策分野ごとに設置されている。

(2) 米国におけるEBPM

米国では、1960年代という早い時期から、教育、保健医療、社会福祉等の分野でRCTによる厳密な有効性評価が実施されてきた経緯がある。しかし、それが「エビデンス」と呼ばれ、施策への反映が重視される契機となったのは、2001年教育改革法であった。この法律では、教育現場に科学的根拠のある指導法の採用を求めており、エビデンスに関する情報提供サイトとしてWhat Works Clearinghouseが設置された（田辺 2006）。

政府全体としてのEvidence-Basedの取組が開始されたのは、2009年発足のオバマ政権からである。オバマ政権では予算編成においてエビデンスを重視する方針が採られ、RCT等による質の高い有効性評価の実施が推奨された。予算編成でのエビデンス重視は、その後のトランプ政権でも踏襲されている（大槻 2018、p.31）。このほか、エビデンスの信頼性の程度に応じて補助金を配

分する階層別補助金（Tiered grant）、社会的課題の解決に民間事業者の能力を活用し厳密な有効性評価と成功報酬を行うPFS（Pay for Success）、英国の経験にならい行動経済学を政策に応用することを目指すナッジ・ユニット[10]の設立等の取組が見られる。

2016年には、超党派の支持により「エビデンスに基づく政策のための諮問委員会法」[11]が成立し、2017年には同法に基づく最終報告書がまとめられた。データへの安全なアクセス、個人情報保護、データインフラ近代化、エビデンス創出能力の強化に関する勧告が含まれており、これを具体化するために「エビデンスに基づく政策形成基盤法」[12]が2019年1月に成立した。

（3）日本におけるEBPM

わが国では、2010年代に入ってから、RCTによる因果推論や政策効果検証の必要性を説く啓蒙書（西内 2013；中室 2015等）が相次いで出版され、エビデンスやEBPMへの関心が高まる契機となった。

国レベルでは、2016年から経済財政諮問会議や与党の会議体における統計改革の議論の中で、エビデンスやEBPMが議論の俎上に載った。2017年5月の統計改革推進会議の最終取りまとめでは、政府全体でのEBPM推進体制の整備が盛り込まれた。これを受け、各府省にEBPM統括責任者が設置され、同職を構成員とするEBPM推進委員会が設置された。政府文書によれば、EBPMを推進する上では、行政事業レビュー、政策評価、経済・財政再生計画の点検・評価が「三本の矢」であると説明されている（内閣官房行政改革推進本部事務局 2018）。EBPM推進委員会では、EBPM推進のためのデータ活用や人材育成が検討されたほか、各府省においてはEBPMの実例創出が進められている。

こうした中、広島県や横浜市等、自治体でもEBPMに積極的に取り組む例が出ている。

（4）日本における問題状況

以上のように、日本におけるEBPMはここ数年で急速に広がりを見せているが、その展開を見ると、先に見た英米の事例とは性格がかなり異なることに気付く。英米の例では、“what works”（何が有効か）という表現がしばしば用いられるように、政策の有効性への関心が高く、有効性を評価する上ではRCTのような厳密な手法を活用しようとする点が一つのポイントとなっている。これに対し、日本のEBPMでは、統計改革からスタートしたこともあり、政策効果への関心よりはデータの利活用と関連付けた取組が多く見られる。また政府のEBPMでは、行政事業レビューや政策評価といった既存の評価制度をベースとしており、特にロジックモデルを作成することが主な作業内容となっている。

杉谷（2019）は、こうしたわが国独自の展開を「日本型EBPM」と称し、その特徴として、第一に開始当初は経済政策が議論の中心であったこと、第二に行政事業レビューという既存制度を活用することでその対象を拡大したこと、第三に研究活用の性格が強い英米と異なり研究とのつながりが弱いことを挙げている。

こうした日本独自の展開となった背景として、エビデンスやEBPMに関する捉え方が人によって大きく異なることが挙げられる（田辺 2018）。政府での関心の高まりと呼応し、マスコミや有識者がEBPMについて取り上げることも格段に増加したが、エビデンスを因果推論を伴う有効性評価に限定する立場と、より広い意味での研究やデータもエビデンスと捉える立場が存在する。また、何をすればEBPMを実践したといえるかについても、合意があるとはいえない。そうした点について議論を整理する上で、先に見たEBMの例は、一つの座標軸となりうると考えられる。

6. 日本のEBPMへの示唆

医療の経験から得られる日本のEBPMへの示唆は、以下のとおりである。

（1）エビデンスとは何か

一点目は、「エビデンス」という用語の理解についてである。EBMにおいては、エビデンスとは基本的に、医療行為の「有効性」に関する信頼性の高い研究を意味する。その信頼性の度合

に応じてエビデンスには強弱があるものの、その主眼はあくまで「有効性」にある。また、医療においても誤解されがちな点であるが、エビデンスは常にRCTでなければならないわけではなく、入手できる範囲で最良のエビデンスを用いればよいと考えられている（岩田 2014）。つまり、重要なのは「RCTかどうか」ではなく、「有効性に関する情報であるか」という点にある。

この点で英米におけるEBPMを見ると、“what works”、すなわち何が有効かを重視する点はEBMと呼応しているといえる。対して日本のEBPMでは、エビデンスの範囲を広く理解する傾向があり、エビデンスに研究全般や統計データを含める論者も見られる。

たしかに、EBPMの場合はEBMと異なり、英米においてもエビデンスは必ずしも有効性のみに限定されず、広く定義されていると指摘される（小林 2018等）。たとえば、家子ほか（2016）では、エビデンスには「現状把握のためのエビデンス」と「政策効果把握のためのエビデンス」があると整理されている。しかし、主眼はあくまで後者の政策効果把握のエビデンスにあり、前者の現状把握のためのエビデンスは、後者の補完として併用するものと理解すべきであろう。なぜならば、「現状把握のためのエビデンス」であれば、程度の問題こそあれ、従来の政策立案でも利用されてきたはずだからである。それに加えて、「政策効果把握のためのエビデンス」を積極的に用いる点こそが、EBPMの核心と考えられる。その際、政策効果を測る方法としては、必ずしもRCT等の厳密な手法でなければならないわけではなく、実施可能な方法の中で最善のものを用いればよいと理解できる。

（2）エビデンスの共有

二点目は、エビデンスは社会全体で蓄積して共有するものだという点である。医師がEBMを実践する場合、その都度、RCTを実施して治療の効果を検証するわけではない。医療界では既存研究で明らかになったエビデンスが共有されており、それを個々の医師が参照している。その際は、単一のRCTよりもエビデンスの総体が重視されており、複数の研究を統合した系統的レビューが最上位のエビデンスとされる。つまり、蓄積されたエビデンスを社会全体で共有し、統合することで、より外的妥当性の高いエビデンスを産出し、それを医療界で活用しているのである。

これに対し日本の現状のEBPMでは、省庁や自治体がそれぞれ、自らの施策についてエビデンスを作ろうとしている。これはEBMに例えるならば、多忙な現場の医師や病院が自前でRCTを実施しているようなものである。たしかに既存のエビデンスの蓄積の薄い日本では、個々の組織が自ら政策効果を評価するのもやむを得ない面はある。また既存のエビデンスがあったとしても、社会・経済状況が異なる別の地域・集団に適用できるかは慎重に検討する必要があり、時には自らの環境下で有効性を検証し直す必要がある（青柳 2019）。しかし、今後蓄積されていくエビデンスは、組織の枠を超えて共有し、他の自治体や機関が参照できるものにしていくべきであろう。社会科学の場合、医療のようなエビデンスの統合が困難な場合もあろうが、可能であればエビデンスを統合し、より信頼性の高い知見を創出することも視野に入れる必要がある。

（3）エビデンスの伝達

二点目とも関係するが、三点目は、エビデンスを「つたえる」役割の重要性である。EBMでは、多忙な医師のためにエビデンスをレビューして提供する診療ガイドラインやデータベースが整備されている。

実は海外のEBPMでは、これに当たる取組を見ることができる。英国では各分野のエビデンス・センターが、エビデンスの創出や利用支援と並んで、わかりやすい形でのエビデンスの集積・提供を行っている。米国でも教育分野でエビデンス情報提供サイトが整備されている。国際的な取組としては、医療のコクランにならったキャンベル共同計画や、国際開発分野の3ie[13]等の例がある。

ところが、日本のEBPMにおいては、この部分の議論がすっぽり抜け落ちている。どういった機関がエビデンスを「つたえる」役割を担うのか、分野ごとに取り組むのか、分野横断的に取

り組むのか等、論点は様々考えられるが、議論は進んでいない。日本ではデータベースにできるようなエビデンスの集積がいまだないという事情はあるが、海外のエビデンスであっても、翻訳して政策現場に伝えることには一定の意味がある。英国では、先行してRCTを実施してきた米国の例を参照し、同様の施策をRCTで検証し導入した例がある（大槻 2018、p.29）。海外のエビデンスをそのまま適用はできないとしても、政策立案の参考とすることはできる。日本でもすでに書籍の形で海外のエビデンスを紹介したものはあり（経済協力開発機構 2018；ハッティ 2018）、さらに発展させて、ウェブサイト等を用いた使いやすいものとしていく余地はあろう。

（4）EBMの意思決定

四点目は、EBMにおいてすら、エビデンスは意思決定に用いられる様々な情報の一つに過ぎないという点である。先に見たように、医療ならばエビデンスを基に機械的に意思決定できるというのは誤解である。図1に示したように、EBMでは研究によるエビデンスに加え、「患者の病状・環境」や「患者の価値観」も考慮に入れた上で、医師が専門的な判断を行う。

これをEBPMに当てはめるならば、「患者の病状・環境」はその施策の置かれた状況、すなわち施策が必要となる社会問題の深刻さや、対象集団の特性、施策の実現可能性やコストということになる。これはすなわち、先に見た「現状把握のためのエビデンス」に当たる。もう一つの「患者の価値観」は政治的な選好に該当し、民主的な手続を経て、それを施策の意思決定に反映させる必要がある。これらの要素は従来の政策立案でも当然考慮されてきたものであるが、EBPMにおいても依然として重要である。

政策立案は医療とは異なり政治的な意思決定を経る必要があるため、EBMのようなわけにはいかないと考える向きもあるかもしれないが、実は構図そのものはEBMにおいても大きくは違わない。大切なのは、既存の意思決定にエビデンスという要素をプラスすることにある。いかに民意が求める施策であったとしても、効果がない、あるいは逆に有害な副作用があるような施策を実施すべきではないだろう。その点は、従来の政策立案では、実は見極められていなかったということになる。

（5）エビデンスの普及

五点目は、たとえ医療であっても、確固たる根拠のある治療法が必ずしも容易に普及はしないということである。たとえエビデンスが提供されても、医師や医療現場の行動を変え、実際に医療のあり方を変えるのは難しい。エビデンスが「ある」ということと、それが現場で「使われる」ということは、決して同義ではない。

このことは実は、政策立案においても過去に経験されている。米国では、1960年代からRCT等による施策の厳密な有効性評価が実施されたが、その報告書は政策現場で利用されず、評価専門家を失望させた。しかし、それを機に、評価の利用や利用の促進策について研究が進んだ（田辺 2019b）。

EBPMにおいても、今後、せっかく生み出されたエビデンスが思うように政策立案に使われないことも十分に起こり得る。その際は、そこであきらめるのではなく再び医療の経験を思い出し、エビデンスを普及する努力をするべきであろう。

（6）業績測定との併用

六点目は、厳密な有効性評価と簡便な指標による評価を適切に併用することである。医療において、エビデンスを普及するために採られた方策は、エビデンスを基にQIという指標を作成し、実際の医療行為とエビデンスとのギャップを測定して改善につなげる取組であった。近年、各国で医療の質向上への関心が高まっているが、そのための車の両輪となっているのが、EBMと、QIによる医療の質評価なのである。

これをEBPMに当てはめるとすれば、エビデンスを基に妥当性の高い指標を設定し、それを業績測定に活用することにほかならない。米国の連邦政府に業績測定を義務付けた政府業績結果法（GPRA）[14]が意図したように機能しなかった一因は、業績測定では施策と効果の因果関係を明らかにできず手法的に限界があったことにある

と指摘されている（津田・岡崎 2018、p.4）。言い換えれば、因果関係に基づく妥当な指標設定ができていなかったということであり、もしエビデンスの裏付けがあれば、より適切な指標を設定することができる。本来、有効性評価に代表されるプログラム評価と指標による業績測定は、それぞれ役割が異なり、両者の性格を理解した上で使い分けることが望ましい（田辺 2014）。

日本では政策評価が1990年代後半から急速に普及したが、厳密なプログラム評価はほとんど普及せず、業績測定による評価が中心となっている。そしてその業績測定においては、指標の設定に困難を抱える例が多い。EBPMは、妥当な指標設定の根拠を提供しうるものであり、それによって業績測定を改善できる可能性がある[15]。

(7) 政府の役割

最後に、特に日本におけるEBMの経験に照らすならば、政府に期待される役割にEBPMのためのインフラ整備があることが挙げられる。日本においては、EBMの導入が諸外国に遅れ、1999年頃に当時の厚生省においてその推進が議論された。その結果、厚生省は、①診療ガイドラインの作成支援、②それを提供するEBMデータベースの整備、③エビデンス創出のための臨床研究の推進等を進めてきた。これらは、EBMの「つくる」「つたえる」局面を支援する取組であり、医療界全体のEBMを支えるインフラ整備に当たるものといえる。

翻って日本のEBPMの現状を見ると、政府による取組は各省庁が自らの政策立案にエビデンスを用いる事例づくりが中心となっており、わが国全体のEBPMのインフラ整備という視点は薄い。

「つくる」局面については、信頼性の高いエビデンス創出を目指すのであれば、行政官が自ら有効性評価を行うのは限界がある。英米のEBPMでは外部委託によるエビデンス産出が行われており（大槻 2018、p.32；津田2019、p.97）、日本でも大学等を含めた政府外でのエビデンス創出を政府が支援する方向性も考えられる。そのための予算としては、EBMでは診療ガイドライン作成に厚生労働科研費という研究予算が活用された例があるし、英米のEBPMでは事業予算の一定割合を評価に割く方法も採られている（大槻 2018、p.33）。

「つたえる」局面については、先に触れたように、わが国の議論では抜け落ちている部分であり、今後どのようにエビデンスの集積・提供を進めていけるのか、まずは議論を始める必要があるだろう。

7. おわりに

本稿では、医療におけるEBMの経験からEBPMが学べることをまとめた。医療をエビデンスに基づいて実践するというアイディアは、実は古くから当然のようにあったものではなく、最初にEBMが提唱された1991年から現在まで、いまだ30年も経っていない。特に日本においてはEBMの導入が遅れ、諸外国へのキャッチアップが目指されてきた。

なお、本稿では紙幅の都合もあり、医療現場におけるエビデンスに基づく「実践」としてのEBMに議論を限定したが、保健医療の政策立案にエビデンスを適用する「エビデンスに基づく医療政策」については、EBMに比べ取組が遅れているとされる（津川 2014）。

わずか30年足らずの間に、EBMというまったく新しいパラダイムを作り出し普及させた医療の経験からは、EBPMが最大限目指しうる理想像とともに、今後直面するであろう課題をも占うことができる。本稿が、EBPMを推進する上で少しでもヒントとなれば幸いである。

付記

本稿における見解は筆者個人のものであり、所属する組織を代表するものではない。

注記

1 Guyatt, G. H.（1991）. Evidence-based medicine. *ACP Journal Club*, 114(2), p.A-16, 中山（2010）p.259に引

用。

2 最近の研究でも、従来広く実施されてきた医療行為にもRCTで有効性が否定されたケースがあることが示されている（岬 2019）。

3 利益相反を含め、診療ガイドラインの質を評価するためのガイドラインとしてAGREE IIがある（AGREE Next Step Consortium, “Appraisal of Guidelines for Research & Evaluation II,” May 2009）。

4 旧名称は「コクラン共同計画」（Cochrane Collaboration）。コクランの経緯については、正木・津谷（2006、pp.4-5）を参照。

5 これに加え、医療の質は医療安全や患者満足度を含むと解釈される場合もある（OECD 2011, p.16）。

6 医療へのアクセス、患者の経験という側面を加える場合もある（福井 2017、p.29）。

7 現在のサイト名は「Mindsガイドラインライブラリ」。https://minds.jcqhc.or.jp

8 この章の記述は、主に津田・岡崎（2018）、大槻（2018）、津田（2018）等による。

9 正式名称はBehavioral Insights Team (BIT)。

10 正式名称はSocial and Behavioral Sciences Team (SBST)。

11 Evidence-Based Policymaking Commission Act of 2016, P.L.114-140.

12 Foundations for Evidence-Based Policymaking Act of 2018, P.L.115-435.

13 International Initiative for Impact Evaluation (3ie).

14 Government Performance and Results Act of 1993, P.L. 103-62.

15 なお、津田（2019）では、本項の議論と異なる形での業績測定との併用として、施策の効果が見込まれるかを業績測定で見極めた上で有効性評価を行う米国の事例を紹介している。

参考文献

青柳恵太郎（2019）「国際開発分野におけるEBPM実践と自治体への示唆」、『かながわ政策研究ジャーナル』、No.13：86-94

東尚弘（2007）「エビデンス－診療ギャップを埋める研究」、『医学のあゆみ』、221（6）：542-546

家子直幸ほか（2016）『エビデンスで変わる政策形成：イギリスにおける「エビデンスに基づく政策」の動向、ランダム化比較試験による実証、及び日本への示唆』、三菱UFJリサーチ＆コンサルティング

岩田健太郎（2014）「エビデンスは常にある：EBM巡る勘違い、RCTそぐわない場合も」、医療維新ウェブサイト、2014.6.18 <https://www.m3.com/open/iryoIshin/article/225637>

上田佳世・五十嵐稔子・中山健夫（2018）「根拠に基づく医療（EBM）や診療ガイドラインとの上手な付き合い方」、『助産雑誌』、72（12）：916-922

内山融ほか（2018）「英国におけるエビデンスに基づく政策形成と日本への示唆：エビデンスの『需要』と『供給』に着目した分析」、RIETI Policy Discussion Paper Series、18-P-018

大槻大輔（2018）「英国、米国におけるEBPMの動向」、『統計』、69（7）：28-33

桐野髙明・松田晋哉（2015）「対談：医療の質指標は何をもたらすか」、『病院』、74（11）：781-786

経済協力開発機構編著、齋藤長行監訳（2018）『世界の行動インサイト：公共ナッジが導く政策実践』、明石書店

厚生科学審議会臨床研究部会（2019）「臨床研究・治験の推進に関する今後の方向性について（2019年版）中間とりまとめ」、2019.3.29

小林庸平（2018）「日本におけるエビデンスに基づく政策形成（EBPM）の現状と課題」、日本評価学会第19回全国大会、横浜市、2018.12.1

小松康宏（2012a）「透析医療のCQI入門　第1回：最善の医療を最善の方法で提供するシステムとは？」、『臨床透析』、28（1）：101-103

小松康宏（2012b）「透析医療のCQI入門　第2回：PDCAサイクルをチームで使う」、『臨床透析』、28（2）：213-216

小松康宏ほか（2016）「医療の質改善の概念と手法：PDCA, six sigmaなど」、『日本内科学会雑誌』、105（12）：2353-2357

杉谷和哉（2019）「『日本型EBPM』の展開と課題：2010年代統計改革の政策過程とエビデンス論」、公共政策学会第23回研究大会、茨木市、2019.6.7-8

鄭丞媛・近藤克則（2010）「医療の質のマネジメントとP4P」、『総合リハビリテーション』、38（10）：929-934

田辺智子（2006）「エビデンスに基づく教育：アメリカ教育省What Works Clearinghouseの動向」、『日本評

価研究』、6(1)：31-41
田辺智子（2014）「業績測定を補完するプログラム評価の役割：米国GPRAMAの事例をもとに」、『日本評価研究』、14(2)：1-16
田辺智子（2016）「エビデンスに基づく医療政策の必要性：医療の質と費用対効果」、『調査と情報―ISSUE BRIEF―』、No.907
田辺智子（2018）「エビデンスに基づく政策立案：その系譜と本質」、日本評価学会第19回全国大会、横浜市、2018.12.1
田辺智子（2019a）「業績測定の利用と影響：公共図書館を事例とした実証分析」、『日本評価研究』、19(2)：3-18
田辺智子（2019b）「医療の質と『実績に基づく支払（P4P）』」、『レファレンス』、(819)：77-98
津川友介（2014）「『医療政策研究』はじめの一歩：根拠に基づいた政策づくりを推し進めるために」、『週刊医学界新聞』、(3066)
津谷喜一郎（2000）「コクラン共同計画とシステマティック・レビュー：EBMにおける位置付け」、『公衆衛生研究』、49(4)：313-319
D. J. トーガーソン・C. J. トーガーソン、原田隆之ほか監訳（2010）『ランダム化比較試験（RCT）の設計：ヒューマンサービス、社会科学領域における活用のために』、日本評論社
内閣官房行政改革推進本部事務局（2018）「EBPMの推進」、2018.1.12、首相官邸ウェブサイト <https://www.kantei.go.jp/jp/singi/toukeikaikaku/kanjikai/dai5/siryou1.pdf>
中室牧子（2016）『学力の経済学』、ディスカヴァー・トゥエンティワン
中山健夫（2010）「エビデンス：つくる・伝える・使う」、『体力科学』、59(3)：259-268
中山健夫（2014）「診療ガイドライン：国内外の動向と展望」、『臨床放射線』、59(3)：373-378
南郷栄秀（2019）「わが国での診療ガイドラインの問題点と改善への試み」、『循環器ジャーナル』、67(1)：16-22
西内啓（2013）『統計学が最強の学問である』、ダイヤモンド社
J.ハッティ著、山森光陽監訳（2018）『教育の効果：メタ分析による学力に影響を与える要因の効果の可視化』、図書館文化社
廣瀬淳子（2019）「【アメリカ】2018 年根拠に基づく政策形成基盤法」、『外国の立法』、No.279-2：2-3
福井次矢（2011）「我が国の臨床研究の特徴と改善の方向性」、『最新医学』、66(7)：1676-1679
福井次矢（2015）「Quality Indicator（QI）を用いた医療の質管理」、『日本病院会雑誌』、62(4)：12-41
福井次矢監修（2017）『Quality Indicator〈医療の質〉を測り改善する―聖路加国際病院の先端的試み 2017―』、インターメディカ
福井次矢・山口直人監修（2014）『Minds診療ガイドライン作成の手引き 2014』、医学書院
伏見清秀（2015）「わが国における医療の質指標導入の意義について―国立病院機構の取り組みを中心に―」、『病院』、74(11)：808-811
星良孝（2013）「臨床研究の常識に異変あり（Vol.4）RCTからビッグデータへ」、医療維新ウェブサイト、2013.8.16 <https://www.m3.com/news/iryoishin/178690>
正木朋也・津谷喜一郎（2006）「エビデンスに基づく医療（EBM）の系譜と方向性：保健医療評価に果たすコクラン共同計画の役割と未来」、『日本評価研究』、6(1)：3-20
岬りり子（2019）「確立された医療行為でも約400件が無益」、Medical Tribuneウェブサイト、2019.7.22 <https://medical-tribune.co.jp/news/2019/0722520842>
よしかわアキ（2017）『日米がん格差：「医療の質」と「コスト」の経済学』、講談社
吉田雅博（2018）「Minds診療ガイドラインとは」、『呼吸器内科』、34(2)：182-187
OECD編著、児玉知子・岡本悦司訳（2011）『医療の質国際指標：OECD医療の質指標プロジェクト報告書2』、明石書店
OECD編、齋藤長行監訳（2018）『世界の行動インサイト：公共ナッジが導く政策実践―』、明石書店
McGlynn, E. A. et al. (2003). The quality of health care delivered to adults in the United States. *New England Journal of Medicine*, 348(26), 2635-2645.

（2020.3.21 受理）

Furthering Evidence-Based PolicyMaking in Japan: Implications from experiences in the healthcare sector

Satoko Tanabe

National Diet Library
stanabe0509@gmail.com

Abstract

Evidence-based policymaking (EBPM) has its origins in evidence-based medicine (EBM). Policymaking is believed to be separate from healthcare, where evidence is created and used in an ideal setting. However, even in healthcare, medical providers are not able to make decisions based only on evidence. Additionally, they face the issue of the evidence-practice gap.

This paper distilled suggestions for EBPM based on EBM and experiences in the healthcare sector, namely, creating a definition of evidence that focuses on effectiveness, accumulating and sharing evidence, understanding the factors to be considered in decision-making, combining the application of rigorous evidence and simple performance measurement, resolving difficulties in diffusing evidence, and realizing the government's role in developing infrastructure for EBPM.

Keywords

Evidence-Based Policy Making (EBPM), Evidence-Based Medicine (EBM),
evidence, randomized controlled trial (RCT), quality of healthcare

【研究論文】

日本におけるエビデンスに基づく政策形成（EBPM）の現状と課題
－Evidence-Basedが先行する分野から何を学び何を乗り越える必要があるのか－

小林　庸平
三菱UFJリサーチ&コンサルティング
y.kobayashi@murc.jp

要　約

日本においても、エビデンスに基づく政策形成（EBPM）が進められている。本稿では、Evidence-Basedに関する議論を概観したうえで、日本のEBPMの現状を整理するとともに、その評価を行う。日本のEBPMは事前分析やロジックモデルの作成に重点が置かれており、EBPMの本来的な意味からの逸脱がみられる。次に、Evidence-Basedが先行する医療等の分野との比較を行う。具体的には、問いの設定、エビデンスの創出、エビデンスの活用という3つの観点から、医療等の先行分野と政策の比較を行う。比較分析を踏まえて、EBPMをどのように捉えどのように進めていくことが望ましいかを検討する。具体的には、問いの設定の重要性や、エビデンスの範囲、つくる・つたえる・つかうのウエイト、エビデンスに対する需要創出の重要性等を指摘する。

キーワード

エビデンスに基づく政策形成（EBPM）、エビデンスの定義、エビデンスづくり

1．はじめに

近年、日本においても、エビデンスに基づく政策形成（Evidence-Based Policy Making：以下、EBPM）が進められている。2017年5月に統計改革推進会議が最終とりまとめを公表し、そのなかで各省におけるEBPM推進統括官（現：政策立案総括審議官）の設置と、官民データ活用推進戦略会議の下に政府横断でEBPMを推進するための会議体としてEBPM推進委員会を設置することが盛り込まれ、前者は2018年度から、後者は2017年度からスタートしている。その一方で、日本におけるEBPMには課題も多く、批判的な声も聞かれるようになってきた（青柳 2018a、田辺 2018）。

そこで本稿では、Evidence-Basedな取り組みが先行する医療等における議論の整理を行ったうえで（第2節）、日本におけるEBPMの現状を整理し評価を行う（第3節）。

次に、医療等のEvidence-Basedが先行する分野と政策の比較分析を行うことで、既存のEvidence-Basedの議論を適用するだけでは不十分な点を整理する（第4節）。そのうえで、EBPMを進める際に先行分野から何を学び、そして何を乗り越えていく必要があるのか考察を行う（第5節）。最後に本稿の結論をまとめる（第6節）。

なおEBPMについては、地方自治体でも同時に進められている[1]が、本稿は主として国の動向に

焦点を当てる。

2. Evidence-Basedに関する議論の整理

そもそもエビデンスやEvidence-Basedな考え方とは何を指すのか。後述するように、国・地域や論者によってエビデンスにはさまざまな定義が与えられているが、その源流はエビデンスに基づく医療（Evidence-Based Medicine：以下、EBM）にあるとする考え方が一般的だろう（正木・津谷 2006、岩崎 2010、Baron 2018、青柳 2018b）。

EBMという言葉が初めて学術誌上に登場したのはGuyatt（1991）だとされるが、その後Sackett et al.（1996）が「EBMとは、個々の患者のケアに関する意思決定過程で、現在得られる最良のエビデンスを良心的、明示的、かつ思慮深く利用することである。EBMの実践とは、個人の臨床的専門技能を、系統的研究から得られる外部の利用可能な最善の臨床的エビデンスと統合することを意味している」（筆者訳）と定義することでより具体的に提示した。ここでいうエビデンスとは、ある医療的な介入行為に関する因果的な検証結果のことであり、例えば、ランダム化比較試験（Randomized Controlled Trial：以下、RCT）によって実証的に明らかにされた介入の因果効果がエビデンスの代表的なものとなる（正木・津谷 2006）。この定義からも明らかなように、EBMとはエビデンスに基づいて治療方法を機械的に選択することではなく、医師の技術や経験、患者の価値観や嗜好も含めた総合的な判断を行うことを意味している。その意味でEvidence-Informed Decision Makingという表現がより正確にEBMの本質を表現していると言える。

EBMが世界的に普及した背景には、1992年にイギリスの国民保健サービスの一環として始まったコクラン共同計画（The Cochrane Collaboration）がある（正木・津谷 2006）。コクラン共同計画はRCTを中心として世界中のエビデンスの評価や統合を行い、EBMのインフラとしての機能を担った。こうしたEvidence-Basedの考え方が医療以外の分野に広がったのが今日のEBPMにつながる流れだといえる。

EBMをその源流に持つEBPMだが、政策の因果効果に着目するという考え方はEBMだけに見て取れるわけではない。例えば政策評価の代表的な手法であるプログラム評価は、公共政策プログラムの有効性を体系的に明らかにする手法だが、具体的には、政策が必要とされる背景を評価するニーズ評価、政策の論理的妥当性を検討するセオリー評価、政策実施段階における問題の早期発見・解決を図るプロセス評価、政策と成果の因果関係を評価するインパクト評価、政策の費用対効果を検証する効率性評価で構成されている（Rossi et al. 2018、秋吉 2017、龍・佐々木 2000）。プログラム評価におけるインパクト評価はEBMにおけるエビデンスそのものであり、政策評価でも同様の考え方があったことを示している。

EBPMに連なるもう1つの大きな流れとして、経済学における信頼性革命（Credibility Revolution）や開発経済学におけるフィールド実験の興隆を指摘できる。信頼性革命はAngrist and Pischke（2010）が付けた名称だが、経済学の実証分析が因果推論の枠組みで解釈されるようになった結果、分析結果に対する信頼性が向上したことを意味しており、因果関係を特定するための実証分析手法が発展した。因果関係を特定することの重要性が高まるなか、開発経済学の分野ではRCTを実際の社会のなかで行うフィールド実験が興隆し、現実社会でのさまざまな制約を考慮しながらRCTを行うための手法が開発され、EBPMの潮流に影響を与えてきた（高野 2020）。

このようにEBPMは、EBMや政策評価、経済学の実証研究手法の発展のなかで政策に取り入れられるようになってきた考え方だと整理できる。

3. 日本におけるEBPMの現状

（1）統計精度の向上がきっかけとなって加速した日本のEBPM[2]

本節では日本のEBPMがどのように進んできたのかを概観する。前述の通り政策の因果効果に

着目する考え方は既存の政策評価にも内在されており、成果志向についてもNPM（New Public Management）のなかで日本の行政の中に取り入れられてきた[3]。政府内での議論はすべてが公開されているわけではないため、何をもって日本のEBPMの起点とするかは簡単ではない[4]。そこで本節では、一定の推測も交えながら、近年（2015年以降）における日本のEBPMの展開を整理していきたい。

近年、日本でEBPMが進展するきっかけとなったと考えられるのが、2015年10月16日に開催された経済財政諮問会議である。同会議において麻生太郎財務大臣と高市早苗総務大臣が、消費統計の正確性について論戦を繰り広げた。麻生大臣は、国内総生産（GDP）の推計に使われている総務省「家計調査」の動きが経済産業省「商業動態統計」の小売業販売額と異なっている点を指摘し、その正確性について疑義を呈した。それに対し、高市総務大臣は商業動態統計にはサービス消費がほとんど含まれていないと反論した。統計改革が重要な政策テーマとして浮上したのである。

こうした統計改革に関する議論とセットになってEBPMの検討が進んだのは、2016年8月に山本幸三氏が行政改革担当の内閣府特命担当大臣に就任し、三輪芳朗大阪学院大学教授が大臣補佐官に就いて以降である。2016年夏以降のEBPMに関する主な会議体・提言を整理したものが表1である[5]。当初は、統計の精度向上や統計データに対するアクセシビリティの向上が議論の中心だったが、徐々に政策効果の測定や各省庁での推進体制強化の方策等、EBPMの議論に発展してきたことが見て取れる。統計改革推進会議によるEBPM推進体制案を示したものが図1である。官民データ活用推進戦略会議の下にEBPM推進委員会を設置し、各省と連携しながらEBPMを進めていくことがうたわれている。

2017年8月からは、内閣官房副長官補を会長、各省のEBPM統括責任者を構成員としたEBPM推進委員会が立ち上げられた。EBPM推進委員会は、2018年春に、「統計等データ提供等に関する基本的なガイドライン」および「人材確保・育成等に関する方針」を決定している。また、2018年度は各府省においてEBPMの観点から政策の検証（適確な課題把握・目標設定、政策効果の予測・測定等）を行い、2018年末をメドにEBPM推進委員会において、実例が各府省につき1つ以上報告された。2019年度も引き続き実例が創出される予定となっている。また2018年度からは、各府省においてEBPM推進のための局長級のポストとして政策立案総括審議官が設置されている。

（2）各府省における取り組み状況

前述の通り、各府省におけるEBPMの取り組みは実例創出が現在行われている段階であり、公開情報のみからではどのようなことが進んできているのか分からない部分も多いが、いくつか整理をしたい。

経済産業省[6]では、中小企業に対する研究開発補助金の効果を、回帰不連続デザイン（Regression Discontinuity Design：以下、RDD）を用いて測定する等の試みを、2017年度に行っている。また2018年度の予算要求プロセスでは、新規事業の立案時にアウトカムを明確化させるとともにロジックモデルの作成を行った。そのなかから、個別施策の効果検証方法を検討し、省内の調査統計グループの統計コンシェルジュ・チームや経済産業研究所のサポートを得ながら、エビデンスづくりに着手している。総務省行政評価局は、EBPMの推進のために各府省と実証的共同研究を実施している。実証的共同研究の対象は、IoTサービス創出支援事業、女性活躍推進、競争施策広報、訪日インバウンドであり、ロジックモデルの作成と、RCTやRDDを用いた効果検証に取り組んでいる（総務省行政評価局 2019）。

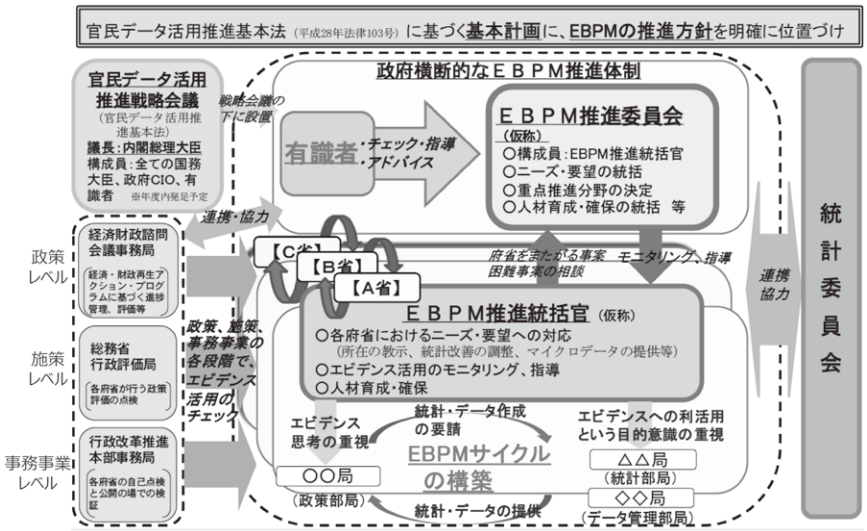
こうした経済産業省や総務省の取り組みは日本においてEBPMを進めるうえで興味深い試みとなっているが、他府省に目を転じると必ずしもそうした状況ではない。例えば、内閣府は2018年10月に「EBPMの観点からの政策の検証等[7]」をホームページで公表しており、30の事業についてロジックモデルを構築している。事業ごとの検証では、「手段と目標の因果関係の検討」という項目が記載されており、エビデンスを参照して政策を検討しようとする姿勢は見られるものの、エビデンスの活用はまだまだ十分とは言

表1　EBPMに関する主な会議体・提言

会議体・提言（実施時期）	開催趣旨・目的	主な結論・方向性
自民党行政改革推進本部 EBPM 提言（2016 年 6 月）	■限られた予算・資源のもとで政策効果を最大限発揮させるために、証拠に基づいて合理的に政策判断を行う EBPM を提言。	■ EBPM 推進体制の構築や統計委員会の機能強化。 ■アウトカム指標を統一化し、EBPM 優良事例を蓄積。 ■統計法の見直しやビジネスレジスターの推進。統計委員会の機能強化。 ■省庁間のデータの利活用を進める。 ■人材育成を進める。
EBPM のニーズに対応する経済統計の諸課題に関する研究会（2016 年 12 月）	■効果的な政策の企画立案の根拠を提供する重要な基盤である経済統計について、施策ユーザーのニーズから見た諸課題とその原因、改善方策を探る。 ■ EBPM の定着等について検討を行う。	■生産面を中心に置いた GDP 統計への移行を図るとともに、サービス分野等の 1 次統計を整備する。 ■ EBPM の実施には、①データが収集されていない、②収集されているのに技術的に使えない、③技術的に使えるがアクセスできない、④アクセスできるのに利用されない、という 4 類型が存在。 ■統計作成における行政情報の活用が不十分。 ■自治体等におけるリテラシーの向上が必要。
経済財政諮問会議（2016 年 12 月に 2 回議論）	■ 2015 年の麻生・高市論争と、上記研究会の検討をうけて、経済財政諮問会議のなかで、EBPM について検討。	■山本大臣から、「EBPM のニーズに対応する経済統計の諸課題に関する研究会」の内容を報告。 ■ 2016 年 12 月 21 日の諮問会議において、統計改革推進会議の設置を含む「統計改革の基本方針」が取りまとめられた。
統計改革推進会議（2017 年 5 月）	■「統計改革の基本方針」を踏まえて、官房長官を議長、行革大臣・財務大臣・総務大臣・外部有識者を構成員として設置。 ■ 2017 年初より議論を開始。	■ 2017 年 5 月に最終とりまとめを公表。 ■基礎統計の改善や生産統計の整備を進め、GDP 統計を供給・使用表体系に移行させ、推計精度を向上させる。 ■統計調査の利活用の促進や、報告者負担の軽減、統計業務の効率化を図る。 ■各省に EBPM 推進統括官（現：政策立案総括審議官）を設置するとともに、政府横断的な EBPM 推進委員会を官民データ活用推進戦略会議の下に設置する。 ■政策レベルは内閣府（諮問会議）、施策レベルは総務省、事務事業レベルは行革本部が所掌する。
EBPM 推進委員会（2017 年 8 月～）	■関係行政機関の相互緊密な連携の下、政府全体として EBPM を推進するために、官民データ活用推進戦略会議の下に設置。 ■内閣官房副長官補が会長、各省の政策立案総括審議官（EBPM 統括責任者）が構成員。	■行政改革推進本部事務局、総務省行政評価局、内閣府経済社会システム担当（EBPM3 本の矢）や、各府省と連携しながら EBPM の推進を検討する。 ■ 2018 年春に「統計等データ提供等に関する基本的なガイドライン」および「人材確保・育成等に関する方針」を決定。 ■ 2018 年度は、各府省において EBPM の観点から政策の検証（適確な課題把握・目標設定、政策効果の予測・測定等）を行い、1 つ以上の実例が報告された。 ■ 2019 年度も引き続き実例が創出される予定。

（出所）青柳・小林（2019）をもとに直近の状況を筆者が加筆

図1　統計改革推進会議によるEBPM推進体制案



（出所）統計改革推進会議最終とりまとめ参考資料（2017年5月）

えない。例えば「公文書管理に関する研修の充実強化」という施策における因果関係の検討は、以下のような記載となっている。

（前略）各行政機関に対して実施した（中略）アンケート調査の結果として、（中略）「行政機関の職員等、受講者のニーズを踏まえ、研修の内容を充実させていくことが重要」との指摘を受けたところ。

こうした各行政機関からの意見・要望を踏まえると、研修の充実強化を図ること（＝手段）は、各行政機関の職員の公文書に対するコンプライアンス意識の改革（＝目標）に繋がるという点で因果関係があるものと見込まれる。

つまり、「アンケートで研修のニーズがあるから研修が効果的である」と述べているだけであり、研修によってアウトカムが改善し得るのかどうかのエビデンスは示されておらず、エビデンスレベルは非常に低い。また家子・小林他（2016）が指摘している通り、EBPMにはそれに適した事業と適さない事業があるが、内閣府のEBPM検証のなかには「栄典事務の適切な遂行」や「専門調査会の実施」等、EBPMには必ずしも適さない事業も数多く取り上げられている。

こうした取り組みは、内閣府以外の他府省においてもおおむね同様であり、現在の日本のEBPMは事業に関するロジックモデルづくりが中心となっていると言える。

（3）Evidence-Basedに関する伝統的な議論との対比からみた日本のEBPMの評価

第2節で整理したEvidence-Basedに関する伝統的な議論を踏まえながら、日本におけるEBPMの評価を行いたい。

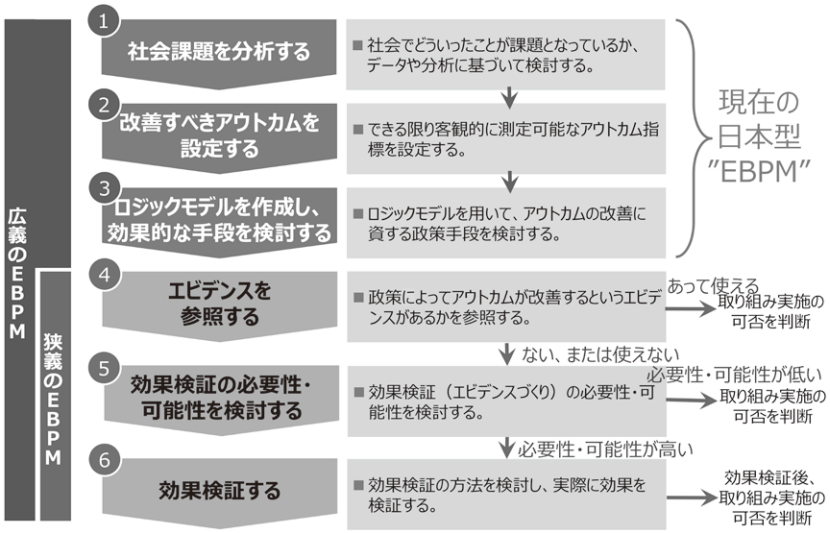
EBPMの流れと現在の「日本型EBPM」の関係性を整理したものが図2である。前述の通り、実証的に検証された政策の因果効果をエビデンスと定義すると、図2における④エビデンスを参照する、⑤効果検証の必要性・可能性を検討する、⑥効果検証するという3つのプロセスが本来的な意味（狭義）でのEvidence-Basedに相当する部分だと考えられる。一方で①社会課題を分析する、②改善すべきアウトカムを設定する、③ロジックモデルを作成し、効果的な手段を検討するという3つのプロセスは、EBPMを行うために必要ではあるものの、通常の政策形成でも必要なプロセスであり、EBPM特有のものではない。

日本のEBPMは統計改革が1つのきっかけとなったため、EBPMの目的もそちらに引き摺られている感は否めない。その結果、図2における、①社会課題を分析する、②改善すべきアウトカムを設定する、③ロジックモデルを作成し、効果的な手段を検討する、というプロセス（広義のEBPM）が中心になっており、エビデンス（政策の因果効果を表す実証的根拠）をつくったり（政策の効果を検証する）、つかったり（意思決定に活用していく）していく発想は弱い。加えて内閣府の事例からわかるように、ロジックモデルづくりも既存事業を当てはめるだけのものとなっており、ニーズ評価やセオリー評価の本来的な趣旨とも異なるものになっている。

こうした日本のEBPMの進展については、批判的な声も上がっている。例えば青柳（2018a）は、現状を把握するための情報（＝統計データ等）をエビデンスと呼んでいる場合があるが、それらはエビデンスではなく単なるデータであり、統計整備やロジックモデルの作成が行われてもEBPMが実現する訳ではないと批判している。田辺（2018）も、Evidence-Basedに関する議論の系譜を踏まえたうえで、日本のEBPMがエビデンスを広義に捉えていることを批判している。

確かにそうした指摘はもっともである。ロジックモデルの作成がEBPMだとみなされてしまえば、ロジックモデルをつくってEBPMが終了してしまう懸念があるし、政府全体で大量のロジックモデルが量産されるだけとなってしまうことが危惧される。また現状把握のための情報活用がEBPMだとみなされてしまうと、データや統計さえ参照していればEBPMであると解釈されるおそれもあり、今までの政策形成と何が違うのか分からなくなってしまう。第2節で整理したEvidence-Basedの本質的な部分（図2の④～⑥）を認識したうえ、EBPMの役割を再検討し、日本においてどのように進めていくべきかを整理していく必要があるだろう。

図2　EBPMの流れと現在の「日本型EBPM」



（出所）小林庸平（2018）「日本におけるエビデンスに基づく政策形成（EBPM）の現状と課題」（日本評価学会第19回全国大会報告スライド）

4. 医療等の先行分野との比較

前節では日本のEBPMの動向とそれに対する評価を行った。日本のEBPMに対しては批判的な声も聞かれるようになってきており、その多くはEvidence-Basedの考え方の基本から逸脱していることに起因すると考えられる。EBPMを進めるにあたっては、先行分野の経験から学び、意味ある形で定着を図っていくことが不可欠である。

しかしながら一方で、医療等の先行分野と政策形成では状況が異なっている部分も多い。そのことが、EBPMに対する認識のギャップを生んでいるものと考えられる。そこで本節では、Evidence-Basedの取り組みが先行する医療等と政策の比較を行う。

(1) 問いの設定

医療等の先行分野と政策の違いを、「問いの設定」、「エビデンスの創出」、「エビデンスの活用」という区分で整理をしたものが表2である。

「問いの設定」は「評価設問（PICO：Population、Intervention、Comparison、Outcome）の設定」と言い換えても良い。問いを設定するためには、何が課題になっているのかを特定し、アウトカムを設定する必要がある。

医療等であれば、既存の方法で治療が困難な病気等があればそれが課題そのものであるため、課題の特定は比較的容易だろう。またアウトカムについても、生存率やQALY（Quality Adjusted Life Year）等を用いるべきことが共通認識となっている。

しかしながら、政策形成の場合、課題の特定や改善すべきアウトカムの設定が容易ではないケースも多い。例えば2000年代初頭、格差の拡大が社会的・政治的に大きな論争になった。橘木（1998）はジニ係数が上がっていることから所得格差の拡大を指摘した。しかし大竹（2005）は高齢化と単身世帯の増加がジニ係数の上昇の原因であり、格差が拡大しているとは言えないと指摘した。このように政策の場合、社会課題を特定すること自体が難しいケースは多い。また、仮に所得格差の拡大が事実だったとしても、それが解決すべき課題なのかどうか判別が難しいケースもある。例えば、社会の流動性が高く、経済格差の世代間連鎖の程度が弱いのであれば、ある時点において格差が大きいことは問題ではないかもしれない。

同時にアウトカムの設定も容易ではない。所

表2　医療等の先行分野と政策の違い

要素		医療等の先行分野	政策
問いの設定	課題の特定	容易	必ずしも容易ではない
	アウトカムの設定	容易	必ずしも容易ではない
エビデンスの創出	アウトカムの測定可能性	高い	低い場合も多い
	SUTVAの成立	基本的に成り立つ	成り立たない場合も多い
	処置群と対照群の分離可能性	高い	低い場合も多い
	（潜在的な）サンプルサイズ	大きい	小さい場合も多い
	介入プロセスへの関与	比較的容易	難しい場合も多い
エビデンスの活用	外的妥当性	高い	低い場合も多い
	意思決定主体	受益者と提供者	社会

（出所）筆者作成

得格差をシンプルに解消するのであれば事後的に所得再分配を行えば良い。しかし事後的な所得再分配が就労や教育投資に対する意欲を弱めてしまうのであれば、むしろ教育機会の均等を図ることによって所得格差を間接的に解消する方が望ましいかもしれない。つまり政策の場合、政策的な介入によってどういったアウトカムを改善すべきかも必ずしも明確ではないのである[8]。

(2) エビデンスの創出

表2の「エビデンスの創出」は効果検証にかかる部分である。

定量的な効果検証を行うためには、アウトカムが測定可能である必要がある。医療の場合であれば生存率やQALY等がアウトカム指標として確立しており測定は容易である。政策でもアウトカム指標が容易に測定可能な場合はあるが、そうでない場合も多い。例えば人々の幸福度が重要なアウトカム指標になる場合もあるし、非認知能力（non-cognitive skills）等の測定方法が確立していないようなものがアウトカム指標となる場合もある。

SUTVA（Stable Unit Treatment Value Assumption）とは、ある主体への政策介入の有無が他の主体のアウトカムに影響を与えないという仮定であり、外部性（Externality）を排除するものである。医療の場合、伝染病等を除けばSUTVAの成り立つケースが多いと考えられる[9]。しかしながら政策の場合はSUTVAが成り立たないケースも多い。例えば近年、GoogleやAmazonといったプラットフォーム企業に対する政策対応の重要性が増しているが、市場において独占力の高い企業への対処は、当該企業のみならず関連企業や市場全体に影響を与えてしまうため、SUTVAは成りたない。また経済学的に言えば、外部性の存在が政策的介入を行う根拠になり得るため、SUTVAが成り立たないような社会課題こそ政策的な対応が求められるとも言える。

質の高いエビデンスを創出するためには、比較可能な処置群と対照群をつくりだし、介入効果を測定していくことが望ましい方法である。そのため、介入を受ける処置群と受けない対照群に分離できなければ、質の高い効果検証は難しい。医療等の先行分野は基本的に対人サービスが中心であるため、対象者を処置群と対照群に分離することは（少なくとも原理的には）容易である。一方で政策の場合は、処置群・対照群への分離が難しい場合も多い。例えば法人税減税の効果に関するエビデンスを得たいとしても、法人税が減税される企業と減税されない企業に分離して効果検証を行うことは難しい。こうした問題は、マクロ経済政策や通商政策等でも同様である。

統計的に誤差の少ない処置効果の推定値を得るためには、一定規模以上のサンプルサイズを確保することが必要となる。前述の通り医療等の先行分野の対象は「人」であるため、効果検証のために、母集団である対象人口全体までサンプルサイズを大きくすることが原理的には可能である。しかしながら政策の場合はサンプルサイズが小さく、誤差の少ない推定値を得ることが難しいケースも多い。例えば、大企業向けの研究開発支援の効果を測定しようとしても、そもそも分析対象にできる企業数が限られてし

まう。同様の理由でインフラ整備等も十分なサンプルサイズの確保は難しい。

質の高い効果検証を行うためには、介入プロセスに関与することによって理想的な検証環境を整えることが有効である。例えば処置群と対照群にランダムに割り当てることが可能であれば、RCTによって質の高い結果を得ることができる。しかしながら政策では、そうした前向き評価（Prospective Evaluation）が可能なケースばかりではない。政策は常に回り続けている状況であり、質の高いエビデンスが必要だからといって、実施している政策をいったん止めて検証環境を整備することはハードルが高い。また生活保護のように憲法や法律で要請されている政策の実施方法を、効果検証のために変更することも難しい。加えて、行政は基本的に年度単位で運営されているため、エビデンス創出のために年度途中で計画を変更することは容易ではない。

(3) エビデンスの活用

表2の「エビデンスの活用」は既存のエビデンスを政策的な意思決定に利用する部分である。

分析対象となったサンプルに対する因果効果の確からしさのことを内的妥当性（Internal Validity）と呼び、内的妥当性のある結果を他のサンプルや他の政策に一般化できるかどうかを外的妥当性（External Validity）と呼ぶ。医療の場合、構造がある程度安定している「生き物としての人間」を相手にしており、特定の環境でつくり出されたエビデンスを他の事例に横展開することは比較的容易だと考えられる。例えば海外で確認された薬効は日本でもある程度妥当すると考えられるし、過去に検証された医薬品の効果は現在でも基本的には同じような効果が得られると考えられるため、外的妥当性は高いと考えられる。しかしながら政策が対象としているのは「社会」や「社会のなかで意思決定を行っている人間」であり、社会的な文脈が異なれば、ある地域で実施した施策が他の地域でも機能するとは限らないし、ある時期にうまくいった施策が将来的には機能しなくなる可能性も高い。また、小規模では効果を上げていた施策が、大規模展開された場合に効果が小さくなることも指摘されている[10]。そのため、政策では外的妥当性は低い場合も多いと考えられる。

次に、エビデンスを活用して意思決定を行うのは誰なのかについて考えてみたい。医療等の場合、意思決定主体はサービスの受益者（患者）と提供者（医者等）となる。つまり受益者と提供者が納得した意思決定を行えれば良いのであり、比較的シンプルな構造だと言える。しかし政策の場合、意思決定主体は社会であり非常に複雑化する。例えば質の高いエビデンスがあったとしても、それが政治家や行政官にとって都合の悪いものであれば使われない可能性が高い。またある政策にエビデンスがあったとしても財源に限りがあれば別の政策をあきらめる必要が出てくるため、不利益を被る主体が反対するかもしれない。政治家や行政官がエビデンスを参照して意思決定を行ったとしても、市民や企業等がそれを支持しなければ、次の選挙ではそうした政治家は力を失ってしまうかもしれない。このように政策は民主主義社会のなかでの意思決定の結果として実施されるものであるため、非常に複雑な合意プロセスが求められる。

5. 先行分野との比較から得られる示唆

前節では、Evidence-Basedな取り組みを進めるうえで、医療等の先行分野と政策の比較を行った。そこからわかることは、医療等の先行分野の経験を踏まえることは重要であるが、政策特有の問題に対する配慮の必要性である。本節では、前節の比較を踏まえてEBPMに対する示唆を検討したい。

(1)「問いの設定」の重要性

医療や犯罪、教育、開発といった先行するEvidence-Basedの事例では、実践者の目の前に解決すべき課題が明らかになっているため「問いの設定」が容易であり、図2における「狭義のEBPM」にスムーズに入っていくことが可能である。そのためこうした分野では、図2の④～⑥がEvidence-Basedの重要な要素であると認識されている。

しかしながら前述したように、政策の場合は「問いの設定」自体が困難であるケースも多い。青柳（2018a）が批判するように、日本のEBPMが統計改革を通じた現状の正確な把握や、アウトカムの設定、ロジックモデルの作成といった点を優先させていることは事実である。しかしながら政策の場合は、事前分析による社会課題の把握や、改善すべきアウトカムの設定、望ましいと考えられる効果的な手段の特定（広義のEBPM）を行ってからでないと、狭義のEBPMが始まらないことも事実である。また日本の政策立案過程はEBPM導入以前の課題が多いからこそ、現状のようなEBPMが進展しているとも言える[11]。

また前述の通り、医療等の先行分野と比較すると、政策の場合はエビデンスの創出力が制約されるケースも多い。そのため広義のEBPMに取り組み、解決すべき課題の特定やロジックの精査を行ったうえで狭義のEBPMに取り組まないと、実践現場での負荷が大きくなり過ぎてしまい、結果として「エビデンス創出の無駄打ち」が多くなってしまう。

以上を踏まえると、EBPMにおいては「問いの設定」により大きな力を割いていくことが正当化され得る。

⑵ エビデンスの定義と範囲の再検討

①エビデンスの定義

次にエビデンスの定義と範囲について考えていきたい。

エビデンスとは何か。論者によってさまざまに定義されているが、前述の通り本来的には、RCTやそれに準じた方法に基づいて明らかになった「介入の因果効果」と定義されることが多い。政策を行う場合、それによってどの程度効果が得られるのかがもっとも不確実な要素であるため、「介入の因果効果」をエビデンスの定義の中心に据えていくことは重要である。その一方で、一般的にエビデンスと言う場合、インタビュー等に基づく定性的な因果関係や、シミュレーション分析の結果、データ・ファクト等を含めて、広義に捉えられるケースも多い。

諸外国におけるエビデンスの定義もまちまちである。例えばアメリカのCEP（Commission on Evidence Based Policymaking）では、エビデンスを「政府の施策・政策を評価する際に有用たり得る情報」と広く定義している。またイギリス政府に対する現地インタビューでもエビデンスは広義に捉えられているケースが多く、英国においてエビデンスに基づく警察活動を進めているWhat Works Centre for Crime Reductionは、取り組みの因果効果以外もエビデンスだと捉えている[12]。

②EBPMでエビデンスをどう捉えるべきか

それではEBPMにおいてエビデンスをどう定義してどう捉えていくべきか。Evidence-Basedが先行する医療等の分野では、エビデンスを狭義に捉えているケースがほとんどだと考えられる。しかし医療の場合、政策と比較して解決すべき課題（アウトカム）が比較的明確であり、SUTVAが基本的に成り立ち、かつサンプルサイズも大きくできるため、RCT等を適用することが容易なシチュエーションが多い。

一方EBPMの場合、そもそも解決すべき社会課題やボトルネックを特定することが難しい場合や、改善すべきアウトカムを設定することが難しいケースも少なくない。つまり前述の通り「問いの設定」が重要であり、そのための統計情報やファクト等も政策形成に有用な情報として位置付ける必要がある。

また前述した通り、EBPMにおいてはSUTVAが成り立たない、処置群と対照群への分離ができない、サンプルサイズが小さい、介入プロセスに関与できないといったケースも多い。そのため、エビデンスを狭義に捉えた（政策の定量的な因果効果）としても、RCTが適用できないようなシチュエーションについても幅広に考えていくことが不可欠である。

③エビデンスの範囲のあり方

以上の点を踏まえて、筆者なりにエビデンスの種類と分析手法の例を整理したのが表3である。大きくは、ファクト、事前分析・将来予測、政策の因果効果の3つに分けたが、政策の因果効果が狭義のエビデンスに相当する部分である[13]。

本稿における今までの分析を踏まえると、EBPMにおけるエビデンスの範囲は以下のように捉えるべきだと考える。

第1に、EBPMにおけるエビデンスの本質的部分は政策の因果効果とすべきである。そうでなければ、単にデータを使っているだけの政策形成もEBPMに位置付けられてしまうことになり、Evidence-Basedな取り組みが蓄積してきた知見を放棄することになってしまう。また、EBPMの重要性が普及してきたのは、政策の因果効果を踏まえたうえで政策形成を行うことの重要性が認知されてきたことと、それを明らかにするための分析ツールが発展してきたからに他ならない。

しかしながら第2に、「政策の因果効果」とは概念を明瞭に区別する必要はあるものの、ファクトや事前分析・将来予測といった情報も政策形成の有用な情報として活用していくことが望ましい。なぜならば、前述の通りEBPMにおいてはエビデンスの創出力に制約がある場合も多く、「問いの設定」が非常に重要だからである。ファクトや事前分析・将来予測は、問いを設定するために非常に有用な情報となる。

第3に、政策の因果効果をエビデンスの本質に位置付けたとしても、Evidence-Basedが先行する医療等の分野で想定されている領域は、政策形成のなかではごく一部にとどまる。具体的には、行動変容を伴い、SUTVAが成り立ち、処置群と対照群の分離が可能で、サンプルサイズが（潜在的に）大きく、介入プロセスへの関与が可能なケースのみである。こういったケースではRCTや前向き（Prospective）なRDDを適用することが容易であるため、質の高いエビデンスの創出は比較的簡単である。しかしながら表3を見ればわかるように、こうした理想的なケースは限定的である。

また、行動変容を伴わない（想定しない）ような政策効果の分析も非常に重要である。例えば、政府活動の大半は再分配政策であるといっても過言ではない。日本の一般政府ベースでの歳出規模は年間200兆円程度だが、そのうち社会保障給付が約110兆円であり、さらにそのうち現金による給付が70兆円程度を占めている。実に、政府歳出の1/3が現金給付である。こうした再分配政策は、各個人の行動を変容させなかったとしても、分配面に対してどのような影響を及ぼし得るのかを把握することは、政策立案上、非常に重要である。また日本のように、少子高齢化や高水準の公的債務残高に直面する国では、特定時点における再分配政策の影響だけでなく、時間を通じた世代間の再分配の状況を把握していくことも重要となる。ある時点の再分配政策の影響を把握するためにはマイクロシミュレーション分析が用いられることが多い。また、世代間の再分配政策の影響を把握するための世代会計モデルを用いて分析を行うことが多い。こうした分析から得られるエビデンスも、政策立案上非常に重要な情報である。

介入プロセスへの関与が難しい場合は、観察データを用いた分析に頼らざるを得ないが、そうした場合、近年は部分識別（Partial Identification）を用いた分析も行われるようになってきた[14]。また、処置群のサンプルサイズが小さい場合は統計的な分析は難しいと考えられてきたが、近年は合成コントロール法（Synthetic Control Method）を用いた分析も普及してきた[15]。処置群と対照群の分離が難しく、介入プロセスへの関与もできない場合や、SUTVAが成り立たない場合等は、例えば構造推定によるシミュレーション分析や一般均衡モデルの活用が考えられる。例えばBessho and Hayashi（2014）は労働供給の構造モデルを推定することによって、配偶者控除が廃止された場合のシミュレーション分析を行っている。またYamaguchi（2019）も構造推定したモデルを用いて育児休業期間が延長された場合の政策シミュレーションを行っている。伊神（2018）は、ハードディスクドライブを例に、イノベーターのジレンマの原因を分析するとともに、イノベーションを促進するための政策手段についても議論している。こうしたモデル分析の強みは、実際に実施したことのない介入であっても政策効果を見積もれることである。例えばBessho and Hayashi（2014）は、労働供給に関連するパラメーターを推定することによって、実際には実施されていない配偶者控除廃止の効果を推定している。

またマクロ経済政策については、VAR（ベク

表3　エビデンスの種類と分析手法の例

エビデンスの種類						分析手法の例
ファクト						・時系列データの集計 ・ミクロデータの集計
事前分析・将来予測						・将来人口推計 ・世代会計モデル・財政推計
政策の因果効果	行動変容を伴わない（想定しない）					・マイクロシミュレーション分析 ・世代会計モデル・財政推計
	行動変容を伴う	SUTVA	分離	ｻﾝﾌﾟﾙｻｲｽﾞ	介入への関与	
		○	○	○	○	・RCT、前向きRDD
		○	○	○	×	・自然実験、DID、部分識別
		○	○	×	○or×	・合成コントロール法
		○	×	○	×	・構造推定による税制シミュレーション分析 ・育児休業延長のシミュレーション分析
		×	○	○or×	○or×	・一般均衡モデル
		×	○	×	×	・構造推定によるシミュレーション分析
		×	×	×	×	・DSGEによるマクロ分析 ・時系列モデル
						・インタビュー、アンケート調査

（出所）筆者作成

トル自己回帰）モデルやDSGE（動学的確率的一般均衡）モデルの利用も不可欠になってくる。

「意思決定時の不確実性を減らし、より精度の高い施策を行うこと」がEBPMの本義なのであれば、エビデンスの類型には留意をしつつも、それをある程度幅広く捉えることを排除すべきではないと考えられる。

(3) つくる・つたえる・つかうのウエイト

EBPMを考えるうえで次に検討すべき点が、エビデンスをつくる・つたえる・つかうのウエイトである。

エビデンスに基づく医療（EBM）では、エビデンスをつくるだけでなく、つたえることやつかうことの重要性が強調されてきた（正木・津谷 2006）。エビデンスはつくることが目的ではなく、それらを使うことによってアウトカムを改善していくことが目的であるため、つたえることやつかうことが強調されてきたのはもっともである。

一方EBPMの場合、つくることをより強調すべきではないかと考えられる。前述の通り、「社会」や「社会のなかで意思決定を行っている人間」を相手にするEBPMでは、エビデンスの外的妥当性が低く、既存のエビデンスに基づいて意思決定できるケースは、医療等に比べると限られてくる。そのため、EBPMでエビデンスをつくることにより重きを置いていく必要がある。

またエビデンスは自動生成されるものではないため、効率的につくっていくためには仕掛けづくりが欠かせない。一般に、観察データからでは因果関係を明らかにすることは難しく[16]、エビデンスをつくっていくための仕掛けづくりが必要となる。施策の効果検証を行うためには、施策実施前に効果検証のための設計を行うことが不可欠である。具体的には、事前・事後でのデータの収集方法や、効果検証のために用いる統計手法の選択等を予め行っておく必要がある[17]。効果検証のために、政府内外において体制を整備していくことも必要となる。

また、効果検証を行ううえでのハードルの1つがデータ取得のコストである。理屈上は効果検証が可能であったとしても、データ取得コストが高すぎると、実際の効果検証は難しい。そのためデータ取得コストを可能な限り抑えることが重要であり、EBPMのための統計改革・行政デ

ータ活用を進める必要がある。例えば、政府は個人や企業の個票データを多数保有しているが、施策対象者リストと個票データを紐付けることができれば、政府が保有するデータを効果検証のために利用することが可能となり、効果検証のコストを大幅に抑えることが出来る。具体的には、税務情報等の行政データの活用や、政府統計と行政データの連携や、マイナンバー・法人番号の活用等が重要となってくる。

(4) エビデンスへのニーズを集約し、得られたエビデンスを共有する仕組みづくり

EBPMではエビデンスの創出能力に制約が生じる可能性があるため、エビデンスを効率的に創出し、得られたエビデンスを公共財として共有する仕組みが必要だと考えられる。

例えば英国では、各省庁がどういったエビデンスを欲しているのかをArea of Research Interestという形で公表している（内山・小林他 2018）。また米国でも、各省庁がエビデンスを創出していきたいテーマをLearning Agendaとして公表している（津田・岡崎 2018）。また地方自治体等の政策の執行現場でも、エビデンスに対するニーズは潜在的に多く存在していると考えられる。EBPMではエビデンスの創出能力に制約が生じるからこそ、ニーズを集約し優先順位の高いエビデンスを重点的に創出していく必要がある。

またエビデンスは一種の公共財であり、創出されたエビデンスを集約していく機能も必要だと考えられる。その際は、英国のWhat Works Centreの役割等が参考になると考えられる（内山・小林他 2018）。

(5) エビデンスに対する需要の創出

第2節で述べたように、EBPMの本質がEvidence-Informed Decision Makingにあるとすると、エビデンスを意思決定に活用していかなければ意味はない。そのためには、研究者主導による供給ドリブンのEBPMだけでなく、政策立案者や意思決定者（政治家・国民等）といった需要側にも同時に配慮していく必要があると考えられる。

前述したように、民主主義社会の意思決定では、質の高いエビデンスがあるからといってそれが使われて自動的に政策が作られるほど単純ではない。それぞれのステークホルダーがそれぞれの目的関数に従って行動をしているため、需要側への目配りがより一層重要になる。政策立案プロセスにおいて、エビデンスの利用や成果志向を社会として求めていくことが不可欠であり[18]、具体的には予算編成プロセスでエビデンスの活用を推奨していくことや、エビデンスを活用した政策形成を行った行政官を評価していく仕組みが必要となるだろう。エビデンスをつくり・つかうための政策担当者のケイパビリティ向上も進めていく必要がある。

6. おわりに

本稿では日本におけるEBPMの現状分析を行ったうえで、医療等の先行分野と比較をしながらEBPMをどのように進めていくべきかを分析した。現状および比較分析結果からは以下のような結論が導き出せる。

第1に、日本のEBPMは統計改革が1つのきっかけとなったため、そちらに議論が集中してしまい、かつロジックモデルの作成に重点が置かれてしまった点は否めない。その意味で、EBPMの本来の意義・目的に立ち返るべきという主張には一定の合理性がある。加えてロジックモデルについても事業ありきのものとなっており、目的を達成するために効果的な手段を検討するというロジックモデル本来の目的からも逸脱が見られる。

しかしながら第2に、Evidence-Basedな取り組みが先行する医療等と比較すると、政策形成はさまざまな面で状況が異なっており、既存分野の知見をそのまま政策に移植することは難しい。

以上の分析を踏まえて、EBPMを意味ある形で定着させるためには、以下の点に留意する必要がある。

第1に、EBPMでは「問い（評価設問）の設定」をより重視すべきである。政策の場合、問いの設定自体が難しい場合も多く、何が真に解決す

べき社会課題なのかを明らかにすることも簡単ではない。的を射た問いが設定できなければ、政策の因果効果がいくら特定されたとしてもその価値は減じてしまう。

第2に、政策の因果効果をエビデンスの本質的な部分として位置付ける必要はあるが、「問いの設定」に寄与する事前分析・将来予測等も意思決定の重要な情報として用いる必要がある。また政策では、SUTVAが成り立たない、サンプルサイズが小さい、介入プロセスに関与できないといったケースも多く、RCT等に代表される前向き評価のみをエビデンス創出のツールとして位置付けることも適当ではない。

第3に、EBPMは「社会」や「社会のなかで意思決定を行っている人間」を相手にしており、エビデンスの外的妥当性が低く、既存のエビデンスに基づいて意思決定できるケースは医療等に比べると限られてくる。そのため、EBPMではエビデンスを「つくる」ことに重点を置く必要がある。またエビデンスをつくるための仕掛けづくりを多面的に進めていかなければならない。

第4に、エビデンスに対する需要を創出していく必要がある。民主主義社会の意思決定では、質の高いエビデンスがあるからそれを使って自動的に政策が作られるほど単純ではない。供給者視点だけでなく需要者視点でのEBPMを進めていく必要がある。

注記

1 例えば石田（2018）は広島県におけるEBPMの取り組みを整理しており、小林（2020）は神奈川県葉山町における取り組みを紹介している。

2 本小節は青柳・小林（2019）をもとにしている。

3 NPMの考え方や動向については、例えば南島（2009）参照。

4 例えば2013年の経済財政諮問会議有識者議員提出資料（伊藤ほか 2013）でも「結果（エビデンス）に基づく政策評価を基礎とするPDCAサイクルの確立に向けて」が盛り込まれており、この時点ですでにEBPMが意識されていたともいえる。ただしこの資料では、政策目標の明確化や数値目標の設定とともにエビデンスに基づく政策評価が論じられており、前述のEvidence-Basedの考え方に立脚しているというよりも、業績測定（Performance Measurement）の発想に近いものと考えられる。

5 この表は政府全体の議論の流れを大まかに把握するために示したものであり、政策論議はそうした会議体のみで行われるものではない点に留意されたい。

6 三浦（2017）参照。

7 http://www.cao.go.jp/others/kichou/ebpm/kensyou.html（2020年2月11日アクセス）

8 本小節では再分配政策や所得格差への対応策という政策レベルの例を示したが、政策の因果効果を明らかにするというEBPMの特性上、事業レベルでの議論がもっとも適していると考えられる。しかしながら問いの設定の段階では、政策レベルを含めた課題の検討が必要となる。

9 ただし、教育におけるピア効果はSUTVAが成り立たないケースである。

10 例えば幼児教育プログラムについてはDuncan and Magnuson（2013）。

11 ただし前述の通り、日本のEBPMにおけるロジックモデルは事業ありきのものになっており、効果的な手段を検討するという発想が弱く、本来的なセオリー評価の趣旨には適っていないと考えられる。

12 内山・小林他（2018）参照。

13 既存の政策評価のフレームワークで考えると、主としてニーズ評価やセオリー評価で活用されるのがファクトや事前分析・将来予測であり、インパクト評価で活用されるのが政策の因果効果となる。ただし政策評価に関する代表的なテキストであるRossi et al.（2018）と比較して、本稿では政策の因果効果をより幅広く捉えている。

14 部分識別については、奥村（2018）参照。

15 合成コントロール法についてはAbadie and Gardeazabal（2003）等参照。

16 観察データを用いた施策の効果検証には、傾向スコアマッチングや差の差分析等が用いられてきたが、最近は信頼性が低いことが指摘されるようになってきた（King and Nielsen 2019、戒能 2017）。ただし近年はLASSO等の機械学習の手法を用いて精度を高めるような試みも始まっており、観察データを用いた効果検証の信頼度も今後高まっていく可能性がある。

17 施策の効果検証のために必要な要素については、小

林（2018）参照。

18 ブレア政権時に社会調査チーム次席を務めたPhilip Davies氏は、英国において質の高い評価が行われるようになった原動力は政府のEBPMに対する強いコミットメントであったと指摘している（Davies 2004）。

参考文献

青柳恵太郎（2018a）「経済学および国際開発の視点から」、第2回エビデンスに基づく実践と政策セミナー報告スライド

青柳恵太郎（2018b）「EBPMが継承すべきEBMの思考法」、『行政＆情報システム』、2018年10月号：28-32

青柳恵太郎・小林庸平（2019）「EBPMの思考法 やってみようランダム化比較試験！ vol.1 EBPMの思考法の「きほんのき」」、『経済セミナー』、2019年4・5月号：66-76

秋吉貴雄（2017）『入門 公共政策学』、中公新書

家子直幸・小林庸平・松岡夏子・西尾真治（2016）「エビデンスで変わる政策形成～イギリスにおける「エビデンスに基づく政策」の動向、ランダム化比較試験による実証、及び日本への示唆～」、『MURC政策研究レポート』、三菱UFJリサーチ＆コンサルティング

伊神満（2018）『「イノベーターのジレンマ」の経済学的解明』、日経BP社

伊藤元重・小林喜光・佐々木則夫・高橋進（2013）「財政の質の改善に向けて～実効性あるPDCAサイクルの構築に向けて」、平成25年第6回経済財政諮問会議提出資料

石田直人（2018）「広島県EBPM推進ワーキンググループの取り組み」、日本評価学会第19回全国大会報告論文

岩崎久美子（2010）「教育におけるエビデンスに基づく政策－新たな展開と課題」、『日本評価研究』、10（1）：17-29

内山融・小林庸平・田口壮輔・小池孝英（2018）「英国におけるエビデンスに基づく政策形成と日本への示唆－エビデンスの「需要」と「供給」に着目した分析－」、RIETI Policy Discussion Paper Series 18-P-018

大竹文雄（2005）『日本の不平等－格差社会の幻想と未来－』、日本経済新聞社

奥村綱雄（2018）『部分識別入門－計量経済学の革新的アプローチ－』、日本評論社

戒能一成（2017）「政策評価のための横断面前後差分析(DID)の前提条件と処置効果の安定性条件(SUTVA)に問題を生じる場合の対策手法の考察」、RIETI Discussion Paper Series 17-J-075

高野久紀（2020）「貧困対策　データで実証　開発経済学に新風」、『週刊エコノミスト』、2020年1月14日号：69-71

小林庸平（2018）「学校外教育バウチャーの効果分析とEBPMへの示唆」、RIETIコラム

小林庸平（2020）「英国では徴税率アップ 日本でもごみ放置減に一役」、『週刊エコノミスト』、2020年1月14日号：74-75

総務省行政評価局（2019）「政策効果の把握・分析手法の実証的共同研究」

橘木俊詔（1998）『日本の経済格差－所得と資産から考える－』、岩波新書

田辺智子（2018）「エビデンスに基づく政策立案－その系譜と本質－」、日本評価学会第19回全国大会報告論文

津田広和・岡崎康平（2018）「米国におけるEvidence-based Policymaking（EBPM）の動向」、RIETI Policy Discussion Paper Series 18-P-016

南島和久（2009）「NPMの展開とその帰結－評価官僚制と統制の多元化－」、『日本評価研究』、9（3）：17-27

正木朋也・津谷喜一郎（2006）「エビデンスに基づく医療（EBM）の系譜と方向性：保健医療評価に果たすコクラン共同計画の役割と未来」、『日本評価研究』、6（1）：3-20

三浦聡（2017）「経済産業省におけるEBPMの取り組み」、RIETI EBPMシンポジウム プレゼンテーション資料

龍慶昭・佐々木亮（2000）『「政策評価」の理論と技法』、多賀出版

Abadie, A. and Gardeazabal, J.（2003）. The Economic Costs of Conflict: A Case Study of the Basque Country. *American Economic Review*. 93（1）, 113-132.

Angrist, J. D. and Pischke, J.（2010）. The Credibility Revolution in Empirical Economics: How Better Research Design in Taking the Con out of Econometrics. *Journal of Economic Perspectives*. 24（2）, 3-30.

Baron, J.（2018）. A Brief History of Evidence-Based Policy. *The Annals of the American Academy of Political and Social Science*. 678（1）, 40-50.

Bessho, S. and Hayashi, M.（2014）. Intensive Margins, Extensive Margins, and the Spousal Allowances in the Japanese System of Personal Income Taxes: A Discrete Choice Analysis. *Journal of the Japanese and International Economies*. 34, 162-178.

Davies, P.（2004）. Policy Evaluation in the United Kingdom. Paper presented at the KDI International Policy Evaluation Forum, Seoul, Korea.

Duncan, G. J. and Magnuson, K.（2013）. Investing in Preschool Programs. *Journal of Economic Perspectives*. 27（2）, 109-132.

Guyatt, G. H.（1991）. Evidence-based Medicine. *ACP Journal Club*. 114（2）, A16.

King, G. and Nielsen, R.（2019）. Why Propensity Scores Should not be Used for Matching. *Political Analysis*. 27（4）, 435-454.

Rossi, P., Lipsey, M., and Henry（2018）. *Evaluation: A Systematic Approach 8th Edition*.

Sackett,D.L., Rosenberg, W.M.C, Gray, J.A.M, Haynes,R.B, and Richardson, W. S.（1996）. Evidence-based Medicine: What it is and What it isn't. *British Medical Journal*. 312（7023）, 71-72.

Yamaguchi, S.（2019）. Effects of Parental Leave Policies on Female Career and Fertility Choices. *Quantitative Economics*. 10（3）, 1195-1232.

（2020.3.23 受理）

Current Status and Issues of Evidence-Based Policy Making in Japan

Yohei Kobayashi

Mitsubishi UFJ Research & Consulting
y.kobayashi@murc.jp

Abstract

Evidence-based policy making (EBPM) has been gradually introduced in Japan. This paper summarizes and evaluates its current status it focuses on ex-ante analysis and logic model building, and deviates from the its primary role. We compare EBPM with the other field such as evidence-based medicine in order to clarify the differences between them. Based on the comparison, we discuss how to promote EBPM. Specifically, we emphasize the importance of setting questions beforehand, reconsidering the definition of evidence, generating evidence, and creating demand for evidence.

Keywords

Evidence-Based Policy Making (EBPM), Scope of Evidence, Generating Evidence

【研究論文】

国の行政機関におけるEBPMの取組実例の研究
－現状と課題－

砂山　裕
総務省[1]
ysunayama@soumu.go.jp

要　約

日本国政府でEBPMの取組が進められる中、国の行政機関の実際の政策を対象としたEBPMの実例を創出し、発信・共有・蓄積していくことが一層重要となっている。そこで、EBPMのリーディングケースの提示を目指し、国の政策を対象として、総務省、関係府省及び学識経験者による政策効果の把握・分析手法の実証的共同研究が行われた結果、「EBPM的要素」として日本の行政運営上有用と考えられる項目がいくつか見出されつつある。また、総務省が行う行政評価局調査の中でも、エビデンスを捉えるため各種のデータ分析手法が取り入れられてきており、個別府省の取組と相まって、政府全体でEBPMの推進が図られている。今後は、EBPMの普及・定着に向け、安定した制度的基盤の整備が必要であり、政策評価制度がその有力な選択肢の一つと考えられる。

キーワード

ロジック・モデル、因果推論、実証的共同研究、行政評価局調査、政策評価制度

1. はじめに

行政が限られた資源を有効に活用し、直面する課題に適切に対応して、国民から信頼され続けるためには、その政策立案が客観的データ等のエビデンスに裏付けられたものである必要があるとされており[2]、こうしたエビデンスに基づく政策立案、いわゆるEBPM（Evidence-Based Policy Making）の取組が日本国政府全体で進められている。EBPMについては、その概念や方法論等について統一的な公式見解がなく、研究者の間でも各々の学問的バックグラウンドの違いにより重点の置き方に相当の差異があるように見える。また、EBPMの中核をなす「エビデンス」についても明確な定義が存在しない。

そうした状況の下、EBPMの政府横断的な推進組織であるEBPM推進委員会が2017年8月に発足した。また、2018年度には、各府省においてEBPMの取組の推進を総括する責任者として「政策立案総括審議官」等[3]が新設され、それを支えるスタッフの新設・増員等が行われた。そして、政府全体におけるEBPMの牽引役としては、内閣官房行政改革推進本部事務局及び総務省行政評価局が実質的な推進官庁となっている。こうして体制は整備されたが、EBPMの概念自体が明確ではなく、現時点では取組内容についても確固たる方針があるわけではないため、当面は政府全体で試行錯誤を続けながら日本の行政運営の

日本評価学会『日本評価研究』第20巻第2号、2020年、pp.49-64

風土や実務に適した取組の確立を目指していくことになると思われる。したがって、EBPMの普及・定着に向け、今後は一般理論や海外事例の研究もさることながら、日本の実際の政策を対象とした実例を創出し、発信・共有・蓄積していくことがより一層重要となる。

本稿では、こうした観点から、国の行政機関におけるEBPMの実例について総務省の取組を題材として概観・考察するとともに、EBPMを国の行政実務に定着させる一つの方策として、政策評価制度をプラットフォームとすることの可能性について検討を行うこととする。なお、本稿中意見にわたる部分は筆者の個人的見解である。

2. EBPMの実例創出に向けた取組

政府におけるEBPMの議論の詳細な経緯は他稿に譲るが、大きな流れとしては、まず政府統計の精度の確保とデータ利活用の重要性が脚光を浴び、行政改革の流れと相まってEBPMと統計改革が「車の両輪」と位置付けられた。そして、前述のとおり各府省において体制整備がなされ、EBPMの実例創出に向けた試行錯誤が行われているという状況である。

こうした各府省の取組に加え、統計改革推進会議が公表した最終取りまとめ[4]において、「政策評価における取組」として、「EBPMのリーディングケースの提示を目指し、総務省、関係府省及び学識経験者による政策効果の把握・分析手法の実証的共同研究を行う」こととされた。ここでEBPMと政策評価制度とがリンクされ、政策評価制度を所管する総務省評価局が関係府省等と共同でEBPMの実例の創出に取り組むこととなった。

総務省では、まずEBPMの考え方を整理するため、2018年に小倉総務大臣政務官（当時）の下、EBPMに造詣の深い有識者が招かれ、政府におけるEBPM推進に向けた課題について意見交換会が開催された。その結果、同年10月に「EBPM（エビデンスに基づく政策立案）に関する有識者との意見交換会報告（議論の整理と課題等）」が取りまとめられた[5]。

その上で、この意見交換会の成果も活用しつつ、政府の実際の政策を題材として、EBPMを念頭に置いた政策効果の把握・分析に関する研究が進められ、2019年4月、その成果物の第1弾が「政策効果の把握・分析手法の実証的共同研究報告書」として公表された[6]。同報告書総論のタイトルには「ver1.0」と表記され、総論の本文には、今後も同様の実証的共同研究を実施し報告書を改訂していく予定であることが明記されている。

3. 具体的な分析実例 －概観と考察－

政府全体におけるEBPMの普及・定着を図るためには、実例の創出・発信・共有・蓄積が重要であることは前述のとおりである。そこで、本節では、上記実証的共同研究で取り組まれた分析の結果及びそこから得られた示唆について考察するとともに、評価専担組織としての総務省が実施した行政評価局調査における分析事例のいくつかについても考察する。

3-1. 関係府省と総務省による実証的共同研究

3-1-1. 共同研究の全体像

政策所管府省と総務省による実証的共同研究の第1弾では、ⅰ）IoTサービス創出支援事業、ⅱ）女性活躍推進、ⅲ）競争政策広報、ⅳ）訪日インバウンドという四つの政策テーマが題材とされた。所管府省や分析アプローチは様々だが、すべてのテーマにおいて、まずロジック・モデルが作成された。ロジック・モデルは、ある政策目的の下で行政資源を投入（input）し、行政活動（activity）を行った結果として生み出された政策手段のアウトプット（output）が、どのようにアウトカム（outcome）である政策効果の発現または政策目的の達成に至るかという論理（ロジック）を図解したもので、データ分析の下準備として、分析の目的や枠組み、重点ポイントを明確化する上で有用なツールである。そして、次にエビデンスを得るための種々の分析が行われ、RCT（ランダム化比較試験: Randomized Controlled Trial）に代表される因果推論の手法の

ほか、各種の定量的または定性的分析手法が適用された。なお、EBPMにおける「エビデンス」については国際的にも明確な定義はないが、本稿では、「政策手段と政策効果の因果関係に関する客観的・合理的根拠」という程度に理解しておく。

以下、実証的共同研究のテーマごとに分析の具体例を見ながら考察を行うこととする。ただし、4テーマのうち「女性活躍推進」については行政評価局調査の節で扱う（3-2-2.参照）。

3-1-2. IoTサービス創出支援事業の分析

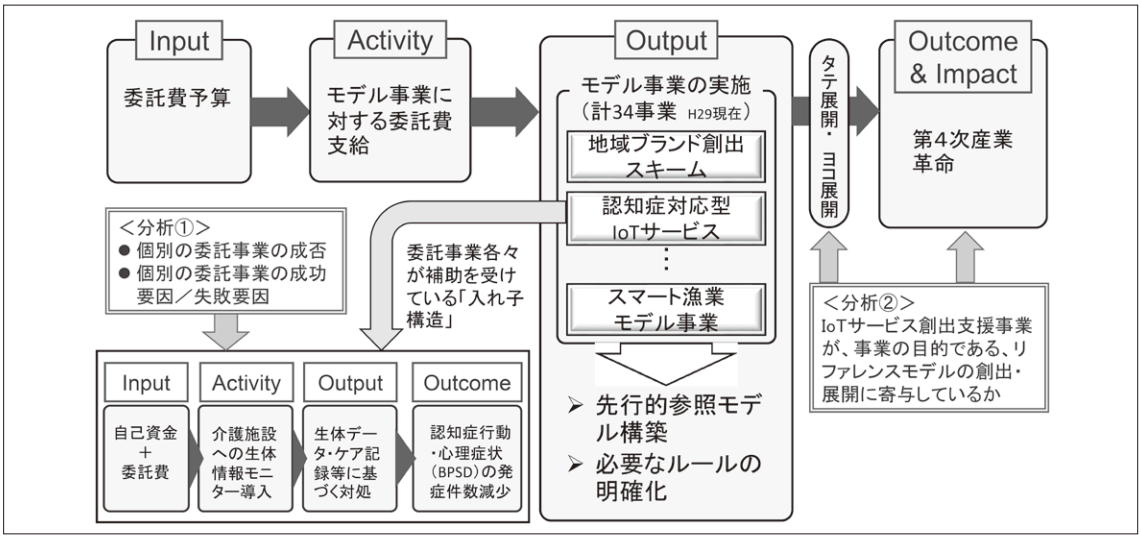
分析テーマの一つであるIoTサービス創出支援事業は、IoT（モノのインターネット: Internet of Things）の技術を用いて地域課題の解決に資するサービスを創出することを目指した一種のモデル事業である。自治体・民間企業・大学・NPO等からなるコンソーシアムに対して国が委託費を支出し、事業委託を行う形で実施される。対象分野は医療・福祉、農林水産業、防災等多岐にわたっており、2017年度までに34件が採択されているが、その事業内容は極めて多様である。本事業全体では、地域課題の解決を目指したIoTサービスをモデル的に多数創出し、事業継続及び他地域への展開を行うことで最終的に第4次産業革命につなげていこうというものであるが、その最終アウトカムに至るまでには無数の政策的要因及び外部要因が作用していると考えられる。こうした特徴をもった政策を前にした時、政策効果の把握・分析を行うに当たり、どのようにアプローチするかというのが本テーマの研究課題の中心である。

まず、本政策の政策効果の発現に至るロジックは、図1のとおり構造化・視覚化された。目を引くのは、本政策に含まれる数十件の各モデル事業それぞれがあたかも独立した政策であるかのように「入れ子構造」の政策ロジックを形成していると捉えたロジック・モデルを構築している点である。

本テーマでは、このロジック・モデルを基に、ミクロの視点として個別のモデル事業についてその成否や成功・失敗要因等を分析するとともに、マクロの視点としてそれらの事業について、同地域での事業継続（「タテ展開」）や他地域への拡大（「ヨコ展開」）に向け、事業のリファレンスモデルの創出・展開につながっているかを検証するという分析の枠組みを設定している。紙面の制約から、ここでは前者の個別モデル事業の分析について、二つの事例を見ていく。

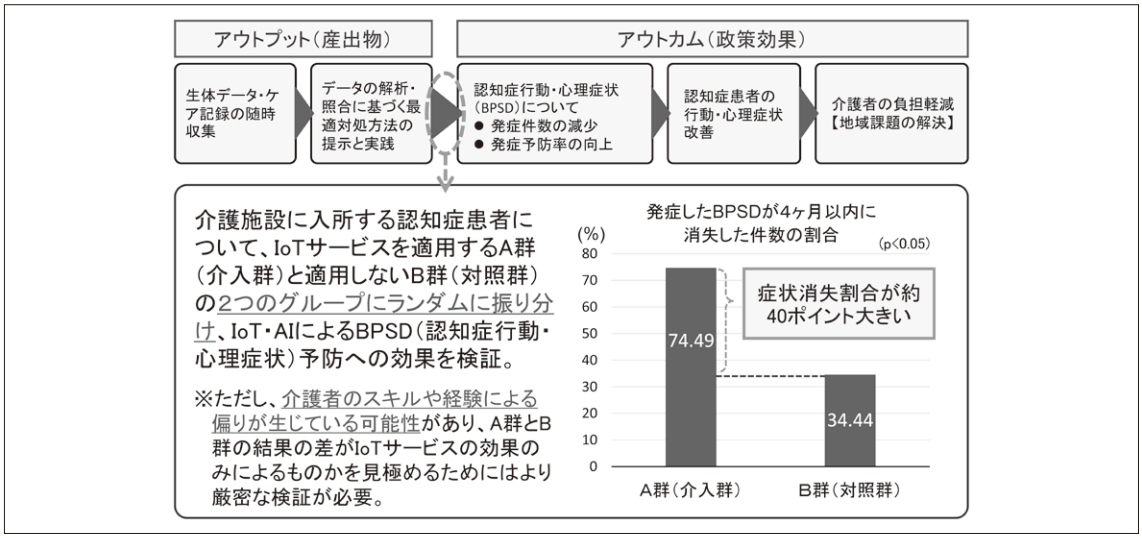
一つ目の事例は、「認知症対応型IoTサービス」

図1　IoTサービス創出支援事業のロジック・モデル



（出所）総務省「政策効果の把握・分析手法の実証的共同研究（IoTサービス創出支援事業の効果測定）報告書」[7]

図2　認知症対応型IoTサービスのロジックと効果測定



（出所）総務省「政策効果の把握・分析手法の実証的共同研究（IoTサービス創出支援事業の効果測定）報告書」

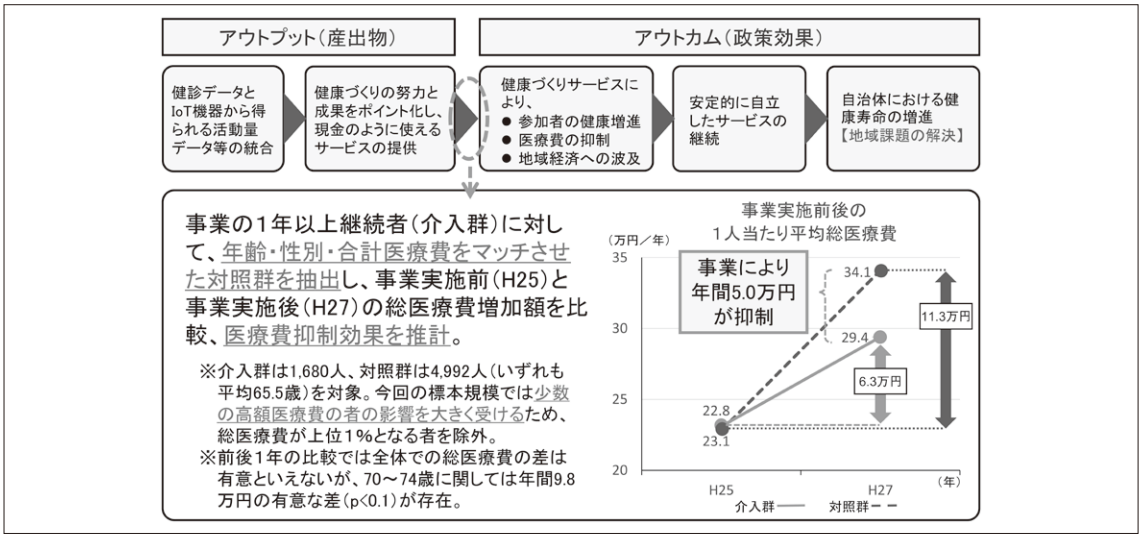
の事業の分析である。これは、認知症患者の介護について、IoT機器から得られる生体データや生活支援記録から得られる情報を基に、最適な支援方法を見出し、それに沿った介護を行うことで、患者の認知症行動・心理症状を改善し、ひいては地域の課題である介護者の負担軽減を目指すというモデル事業である。

この事業の分析に当たり、「介護者の負担軽減」をアウトカムとしての政策効果と捉え、そこに至るまでのメカニズムをロジック・モデルに整理した上で、分析の焦点を「IoTを活用した介護が認知症患者の症状改善につながるか」ということのエビデンスを得ることに当てている。すなわち、その部分こそが本事業の政策ロジックの肝であり、そのエビデンスさえ得られれば、ロジック全体がアウトカムまでつながることが合理的に推定できるという立論である。そこで、認知症患者の症状改善を測る指標として、一度発症した症状が4ヶ月以内に消えた割合を用いることとし、RCTの手法を適用し、IoTを活用した介護を受ける患者群（介入群）と受けない患者群（対照群）をランダムに振り分けて両者の間に差異が生じるかを検証した。その結果、対照群に比べ介入群の方が、統計的に有意な形で症状消失割合が大きいという結果が得られた（図2）。ただし、報告書にもあるとおり、ランダムに振り分けられたのは患者のみで、介護者については各患者とその担当者とのペアリングがそのまま維持された。科学的な厳密さを追求するのであれば、これら介護者の技能が症状の消失に影響した可能性があるため、介護者についてもランダムに振り分けるべきかもしれないが、認知症患者の介護は患者と介護者の相性の問題等もあり、介護の質を確保するためには両者のマッチングが極めて重要である。効果測定の実験のために介護の質を犠牲にするのは現場の実務としては本末転倒であり、上記のような結果が得られれば、現実的なエビデンスとして当面は十分だと考えるべきだろう。ただし、他の事例を対象とした類似の分析の積み重ねにより、エビデンスの精度を高めていくことが望ましい。

二つ目の事例は、「インセンティブ付きIoT健康サービス」の事業の分析である。この事業は、健康診断データやIoT機器から得られる歩数記録等を統合・見える化し、ポイント化して現金のように使えるサービスを提供することにより、住民の健康づくりの努力を促して自治体における健康寿命の増進につなげようという試みである。

この事業の分析でも、まずは政策ロジックを

図3　インセティブ付きIoT健康サービスの有料化挑戦事業の効果測定手法



（出所）総務省「政策効果の把握・分析手法の実証的共同研究（IoTサービス創出支援事業の効果測定）報告書」

明らかにすることからスタートし、健康づくりサービスに参加した住民の医療費抑制効果をエビデンスとして把握することが試みられた。分析手法としてはマッチング推定とDID（差の差分析: Difference-in-Differences）を適用し、図3のとおり、サービスの参加者について、加齢に伴う医療費の伸びが抑制される傾向が確認できた。全体の平均では5万円の医療費の抑制効果が算出されたものの統計的に有意ではなかったが、年齢階層を区切ると、70～74歳では年間約10万円の有意な差（10%水準）が見られた。

以上、IoTサービス創出支援事業の二つの分析事例を概観した。いずれもエビデンスを定量的な手法で得るものであったが、分析にかかるコストや時間も考慮し、最初に政策のロジックを明らかにした上で必要最低限のデータ分析に絞っているところがポイントである。行政実務におけるEBPMの実践を考える上で一つの重要な示唆を与えるものと考えられる。

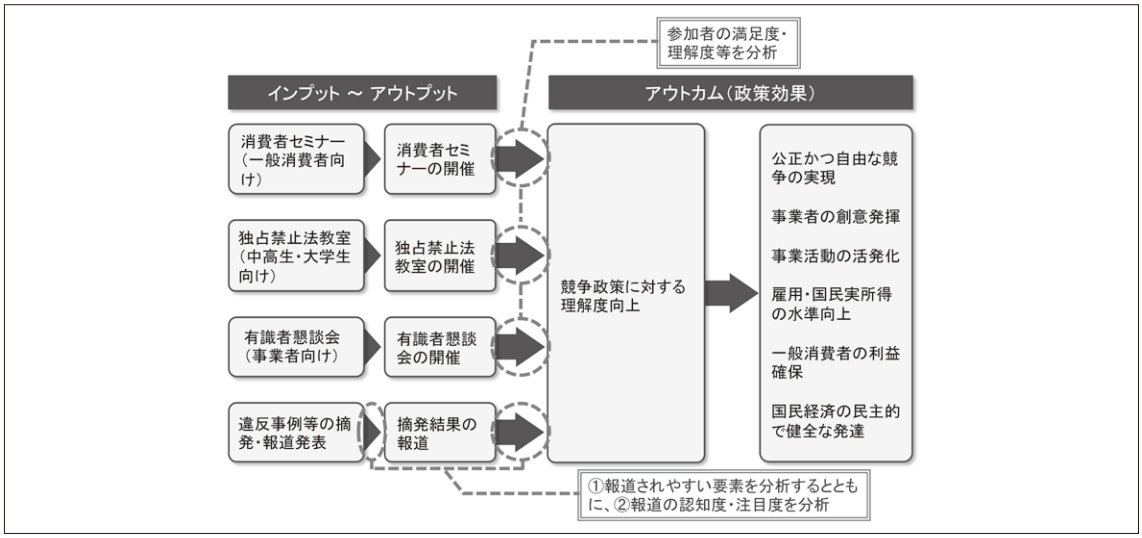
3-1-3. 競争政策における広報施策の分析

次のテーマは、公正取引委員会の独占禁止政策に関する広報施策である。公正取引委員会では、消費者セミナー、独占禁止法教室、有識者懇談会といった各種啓発広報事業を展開しており、本テーマでは、これらの広報効果や事業運営上の改善点を把握するための分析が試みられた。政策のロジックと分析のポイントは図4のとおりである。このうち本稿では、紙面の制約のため、消費者セミナーに関する分析に絞って概観する。

消費者セミナーは一般消費者向けのセミナーであり、独占禁止法や公正取引委員会の活動について分かりやすく説明する事業である。まず、セミナー自体の政策効果を把握するため、セミナー参加者における競争政策の理解度が確認された。参加者及びそれ以外の一般国民に対して競争政策に関するクイズを実施し、その正解率を比較したのである。その結果、前者の正解率は64.8%（個人属性等調整後は62.3%）、後者は56.2%となり、報告書は「消費者セミナーには参加者の理解度を高める効果があることが示唆される」としている。

次に、セミナーの実施方法の改善に資する示唆を得るため、満足度や理解度を目的変数とし、参加者属性やセミナー属性に関する多数の項目を説明変数として回帰分析を行った結果、表1のような結果が得られた（ポイントとなる主要な説明変数の係数のみ表示）。これによれば、参加者数と満足度の関係について、統計的に有意な

図4　競争政策広報のロジック・モデル



（出所）総務省「競争政策における広報の効果測定に係る調査・分析　結果報告書」[8]を基に筆者作成

形でマイナスの係数となっていることから、参加者数が多いほど満足度が低いことが示唆される。また、セミナーの開催に際し、コンテンツとしてシミュレーションゲームを組み込んだ場合、満足度との関係を見ると有意なプラスの係数となっており、シミュレーションゲームの実施は参加者の満足度を向上させることが示唆される。理解度についても、統計的に有意ではないものの、同様の傾向があることが示されている。なお、シミュレーションゲームについて、より厳密な分析のためクラスターRCT[9]を用いた効果の検証を行ったところ、有意ではないが満足度や理解度を向上させる効果が示唆される結果となった。

このように、本テーマは、セミナーの開催等

表1　回帰分析の結果（主要な説明変数のみ抜粋）

説明変数＼目的変数	満足度	理解度
参加者数	−0.00540 (**)	−0.00114
シミュレーションゲームあり	0.407 (***)	0.0272

（注）表中の（　）は統計的有意性の水準を示す。
（*** p<0.01, ** p<0.05）

（出所）総務省報告書を基に筆者作成

個々の広報メニューが全体として参加者の理解度向上というアウトカムに寄与しているかについてのエビデンスを把握するだけでなく、その広報事業の効果的なやり方を示唆するエビデンスをも提示するものとなっている。一般論として、RCTのような厳密な手法の実施に当たっては、コストや時間、実験可能な環境の確保等様々な条件をクリアする必要があるが、そういった手法でも、他の分析手法と組み合わせつつポイントを絞って適用することによって、効果的・効率的に有益な示唆を引き出し得るということの一例であろう。

3-1-4. 訪日インバウンド施策の分析とその限界

三つ目のテーマは、内閣官房及び観光庁が推進する訪日インバウンド施策の分析である。これは、地域における観光づくりの拠点となるDMO（Destination Management Organization）が行う各種広報その他様々な活動が、訪日観光客の呼び込み等に効果を上げているかを把握しようとするものである。その際、民間事業者が保有するビッグデータを活用することによって政策効果の把握が可能かどうかを検証することが試みられており、言わばEBPMにおける民間ビッグデータの活用の可能性を探ることを主眼として

いる。

他のテーマ同様、最初にロジック・モデルが作成された。極力単純化してしまえば、DMOが観光プロモーションを行った結果、その地域で訪日外国人客の数や消費額が増える、というのがマクロ的視点に立ったロジックである。そして、本テーマではさらに踏み込み、ミクロの視点に立ったロジックも提示している。すなわち、DMOによる広報活動が個々の外国人の日本への旅行意思決定に与える影響のプロセスをモデルに組み込み、マクロとミクロのロジックの流れがやがて「訪日観光客の増加」や「消費額／消費単価の増加」というアウトカムで合流するというロジック・モデルである（図5）。

このロジック・モデルを基に分析の枠組みが設計された。Wi-Fiアクセスポイントのログデータ、クレジットカードの購買データ、SNSへの投稿情報といった民間保有ビッグデータを活用し、DMOによる情報発信の前後において、訪日外国人の滞在人数・滞在日数や購買額、SNSへの投稿がどのように変化したかの把握が試みられたのである。

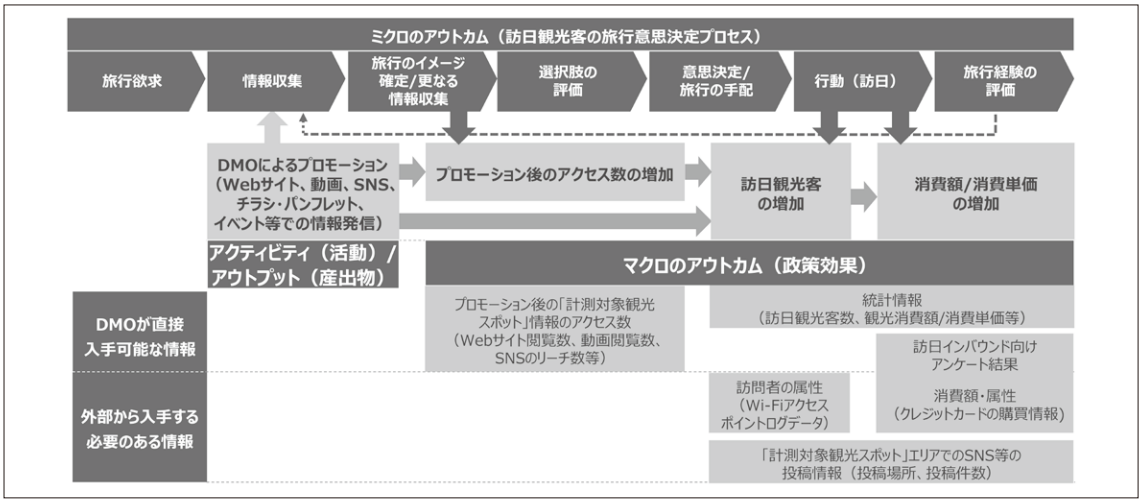
ところが、政策効果を把握するというニーズに対し、得られたデータの質と量には多くの課題があることが判明し、十分な分析ができない結果となった。分析を行う上での難点としては、例えば、生の個票データを入手して自ら自由に分析を行うことができず、データ提供元による分析結果が提供されるのが原則で、それも国別平均・月単位のように大くくり化された形でしか提供されなかった。また、全体はビッグデータでも、「DMOがプロモーションの対象とする地域」のようにデータのエリアを絞っていくと、該当データが数件にしかならないこともあった。しかし、考えてみればこれは至極当然のことかもしれない。なぜなら、そもそも民間保有データは行政目的のために集められたものではない。「ビッグデータの活用」という言葉が独り歩きし、ともすると魔法の杖のようにもてはやされているが、必ずしも有益な政策分析につながらないこともある。販売業者が自身の取扱商品の販売実績データを分析して販路の開拓に活用するというのとは異なるのである。

行政実務で民間ビッグデータを活用する際は、十分な吟味を行った上でデータを購入するよう注意が必要であろう。その意味でも、本テーマから得られた教訓は、実務の現場に警鐘を鳴らすものとして価値があるものと考える。

3-1-5. 実証的共同研究のまとめ

2018年度の実証的共同研究の結果、「EBPM的要素」として日本の行政運営において有用と考

図5　DMOによるプロモーションのロジック・モデル



（出所）総務省「「訪日インバウンド施策に関する調査・分析」報告書」[10]

えられるものがいくつか見出されつつあるように思える。

要素の一つは、政策のロジックの整理である。これは、ロジック・モデルの作成という形で具体化する。既述のとおり、実証的共同研究では四つのテーマの分析に当たり、まずそれぞれの政策のロジック・モデルを作成し、政策効果の発現メカニズムを構造化・視覚化している。ここで強調しておきたいのは、美しいモデル図を作成すること自体が重要なのではなく、その過程で政策の目的や本質的な要素等を深く議論した上で政策効果の発現に向けた戦略やその測定方法等を練ったり再確認したりすることに意義があると考える。唯一絶対の正解があるわけではないし、ロジックがスムーズに組み立てられるとも限らない。最終成果物ができるまで議論の紆余曲折が大いに想定されるが、その過程が政策立案を行う上で決定的に重要なのであり、整理されたロジック・モデルは議論の結果の共有手段にすぎない。行政実務としてのEBPMの実践的意義の一つがここにあると考える。

もう一つの要素がエビデンスの追求である。ここで注意が必要なのは、行政実務の現場でエビデンスという場合に、政策効果に関する厳密な因果関係の追求を至上命題としないこと、特に、RCTに代表される厳密な因果推論の手法といった特定のツールの適用のみに拘泥すべきではないということである。総務省の研究報告においても、分析手法から入るのではなく、最初に「施策の質の向上を図るためにはどのような情報が必要かつ有用か」という問題意識が重要であり、そうした情報を得るため、施策の特質や費用対効果等を踏まえた上で合目的的に方法論を選択すべきことが指摘されている。すなわち、多変量解析のような計量分析的な統計手法やシンプルな記述統計でもよいし、さらには定性的な分析であっても十分エビデンスたりえるということである。この点においてエビデンスに基づく（evidence-based）政策立案はデータ駆動型（data-driven）政策立案とは区別されるべきである。データを使った定量的な分析は重要であり追求すべきではあるが、それがすべてではない。

いずれにせよ、行政実務としてのEBPMは、より質の高い政策立案の実現を目指す営みであることを常に念頭に置くべきであろう。「入手できるデータ」をかき集めて定量的な分析を行ったり、説明材料として有利なデータを多数散りばめたりすることが大事なのではない。「よりよい政策の実現のために真に何が知りたいのか」をまず明らかにした上で、データ取得と分析にかかるコストや時間も勘案し、論点を絞って合目的的に取り組むことが重要である。

3-2. 総務省の行政評価局調査における実例

3-2-1. 概況

行政評価局調査は、総務省行政評価局が評価専担組織として実施する調査であり、政策評価と行政評価・監視という二つの大きな枠組みがある。2001年の中央省庁再編前の総務庁行政監察局が実施していた行政監察の流れをくみ、中央省庁等改革の一環として導入された枠組みである。方法論としては、各府省の様々な政策や行政運営について特定のテーマを定め、様々な角度から掘り下げて調査分析を行い、その結果を最終的に報告書にまとめるという、いわばプログラム評価のイメージである。特徴としては、ⅰ）各府省に対する強力な調査権限を有しているほか、地方自治体その他の公的団体や民間企業等あらゆる対象をターゲットにして調査を行えること、ⅱ）評価した結果に基づき、勧告・意見等という形で関係府省に対し政策・行政運営の課題や改善点を具体的に指摘し、実効性のある改善を実現していること、ⅲ）組織の強みとして全都道府県に出先機関を配置して全国調査網を形成しており、そのネットワークを駆使して全国津々浦々の行政現場の生の情報を収集し、机上の空論ではない現場実態に即した調査分析を行っていること、などが挙げられる。

本節では、最近の行政評価局調査の中からいくつかの事例をピックアップして考察する。いずれの事例も、データ分析のみで結論を導くものではなく、ケース・スタディのような定性的な手法も組み合わせながらエビデンスを形成し、政策や行政運営の改善に向けた種々の指摘を行っているが、本稿ではデータ分析部分のごく一

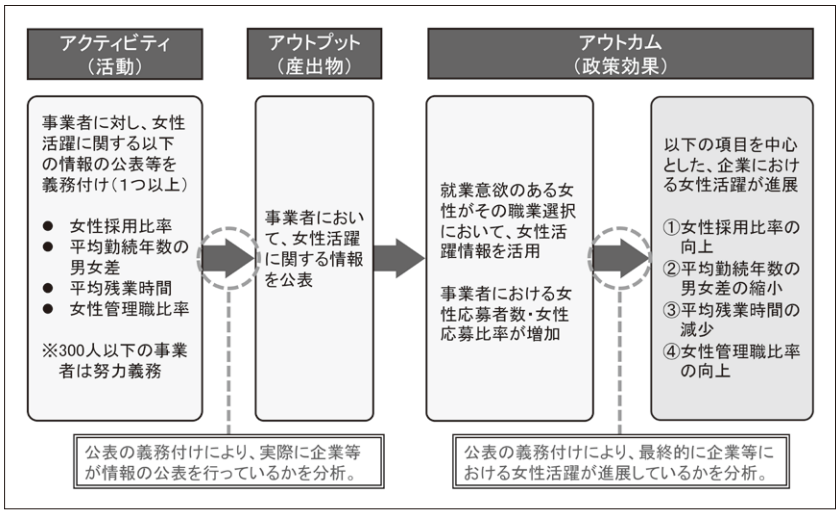
部をサンプル的に抽出して考察する。

3-2-2. 女性活躍推進施策の分析

最初の事例は、内閣府及び厚生労働省が取り組んでいる女性活躍推進施策をテーマとした調査（2019年7月「女性活躍の推進に関する政策評価」）である。この調査は、行政評価局調査であると同時に上述3-1.で言及した関係府省との実証的共同研究のプロジェクトでもある。この調査では、様々な角度から数多くのデータ分析が試みられているが、紙面の制約のため、政府の施策として女性の活躍状況に関する各種指標の公表を事業者に義務付けたことの政策効果に係る分析に絞って見ていくこととする。これは、一定規模以上の事業者に対して女性採用比率、女性管理職比率等の指標の公表を義務付けた結果、実際に事業者側がこれらの情報の公表を行っているか、また、アウトカムとして現実にこれらの指標が向上しているか等に関する分析であり、その方法論としてRD（回帰不連続デザイン: Regression Discontinuity design）やDID等因果推論の手法が多く適用されている。筆者の知る限り、行政評価局調査で大々的に因果推論の手法が採り入れられたのは本件が初めてである。

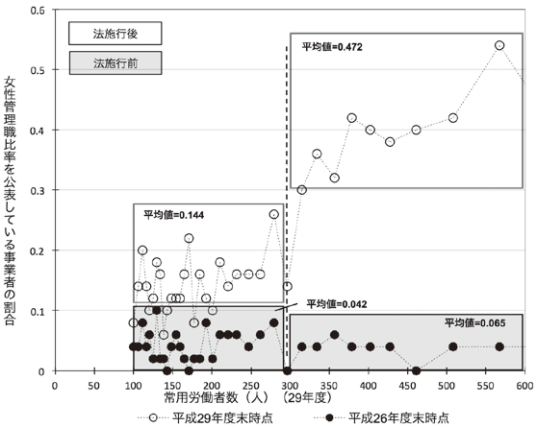
まず、指標の公表を義務付けた結果として、事業者がそれに従って公表を行っているかどうか。ロジック・モデル（図6）を見ると、義務付けという直接的な手段であることを考慮し、事業者による公表を「アウトプット」と位置付けている。他方、義務違反に対する罰則のない緩やかな義務付けであるため、公表状況は義務付けという政策手段の効果として「初期アウトカム」と捉えることもできるだろう。いずれにしても、女性採用比率、平均勤続年数の男女差、平均残業時間、女性管理職比率という四つの指標について、公表の義務付け前後で義務付け対象となっている事業者（常用労働者数300人超）とそれ以外の事業者（同300人以下）の間で差が生じているかを前後比較したところ、四指標すべてについて明確な差が見られた。一例として、RDで分析した女性管理職比率の公表に関するグラフ（図7参照）を見ると、義務を課す法の施行前後で義務付け対象事業者における公表割合が大きくジャンプしている（グラフの右半分）。また、興味深いことに、義務付け対象ではない事業者においても、法の施行前後でわずかながらジャンプが見られる（グラフの左半分）。以上のことから、違反に対する罰則がない状況の下でも、義務付け自体は指標の公表に効果があるものと考えられる。

図6　女性活躍推進施策（指標の公表義務付け）のロジック・モデル



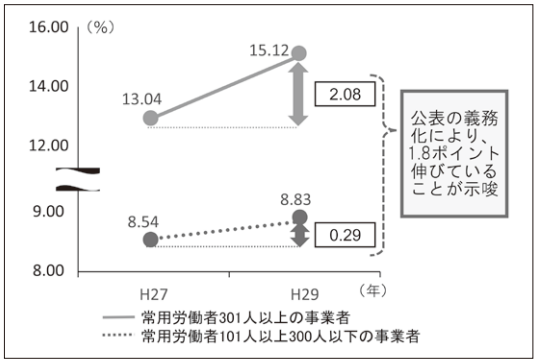
（出所）総務省「政策効果の把握・分析手法の実証的共同研究－報告書総論 ver1.0－」[11]及び「女性活躍の推進に関する政策評価書」[12]を基に筆者作成

図7　法施行前後の変化（女性管理職比率の公表）



（出所）総務省「女性活躍の推進に関する政策評価書」P.67

図8　法施行前後の変化（女性応募者数）



（出所）総務省「「女性活躍の推進に関する政策評価」の結果に基づく意見の通知（概要）」P.11

次に、これらの指標が公表されることにより、その情報が女性求職者に活用され、事業者への女性の応募者数が増加しているか。常用労働者数で基準化した女性応募者数について、公表義務を課す法の施行前後で、義務付け対象事業者とそれ以外の事業者の間で変化の差があるかをDIDで検証した結果、統計的に有意ではないものの、公表の義務付けにより女性応募者数が伸びる傾向があることが示唆された（図8）。この分析の結果からただちに指標の公表が女性応募者数の増加につながると即断することはできないが、一定の傾向は示されており、今後さらなる時系列データの蓄積・分析により、エビデンスとしての精度が高まるものと考えられる。

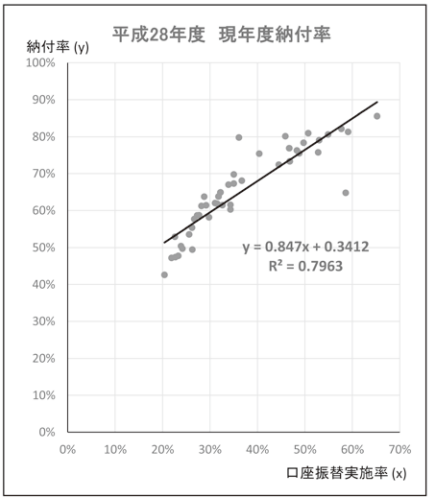
それでは、政策ロジック上さらに先にある女性採用比率や女性管理職比率等への影響についてはどうか。DIDで分析した結果、採用比率・管理職比率ともに統計的に有意な結果は出なかったものの、公表を義務付けられた事業者において増加傾向を示唆する結果が得られている。しかし、元来これらは政策効果の発現までに一定の期間が必要と考えられ、本分析の結果のみをもって政策効果の有無を即断することは困難である。中長期的な視点に立って継続的にモニタリングし、別の視点からの分析も積み重ねながらエビデンスを補強ないし修正していくことが必要であろう。本プロジェクトはそのきっかけを与える第一歩となっている。

3-2-3. 年金業務、ワーク・ライフ・バランス推進施策、高度外国人材受入れ施策の分析

次の事例は、国民年金の業務運営をテーマとした調査である（2018年12月「年金業務の運営に関する行政評価・監視」[13]）。これは、国民年金業務を中心として、日本年金機構の中期計画等に基づく業務運営を評価するとともに、保険料納付率の向上等を図る観点から、国民年金保険料の収納等の業務運営の状況を調査したものである。

この調査では、多岐にわたる調査項目の一つとして、国民年金保険料の収納対策についてデータを用いた分析を行っている。具体的には、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との仮説を検証するため、全国の年金事務所から45事務所を抽出し、口座振替実施率と現年度納付率の相関関係について回帰分析が行われた。モデルとしては、目的変数に現年度納付率、説明変数に口座振替実施率を設定した最も単純な単回帰モデルだが、強い相関関係が見られるとともに（$R^2 = 0.80$）、統計的に有意な回帰係数（0.85）が得られ、口座振替の効果が高いことが示唆された（図9）。そして単にデータ分析による理論的なアプローチのみならず、全国調査網を擁する組織の強みを生かして実際にその45年金事務所を訪れ一つ一つ丁寧に精査したところ、その示唆を裏書きするような状況が見られた。そこで、口座振替の推進

図9　保険料納付率と口座振替率の関係



▲ 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。

（出所）総務省　「政策効果の把握・分析手法の実証的共同研究－報告書総論 ver1.0－」P.59

は国民年金保険料の収納対策として有効であると結論付け、厚生労働省に対し、「日本年金機構の次期中期目標で口座振替による保険料納付を促進する旨を定め、口座振替の利用促進を図るための取組を強化すること」を旨とする勧告を行っている。この事例は、経験的・直観的には当たり前と思える事象についてデータ分析と個別調査により裏付けを行い、エビデンスを提示したという点において価値あるものと言えよう。

その他の事例として、ワーク・ライフ・バランス推進施策及び高度外国人材受入れ施策をテーマにした調査がある（それぞれ2013年6月「ワーク・ライフ・バランスの推進に関する政策評価」[14]、2019年6月「高度外国人材の受入れに関する政策評価」[15]）。前者は行政評価局調査において統計的なデータ分析に本格的に取り組んだ草分け的な調査であり、ワーク・ライフ・バランスの推進に関して設定された16の政府目標それぞれについてロジック・モデルを作成するとともに、データ分析及び事例研究を行い、各目標を達成するための政策手段の効果を分析している。一例として、自己啓発を行う労働者の割合を増

表2　労働者の自己啓発実施に影響を与える要素

目的変数	説明変数	回帰係数	統計的有意性
自己啓発を実施しているか否か（質的変数）	キャリア・コンサルティング経験の有無	0.23663	＊＊＊
	年齢	0.03892	
	男性	0.05144	＊
	年収	0.05330	＊
	正社員	0.08013	＊＊
	1ヶ月当たり所定外労働時間数	0.03568	
	大都市在住	−0.01770	
	勤務先に自己啓発休暇or自己啓発受講料補助制度がある	0.42709	＊＊＊

「統計的有意性」欄は、「＊」の数が多いほど統計的有意性が高いことを表す。
「＊＊＊」: 0.1%有意、「＊＊」: 1%有意、「＊」: 5%有意。

（出所）総務省「ワーク・ライフ・バランスの推進に関する政策評価書」を基に筆者作成

加させるとの政府目標に関し、その政策手段の一つであるキャリア・コンサルティングの効果を検証するため、ロジスティック回帰モデルにより分析を行った。その結果が表2であり、キャリア・コンサルティングが一定の効果を挙げていることが示唆されている。

また、高度外国人材受入れ施策の調査では、高度な専門的能力・知識を有する外国人材にできる限り多く長く日本で活躍してもらい、日本の経済・社会の活性化に貢献してもらうという政策の意図を踏まえ、どのような属性を備えた高度外国人材がより長く就労する傾向があるかを分析している。具体的には、法務省の入国管理に関する台帳データ等を利用し、就労予定期間が5年以上であるか否かのダミー変数を目的変数としてロジスティック回帰分析が行われた。その結果が表3である。これによれば、より長期の就労が期待できる高度外国人材は、中国国籍であること、社会科学または人文科学専攻であること、職歴が3年以上5年未満であることといった属性があることが示唆された。逆に、博士号を保有していること、就労先が大学等であることといった要素があると、より短い就労予定期間になることが示唆されている。この事例は、

表3　就労予定期間５年以上の人材の属性

目的変数	説明変数		回帰係数	有意性
就労予定期間が5年以上か否か（質的変数）	国籍	中国	1.2908	**
		アジア（除・中国）	0.5963	
		G7加盟国	0.1035	
	取得学位	博士号	−0.7350	*
		修士号	−0.0061	
		学士号	1.3797	
	専攻分野	自然科学	0.7232	
		社会科学	1.1010	*
		人文科学	2.1807	*
	就労先	大学等	−2.7256	*
		大企業	−0.1014	
		中小企業	−0.3456	
	職務に関連する過去の職歴	3年以上5年未満	1.0875	*
		5年以上10年未満	0.2124	
		10年以上	−0.2665	

「有意性」欄は、「*」の数が多いほど統計的有意性が高いことを表す。「***」: 0.1%有意、「**」: 1%有意、「*」: 5%有意。

（出所）総務省「高度外国人材の受入れに関する政策評価書」を基に筆者作成

法務省が周知・広報をはじめとする高度外国人材の受入れ推進政策に取り組むに当たり、必ずしも新たにデータ収集のための調査を行わずとも、自ら保有する台帳データという既存データを有効活用することにより、進むべき方向性について有益な示唆が得られることを示す好事例であると考える。

3-3. 実例創出の取組のまとめ

以上、国の政策に関するEBPMの実例創出の取組として、①各府省単独の取組、②関係府省と総務省の実証的共同研究、③評価専担組織としての総務省の取組のうち、②と③を重点的に考察してきた。②からは、政策マネジメントの主体となる政策所管官庁の立場では、政策の「ロジック」と政策効果の「エビデンス」が政策立案の質の向上のために重要な要素となること、③からは、政策所管官庁とは異なる立場からプログラム評価を行う場合でも、エビデンスを得るためのデータ分析が評価の質の向上に寄与し得ることが示唆されていると言える。今後も政府におけるEBPMの確立に向け、こうした実際の政策を対象とした取組の実例を創出・蓄積していくことが重要であると考える。

4. 今後の課題

現在試行的に行われているEBPMの実例の創出・蓄積を今後政府全体に普及・定着させていくためには、第一に取組の土台となる安定した制度的基盤を整備することが不可欠である。しかしながら、行政運営の基盤となる全く新たな制度を構築するためには膨大な時間とコストが必要になる。そこで、EBPMを単体で制度化するというよりも、既存の仕組みの中にEBPMの取組をうまく取り込んでいくことが現実的かつ効率的であると考える。以下では、その選択肢の一つとして、政策評価制度への取り込みの可能性について考察する。

4-1. EBPMと政策評価制度の関係

政策評価制度は、中央省庁等改革の一環として導入された制度であり、2001年1月の中央省庁の再編と同時にスタートした。当初は「政策評価に関する標準的ガイドライン」という各府省間の申合せを基に運用されていたが、同年、行政機関が行う政策の評価に関する法律、いわゆる政策評価法が成立し、同法が施行された2002年からは法律に基づく制度となった。したがって、法的基盤がしっかりした安定的な制度といえる。

EBPMはその名のとおり政策立案についての取組であり、一見すると政策評価とは政策マネジメント上のフェーズが異なる別物のように思える。しかし、政策評価法の規定を吟味すると、実は両者の間には密接な関係があることがわかる。すなわち、「政策評価」を定義する政策評価法第3条によれば、各行政機関にはその所掌する政策について自ら評価し、その結果を当該政策に適切に反映させることまでが義務付けられており、ここでいう政策への反映は政策立案を伴うことが想定されている。つまり、政策評価制度は政策のマネジメント・サイクルを確立することを一つの眼目としており、そこには政策立案も含まれていることになる。

また、同条によれば、政策評価を行う際には政策効果を把握することとされているが、この政策効果とは行政上の一連の行為（行政活動）

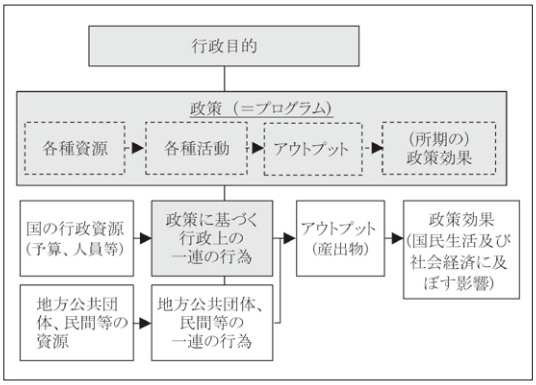
◎行政機関が行う政策の評価に関する法律（平成13年法律第86号）（抄）　※下線は筆者が追加

（政策評価の在り方）
第三条　行政機関は、その所掌に係る政策について、適時に、その政策効果（当該政策に基づき実施し、又は実施しようとしている行政上の一連の行為が国民生活及び社会経済に及ぼし、又は及ぼすことが見込まれる影響をいう。以下同じ。）を把握し、これを基礎として、必要性、効率性又は有効性の観点その他当該政策の特性に応じて必要な観点から、自ら評価するとともに、その評価の結果を当該政策に適切に反映させなければならない。
２　前項の規定に基づく評価（以下「政策評価」という。）は、その客観的かつ厳格な実施の確保を図るため、次に掲げるところにより、行われなければならない。
一　政策効果は、政策の特性に応じた合理的な手法を用い、できる限り定量的に把握すること。
二　政策の特性に応じて学識経験を有する者の知見の活用を図ること。

が国民生活や社会経済に及ぼす影響、すなわちアウトカムを指している。そして、その把握に当たっては、政策の特性に応じた合理的な手法を用いること、できる限り定量的に行うこと、学識経験者の知見を活用することが定められており、これはEBPMにおけるエビデンスの創出に関する考え方と軌を一にしている。言い換えれば、EBPMで求められている内容は、既に法律レベルで各行政機関に義務付けがなされていると考えられるのである。

さらに、政策評価法において「政策」を定義する際、政策の構造を整理する上でロジック・モデルの考え方が下敷きにされている。政策評価法第2条第2項により、「政策」とは、「行政機関が、その任務又は所掌事務の範囲内において、一定の行政目的を実現するために企画及び立案をする行政上の一連の行為についての方針、方策その他これらに類するもの」と定義されているが、この規定は、政策の構造を図10のようなものとして理解している。これによると政策は、資源（input）、各種活動（activity）、アウトプット（output）、政策効果（outcome）というロジックを想定したプログラム（方針・計画・方策等）であり、これに基づいて国の行政資源を使って行政上の一連の行為を行い、地方公共団体や民間等の活動とも相まって、最終的に政策効果を

図10　政策評価法における「政策」の構造



（出所）総務省作成資料（2000年）を基に筆者作成

発現させるもの、という整理である。ロジック・モデルの考え方は、政策評価制度が評価の対象となる「政策」を認識する際の土台となっているものであり、ここでもEBPMの主要な要素の一つが政策評価制度の中に組み込まれていると言えるのである。

4-2. 政策評価制度への具体的な取り込み方法

以上の考察から、本節の冒頭に提起したEBPMの考え方を取り込む既存の制度としては、この政策評価法に基づく政策評価制度が最有力候補の一つと考えられる。以下では、試論として、政策評価制度における既存の取組への取り込み方法について検討する。

政策評価法上、政策評価の具体的なやり方については何ら規定がない。評価を行う時点により、既存の政策について行う「事後評価」と未だ存在しない新規の政策について行う「事前評価」の区別があるだけである。政策評価は政府の内部マネジメントの仕組みであるため、その具体的なやり方は国会の統制に服す法律ではなく、閣議決定以下の政府部内の取り決めの中で方針が定められており、比較的柔軟に取り組むことができるのである。現在の方針の下では、政策評価は、政策の特性等に応じて合目的的に、「事業評価方式」、「実績評価方式」、「総合評価方式」という三つの方式や、これらの主要な要素を組み合わせた方式等で行うこととされている。これら三方式は標準的な取組にすぎないため、

EBPMを念頭に各府省が全く独自のやり方をしても構わないし、政府全体で新たな方式を構築してもよいが、その場合にはそれなりのエネルギーとコストがかかる。そこで、まずは既存の三方式をベースとしたEBPMの取り込み方法に関するアイデアを示したい。

三方式のうち、EBPM的な分析アプローチを最も組み込みやすいと考えられるのが、自由度の高い総合評価方式である。総合評価方式は、特定の政策テーマについて、政策効果の発現状況を様々な角度から掘り下げて分析し、政策に係る問題点を把握するともに、その原因を分析するなど総合的に評価する方式であり、いわゆるプログラム評価に相当する概念である。現時点では、総合評価を実施している府省は僅少だが、各府省では、中長期計画の改定時や大きな制度改正等に当たり、政策の効果について様々な情報・データを集めて検討しているはずであり、その営みをEBPM的視点で改良し、総合評価方式の政策評価として取り組むのが現実的であると考える。

また、実績評価方式の一種である目標管理型評価は各府省で最も普及している方式であり、この目標管理型評価にどのようにEBPMの要素を取り込むかが大きな課題となる。この方式は、評価対象とする施策を分析した上で目標を設定し、その達成状況をモニタリングするとともに、3 ～ 5年程度経過した後に総括的な評価を行うものである。例えば、目標設定に当たり事前分析表を作成する段階で政策のロジックを整理しエビデンスとなる指標を厳選する、あるいは、総括的な評価を行う際に踏み込んだ分析を行うなど、いくつかパターンが考えられる。各府省のEBPMの普及・定着状況を見つつ合理的な取り込み方策を考えることが肝要である。

さらに、事業評価方式についてもEBPMの視点から光を当てることが可能である。事業評価方式は、研究開発・公共事業・政府開発援助（ODA）・規制・租税特別措置の5分野に義務付けられている事前評価と密接にリンクした方式であり、例えば、公共事業評価では、事業採択前の段階で事業に要する費用と便益を金銭価値化して比較する費用便益分析の手法が確立しているが、これなどはまさにエビデンスを創出するための取組にほかならない。後は、上位の施策や同一地域の他事業との整合性確保等、ロジック面での取組を強化することでEBPMの考え方を取り込むことができる。一方、規制や租税特別措置の評価では、政策効果の把握が十分でない評価書が数多く見られるため、まずはエビデンス面での質の向上が課題であろう。

5. まとめ

政府において「EBPM」と銘を打っての取組は緒についたばかりである。一方、EBPMの取組を安定的に普及・定着させるプラットフォームとして有力な政策評価制度は2001年からの積み重ねがある。この両者の関係について、EBPMを推進する立場からは、EBPMの普及・定着のための苗床として法律に基づく安定的な仕組みである政策評価制度を利用し、政策評価を推進する立場からは、「低位安定」が指摘される政策評価制度の活性化の起爆剤としてEBPMを活用することで、相互に恩恵が受けられるものと考える。特に、これまで結果として説明責任を果たすという目的を重視した運用がなされてきた政策評価にとっては、もう一つの目的である効果的・効率的な行政の実現に向け、真に政策マネジメントの機能向上を図るツールとして脱皮するきっかけになり得るのではないか。

もともと政策評価制度は、完成形というものがなく、評価手法に関する国内外における研究の動向や各府省の取組の進展等に伴って進化し続ける制度である。EBPMのような考え方を取り入れて進化することは、政策評価制度が元々予定していたところのものと言える。当面はEBPMの実例創出に注力しつつ、政府全体の取組状況を見ながら政策評価制度を活用してEBPMの制度的基盤を確保することで、EBPMと政策評価制度の間でWin-Winの関係が構築できるものと考える。

注記

1 〒100-8926 東京都千代田区霞が関2丁目1－2 ysunayama@soumu.go.jp
2 2018年11月13日衆議院総務委員会における石田総務大臣の所信表明演説ほか。
3 官の名称は府省によって異なる。
4 「統計改革推進会議最終取りまとめ」（2017年5月統計改革推進会議）
5 http://www.soumu.go.jp/main_content/000579366.pdf
6 http://www.soumu.go.jp/menu_news/s-news/01hyoka01_02000031.html
7 http://www.soumu.go.jp/main_content/000617155.pdf
8 http://www.soumu.go.jp/main_content/000617157.pdf
9 ここでは、個人単位ではなく、セミナーというクラスター単位でランダム化して効果を測定している。
10 http://www.soumu.go.jp/main_content/000624754.pdf
11 http://www.soumu.go.jp/main_content/000617153.pdf
12 http://www.soumu.go.jp/main_content/000630779.pdf
13 http://www.soumu.go.jp/main_content/000590974.pdf
14 http://www.soumu.go.jp/main_content/000233774.pdf
15 http://www.soumu.go.jp/main_content/000627735.pdf

参考文献・ウェブサイト

総務省（2019）『女性活躍の推進に関する政策評価書』
総務省（2019）『高度外国人材の受入れに関する政策評価書』
総務省（2019）『政策効果の把握・分析手法の実証的共同研究報告書』
総務省（2018）『年金業務の運営に関する行政評価・監視－国民年金業務を中心として－結果報告書』
総務省（2018）『EBPM（エビデンスに基づく政策立案）に関する有識者との意見交換会報告（議論の整理と課題等）』
統計改革推進会議（2017）『統計改革推進会議最終取りまとめ』
総務省（2013）『ワーク・ライフ・バランスの推進に関する政策評価書』
政策評価ポータルサイト http://www.soumu.go.jp/main_sosiki/hyouka/seisaku_n/portal/index.html

（2020.2.5 受理）

Study on Evidence-Based Policy Making (EBPM) in practice in the Government of Japan

Yutaka Sunayama

Ministry of Internal Affairs and Communications
ysunayama@soumu.go.jp

Abstract

The Japanese government has engaged in creating, disseminating, sharing and accumulating concrete examples of EBPM, which is one of the most immediate issues to address concerning administrative management in Japan. In the process of the on-going joint study on EBPM among the Ministry of Internal Affairs and Communications (MIC), related ministries and academics, some elements that are effective and useful for EBPM of the government are being identified. In addition, the MIC has conducted several surveys of government programs and their operations with empirical analyses using statistical data to get evidence, which contributes to boosting the government's efforts in promoting EBPM. It is indispensable hereafter to arrange a stable institutional infrastructure to establish EBPM within the government. In this respect, the policy evaluation system introduced in 2001 could be one of the most prospective options.

Keywords

logic model, causal inference, empirical joint study,
Administrative Evaluation Bureau survey, policy evaluation system

【実践・調査報告】

地方自治体におけるEBPMの組織への定着に向けた課題の考察

西畑　宏治[1]
広島県
k.j.nishihata@gmail.com

要　約

近年、地方自治体においても「EBPM」を進める動きが見られるようになった。今後もこうした動きが地方自治体の中で広まるものと考えられるが、その内容が各地方自治体による散発的なエビデンスの創出に留まることなく、各地方自治体において仕組みとして組織に定着が図られることが必要であると考える。そこで、EBPMの組織への定着に向けて、EBPMが要請する行動様式を地方自治体の行政運営に適用しようとした場合、地方自治体のどこに課題が生じ、それを解決するためにどのようなことを検討する必要があるのか、また、どのような解決策が考えられるのかを明らかにしたい。その際、組織への定着に着目した動きを進めている広島県の取組も参考にする。

キーワード

地方自治体、EBPM、組織への定着、広島県

1. はじめに

近年、行政の現場で「EBPM（Evidence-Based Policy Making：エビデンスに基づく政策形成）」という言葉が聞かれるようになり、取組が少しずつ進められている。

国においては、2017年8月のEBPM推進委員会発足といった統一的な動きの下、各府省でEBPMに関する取組が進められている。

地方自治体（以下「自治体」という。）[2]においては、国のような統一的な動きはないものの、「EBPM」の言葉の下に取組を進めるところも出てきている。現時点では、取組を進めている自治体は少数であるが、今後、取組を進める自治体の数は増加すると予想される。

本稿では、後述するように、エビデンスとは、政策介入の有効性を検証する因果推論の実証結果、EBPMについては、政策[3]を形成する際に、このエビデンスを十分に意識することで、より妥当な解決策を選択していこうとするものと理解することとする。

こうした政策過程の科学化・合理化は過去にも試みられており、その経験を踏まえつつ、政策をより良いものにしていくため、自治体はEBPMが求める行動様式をどう取り込んでいくことができるのか、本稿では前向きに考えてみたい。

自治体がEBPMに取り組む場合、その取組方に関して、ある特定の政策における実践のみに終わらせるのではなく、組織への定着を目指して取組を進めるべきであろう。

日本評価学会『日本評価研究』第20巻第2号、2020年、pp.65-76

そこで、EBPMの組織への定着に向けて、自治体のどこに課題が生じ、それを解決するためにどのようなことを検討する必要があるのか、どういった解決策が考えられるのかを明らかにしたい。これによって、今後EBPMに取り組もうとする自治体は参考にすることができる。

このため、EBPMが要請する行動様式を基に、従来の自治体における行政運営との違いが表れる点を考察する。解決策の検討に際しては、組織への定着に着目して取組を進めている広島県の取組も参考にする。

2. 自治体におけるEBPMの動向

まず、昨今の「EBPM」という言葉が出始めて以降[4]の自治体におけるEBPMの動向について、確認しておきたい。

積極的な自治体は早い時期からEBPMの推進を模索してきたと思われるが、そうした動きが広く知られるようになってきたのは、2018年度後半以降と考えられる。

それは、自治体自らがシンポジウムの開催や報告書の公表等を行って、情報発信を積極的に進める動きを見せるようになってきたからである。

例えば、2018年11月30日に広島県がEBPMに関するシンポジウムを開催し、12月14日に独立行政法人経済産業研究所主催で開催されたシンポジウムでは、埼玉県の取組が紹介されている。2019年1月29日には神奈川県がフォーラムを開催し、3月には北海道が報告書をまとめ、公表している。また、総合計画にEBPMという言葉を入れる自治体も出てきている（例えば、生駒市）。

このように情報発信も含め積極的な動きを進める自治体が出てきている一方、EBPMに関する動きが見えない自治体も未だ多いと思われる。例えば、三菱UFJリサーチ＆コンサルティング自治体経営改革室（2019）によると、2018年12月～2019年1月の調査時点で、回答のあった都道府県・市・特別区合わせて465団体のうち、EBPMを「推進している」と回答したところは5.6％に留まっている。

なお、積極的な動きは、広域的自治体[5]の他、基礎的自治体の中では政令指定都市や特別区の中で多く見られる（三菱UFJリサーチ＆コンサルティング自治体経営改革室2019）。

以上から、自治体の動向として、EBPMの取組は緒に就いたばかりであること、積極的な少数の自治体と動きが見えない多数の自治体という二極化の様相を呈していること、積極的な動きは、広域的自治体や基礎的自治体の中でも規模の大きな自治体の中で目立つということがわかる。

また、各自治体が「EBPM」と呼んでいる取組の中身を見ると、自治体によって意味することが異なることもわかる[6]。

こうした現状にある自治体のEBPMであるが、国や先行する自治体の動き等もあり、自治体の間で「EBPM」という言葉は1つのキーワードとして認知されつつあると考えられる。また、2020年度を初年度とし今後5か年の目標や施策の方向性等を内容とする「第2期「まち・ひと・しごと創生総合戦略」」においても「EBPM」という言葉が引き続き使用されていること等を踏まえると、しばらく「EBPM」の流れは変わりそうにない。このため、今後も「EBPM」の言葉の下、取組を進める自治体の数は増加すると予想される。

3．EBPMと自治体の関係

(1) 本稿におけるEBPMという言葉の定義

前述のとおり自治体内で「EBPM」は様々な理解がなされている状況にあると思われるが、本稿では、「EBPM」という言葉の根幹にあるエビデンスを因果推論の実証結果（津富2000）、公共政策の文脈で言えば、政策介入の有効性を検証する因果推論の実証結果と定義する。すなわち、インパクト評価による検証結果である。つまり、ここでのエビデンスは、政策について、ある特定の文脈における有効、無効、ないし逆効果を明らかにしているものであり、現状を把握するためのデータとは区別することとする。

そして、EBPMについては、青柳（2019）が述べるように、政策を形成する際に、このエビデンスを十分に意識することで、より妥当な解決

策を選択していこうとするものと理解する。

この理解では、具体的には、エビデンスを十分に参照する、もし既存のエビデンスが不十分であれば、必要なエビデンスを自らつくるという行動が求められる。

こうした定義により、EBPMのもたらす新規性・革新性や付加価値が明らかになる。それは、青柳（2018b）が源流であるEBMに照らして指摘するように、政策介入の有効性を明らかにした実証結果を政策形成・行政運営の場に持ち出したということである。これにより、エビデンスを踏まえることで、この政策を実施しても公共的問題が解決（改善）しないかもしれないという不確実性をできる限り低減させることが可能になると考えられる。

なお、エビデンスは意思決定に用いられる重要な情報となるが、このEBPMの理解では、エビデンスが全てではないとされており、実施コスト、関係者の意向等も踏まえた総合的な見地から意思決定がなされるという点にも留意しておく必要があるだろう（青柳2018b）。

(2) EBPMと自治体の関係についての整理

こうしたエビデンス、EBPMの定義を踏まえたとき、自治体の政策形成との基本的な関係を2つ考えておきたい。

① 自治体がEBPMに取り組むことの意義

自治体は、地方自治法に規定されるように、住民の福祉の増進を図ることを基本として（第1条の2）、その事務を処理するに当たり、最少の経費で最大の効果を挙げるようにしなければならない（第2条第14項）とされている。政策は、まさに住民の福祉の増進を図るために実施されるものである。そして、前述のようにEBPMは、その政策形成に当たり、政策介入についての有効性、すなわち効果に対して正面から向き合うことを求めるため、自治体がEBPMを進めることは、自治体の存立目的や運営の基本原則とも合致すると考えられる。

また、自治体は、住民に身近であるからこそ、対処が必要な公共的問題を的確に捉えて、問題の分析から解決策の考案までを行うことができる立場にあり、そこに有効性の観点を加えることで、政策形成の質が向上することも期待できる。

このようなEBPMと自治体の関係を踏まえると、EBPMに取り組む自治体が増えることは望ましいと言えるだろう。

② 取り組む上での歴史的な留意点

しかしながら、過去の歴史を踏まえれば、取り組む上で留意すべきことがある。

EBPMは、政策過程の科学化・合理化を推進していく考え方なり手法の1つとされる（山本2019）。

こうした科学化・合理化の取組は、過去にもあった。アメリカにおけるPPBS（Planning Programming Budgeting System）である。このPPBSは、1968年度予算から連邦全省庁に導入され、1971年には廃止されることになったが、その失敗の原因には、予算の政治的意思決定過程の考慮が不十分であったことや政府活動の実際との不一致、実施上の負荷等が挙げられている（秋吉・伊藤・北山2010）。

こうした経験からは、山本（2019）が指摘するように、行政の決定は、政治性と客観的証拠等の合理性の2つのバランスの結果であることに留意する必要があるだろう。また、実際の行政活動を踏まえた実現可能性の模索が求められるだろう。

現在、自治体が取り巻かれ、しばらく変わりそうにない「EBPM」の流れの中で、政策をより良いものにしていくために、自治体はEBPMが求める行動様式をどう取り込んでいくことができるのか、こうした過去の経験も踏まえながら、本稿では前向きに考えてみる。

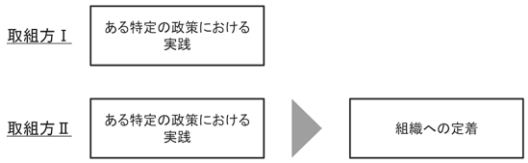
4. EBPMの組織への定着に関する考察の準備

(1) 自治体によるEBPMの取組方

① 想定される取組方の選択肢

自治体がEBPMを推進する場合、図1のような2つの取組方が想定される。

図1　EBPMの取組方



（出所）筆者作成

取組方Ⅰは、ある特定の政策における実践である。

取組方Ⅱは、組織への定着を見据えながらの取組方である。組織への定着とは、次のように考えておく。組織内の職員にEBPMの必要性が認識され、エビデンスを意識し扱う行動様式（エビデンスを巡ってPDCAサイクルを回すことができるとも言える）が身についている状態である。意思決定者においては、様々な分野で数多く実施される自治体の政策のうち、どの政策についてエビデンスを求めるかといった判断ができ、検討に上ったエビデンスを意識した上で政策案選択の意思決定ができるといった状態である。

②　2つの取組方の共通点と相違点

どちらの取組方もある特定の政策における実践が必要である点は共通している。つまり、特定の政策について効果があるものを選択していこうとし、効果がないものは選択しないという意識を持っている点で共通している。

違いは、自治体がその実践のみに留まるか、先を見据えているかである。

前者の場合、どのようなことが起こり得るだろうか。1つには、自治体が住民の福祉の増進を図っている中、ある政策でのみエビデンスを創出したとしても、エビデンスを意識し扱う行動様式がその組織内に定着していかないようであれば、その自治体の他の政策ではエビデンスが意識されず、結局、当該自治体総体として、継続的に住民の福祉の向上に結び付く可能性がなくなってしまうと考えられる。

もう1つは、現在の「EBPM」という言葉の定義がはっきりしないまま進んでいる状況は、自治体における「EBPM」が今後どこに進んでいくのかわからない状況を表しているとも言え、ここに、エビデンスを意識し扱う行動様式が各自治体組織内に定着しない状況が加わるとすれば、「EBPM」という言葉が意味するものの換骨奪胎が進み、形骸化する恐れも出てくると考えられる。

これらを踏まえ、自治体がEBPMに取り組む際に後者の組織への定着を目指す姿とするとしても、この場合もそう簡単に話が進むわけではない。これは、組織を変えていくという性質の話であり、困難さが増すことや時間がかかることが想定される。

以上から、自治体は組織への定着という目指す姿を共有した上で、それをどのように実現していくかに知恵を絞ることが必要になるだろう。

(2) 組織への定着に向けた検討箇所

組織への定着に向けて、図1の取組方Ⅱを採用したとき、図2のように、検討が必要な箇所は大きく2つ考えられる。

図2　組織への定着に向けて検討が必要な箇所

（出所）筆者作成

「検討箇所①」は、ある特定の政策における実践に関して必要な検討が該当する。図1の取組方Ⅰとも共通するところである。

「検討箇所②」は、個別の実践から組織への定着へとつなぐために必要な検討が該当する。

(3) 検討箇所における課題等の考察方法

組織への定着に向けては、これらの検討箇所について、どのような課題が生じるのか、それを解決するためにどのようなことを検討する必要があるのか、どういった解決策が考えられるのかを明らかにする必要があるだろう。

このため、EBPMが要請する行動様式を基に、従来の自治体における行政運営と違いが表れる点を考察する。解決策の検討に際しては、実際

の自治体の取組も参考にする。

自治体の取組としては、前述のエビデンスやEBPMの定義を踏まえた上で、組織への定着といった点に焦点を当てているものを探す必要がある。この観点で現在の自治体の取組を確認すると、広島県が候補になると考えられる[7]。

5. 広島県の取組の確認

ここで、考察に入る前に、広島県の取組を確認しておきたい。

第2節で述べたとおり、広島県は、2018年11月30日に、EBPMに関するシンポジウムを全国の自治体に先駆けて開催している。

そこでの説明や2019年度の業務委託の内容から取組に次のような特徴を見出すことができる（広島県2018a、2018b、2019）。

第1に、エビデンスを政策介入の有効性を検証する因果推論の実証結果と捉え、エビデンスの活用は、論理とファクトに裏付けられた政策案に実証性を与えるものとして、従来の政策形成と一体的なつながりの中で理解がなされている。

第2に、組織への定着を見据えた検討とそのための取組が進められている。

例えば、これまでに広島県が有している制度的基盤（「施策マネジメントシステム」）にエビデンスの活用をどう織り込んでいくかが具体的に検討されている他、エビデンスに高い質を求めることで得られるベネフィットと費用や迅速性とのバランスを重視し、どのようなプロセスが望ましいかが検討されている。また、職員の育成も重視されている。

実際の取組としては、まず、県内部のプロセスの整備が進められている。例えば、2019年度は、職員が新規の政策の形成や既存の政策の再構築を行うに当たって参照可能なロジック構築に係る留意点を整理した「ロジック構築プロセス書」の作成や既存のエビデンスを参照するための手順・運用方法を整理した「既存エビデンス活用プロセス書」の作成が予定されている。

また、職員の育成の取組も進められている。例えば、企画部門の職員らを対象とした政策形成やデータ分析の研修の他、管理職を対象としたEBPMに関する基礎的な内容の研修等である。

第3に、具体的に取組を進めるための仕掛けとして、ワーキンググループ（「広島県EBPM推進ワーキンググループ」）を設置している。

このワーキンググループは、県の内部部局のうち、関係する複数の所属（関係事業局、経営企画チーム、統計課、財政課、人事課等）と外部の専門家（学識者、コンサルティング会社）で構成されている。県の複数の政策を対象にインパクト評価を試行するとともに、この結果や取組状況を踏まえ、前述のプロセスの整備に反映させる他、前述の職員育成の取組を担っている。

以上から、広島県の取組を改めて整理すると、図2の検討箇所①に対応する取組（個別の政策を対象としたインパクト評価の試行）と検討箇所②に対応する取組（複数の政策を対象としたインパクト評価の試行、プロセス整備、ワーキンググループ関係者に留まらない人材育成等）を進めていると言えるだろう。

6. EBPMの組織への定着に向けた課題等

ここからは、EBPMが要請する行動様式を自治体の行政運営に適用しようとした場合に、自治体にどのような課題が生じ、それを解決するためにどのようなことを検討する必要があるのか、また、どのような解決策が考えられるのかについて考察する。その際、広島県の取組も参考にする。

なお、1,700を超える自治体のできるだけ多くに共通的に当てはまると考えられる主な点を挙げることとする。

（1）自治体内部における課題等

① EBPMに取り組むための前提となる発想

青柳（2019）は、国際開発分野での実践を踏まえて、EBPMの前提として、アウトカム志向になることが不可欠な第一歩であると指摘する。

たしかに、そもそも公共政策とは、「公共的問題を解決するための解決の方向性と具体的手段」

（秋吉・伊藤・北山2010）であるから、政策の形成に当たっては、ロジックモデルで言うところの問題が解決（改善）された望ましい状態を表すアウトカムを達成することに焦点が置かれるべきであろう。この場合、例えば、政策案を議論する際は、「この政策案が果たして問題を解決ないし改善するのか」が基本的な視点となる。このようなアウトカム志向での議論や発想が自治体においてなされているかどうかがまず問われるべきだろう。

例えば、広島県を見てみると、企業理念に相当する「広島県職員の行動理念」を策定し、これの職員への浸透を図る取組をこれまで積極的に行ってきているが、この中に、県民を起点とすること、成果を志向することが明確に謳われている（広島県マネジメント研究会2016）。

アウトカム志向を持つことで、政策案は公共的問題を解決（改善）しないかもしれないという不確実性を内包すること、また、第3節で述べたように、エビデンスを活用すれば政策案が内包するこの不確実性を低減させることが可能になるということが、エビデンスという言葉の定義も含め、違和感なく認識されることになるだろう。

②　EBPMに取り組むきっかけ

前述のアウトカム志向を職員が持ったとしても、それだけで組織がEBPMに取り組むようになるわけではなく、何かのきっかけが必要となるだろう。

EBPMに取り組むきっかけとして、1つは、議会等、外部からの働きかけが考えられる。一方、自治体自らがEBPMに取り組むきっかけとして、他自治体との競争もない中にあっては、自治体内部のトップ、マネジメント層の必要性の認識、理解とその上でのリーダーシップがあると考えられる。

本稿で焦点を当てているEBPMの組織への定着という視点に立てば、従前の自治体組織の文化に変革を迫るものとも言えるため、この点からも、トップやマネジメント層のリーダーシップは不可欠と考えられる。

例えば、広島県を見てみると、広島県知事の湯﨑英彦氏の執筆した2017年6月15日の記事ではEBPMの考え方や必要性が述べられている（湯﨑2017）。また、前述の2018年11月30日の広島県主催のEBPMに関するシンポジウムでは、知事自らがEBPMやEBPMの組織への定着の考え方について講演しており[8]、知事のEBPMの必要性の認識と理解、リーダーシップを読み取ることができよう。

なお、こうしたトップ、マネジメント層の必要性の認識、理解とその上でのリーダーシップは、取り組むに当たってのきっかけとして機能するのみならず、取組を進める過程で、その推進を支える機能も果たすことになると考えられる。

③　関与する主体、主体間の関係

まず、EBPMは政策形成そのものを対象としていることを踏まえると、各政策を担当する事業系組織が主体であることは当然であるが、そこに統計を所管する部署のみが関与すれば良いというわけではなく、政策形成に関与する企画部門や財政部門といった総務系組織が積極的に関与することが欠かせないと考えられる。

例えば、広島県を見てみると、前述のとおり、ワーキンググループを設置し、統計課のみならず、企画部門や財政部門、人事部門といった総務系組織もメンバーとする体制を整えていた。

次に、企画部門や財政部門という総務系組織の中を見ると、前者は自由な思考から行動をとりがちであるが、後者はある程度現実主義的な行動をとると言われている（柴田・松井2012）。

こうした中、自治体が自ら政策介入の有効性の検証を行う動機の1つとして、その結果を活かして政策形成の意思決定を行うことにあると考えた場合、多くは、組織的な意思決定の1つである予算編成に関心が向くこととなる。この予算編成は財政部門が中心となるが、財政部門の主たる関心は、一般財源のやりくりと言われ、財源総額である「枠」を重視することとなる（磯崎・金井・伊藤2014）。この結果、財政部門では財政的な厳しさも加わり、予算削減の方向を持ちやすくなると考えられる。

しかしながら、エビデンス自体は、効果があ

る場合、ない場合、逆効果の場合も含めた結果で、価値中立であり、必ずしも削減のために使われるべきものではないことに留意すべきだろう。また、削減ありきでEBPMを進めるのであれば、主役である事業系組織は、EBPMを進めるがゆえに予算を削減されることとなり、誰も取り組まなくなるだろう。

EBPMを推進する中では、事業系組織が向き合っている公共的問題をいかに解決するか、そのための手段は何が良いのかとアウトカム志向の議論を行う過程で、事業系組織には必要なエビデンスが供給され、総務系組織にはそのサポートやエビデンスを参照しながら議論を交わすといった在り方が求められていると言えよう。

財政部門や企画部門の在り方を改めて考える必要があるだろう。

④ EBPMの前提となる従来の政策形成

EBPMでは解決策の選択に当たり、エビデンスを十分に意識するということが求められる。そうすると、その解決策選択の前提として、対象とする公共的問題の特定や分析、解決策の選択肢の導出までが十分になされておく必要がある。ここは、従前の政策形成でも検討がなされてきた範疇であるが、既存のエビデンスを参照する段階やエビデンスをつくる段階に至るためには、その土台として、これまで以上に、この従来の政策形成で行われてきた検討を強化しておく必要がある。

このため、EBPMの推進に当たり、まず、当該自治体における政策形成の現状を振り返る必要が出てくるだろう。それは、組織の中に方法論が存在しているか、自立的な政策形成ができているか、政策の形成プロセスでボトルネックはどこか、といったことを見つめ直すことになる。また、公共的問題には悪構造性という特性があり、問題の構造化が重要な分析段階となるため、この段階の検討が十分かといった確認も必要であろう（秋吉・伊藤・北山2010）。例えば、広島県を見てみると、これまでの政策形成プロセスの中で、問題の構造化というプロセスを重視していることがわかる（広島県2018a）。

政策の形成プロセスの上流に当たる部分の強化がなされるようになれば、問題に対する理解と解決策の検討も十分になされており、既存のエビデンスを参照する際の検討やエビデンスをつくる段階の評価設問の整理に資することになるだろう。

⑤ 既存の政策評価制度との関係

自治体の多くは、政策評価の仕組みを持ち、その評価手法は業績測定である[9]。この政策評価とEBPMは混同されることが多いと思われるが、政策評価、業績測定とEBPM、そして、その根幹であるエビデンスを作り出すためのインパクト評価との関係をあらかじめきちんと認識しておくことが重要であろう。

政策評価は、「アジェンダ設定」、「形成」、「決定」、「実施」、「評価」という政策プロセスのうちの「評価」に力点を置き、「形成」や「実施」へのフィードバックを意図している。「評価」の段階で、業績測定が行われるため、事後評価となる。また、併せて、目の前にある政策の実施結果をもたらした原因を過去に遡って検証しようという思考がなされることになる。

一方、EBPMは、「形成」の段階に力点を置いて妥当な解決策を選択しようとするものであり、そこで検討がなされるインパクト評価は、政策という介入（原因）の結果を探ろうとする発想が基本となることから、発想の力点は事前評価に置かれることとなる。

このように、自治体における現行の政策評価、業績測定とEBPM、インパクト評価は異なる営みであることをまず認識する必要がある。

現在、多くの自治体は、政策評価、業績測定の発想になじむとともに、インパクト評価を適用する枠組みを内部に明確に設けていないと考えられるため、EBPMを進める際に、概念の混乱が生じることになると考えられる。

こうした状況を乗り越えつつ、EBPMを推進する中で、両者の折り合いをどうつけるかを検討する必要がある。それは、つまり、まず、どの政策についてインパクト評価を行ってエビデンスをつくるのか、つくる場合、どの程度の質（信頼性）のエビデンスを求めるのかといった判断やその判断を行う職員、手続きの決定、その判

断が活かされる仕組みの構築等の体制作りが必要である。そして、それ以外の政策は、従前の政策評価を行うということになるだろう[10]。

例えば、広島県を見てみると、エビデンスに高い質を求めることで得られるベネフィットと費用や迅速性とのバランスをとるために、高い質を求める政策を選定した上で、メリハリのある形成プロセスを検討している（広島県2018a）。

そうして出来上がった体制の中で、それぞれの評価の役割や思考法が混同されることのないよう、理解や認識を組織内で共有し、継承していく必要がある。

⑥　エビデンスをつくる場合の時間や場

エビデンスをつくるためには、一定の「時間」と「場」を確保する必要がある。自治体においてエビデンスをつくる意思決定がなされた場合、効果的な解決策を開発するためのテストに要する時間と場を確保する必要があるということである。

ここから、まず、こうした時間が自治体の行政運営に求められている時間とマッチするのかの検討が必要になるだろう。また、場を設けることも自治体の行政運営の方法にマッチするのかの検討が必要である。

これまで、政策実施と言えば、政策の適用範囲を自治体の有する区域全体や区域内の一定の条件を満たす全ての住民といった形で実施することを指す場合が多い。一方、エビデンスをつくる場合、こうした本格的な実施の前段階に、考案した効果が実証されていない解決策を小規模に実施し、効果の有無を検証して解決策を開発するというテスト段階を置くことになる。

自治体は、こうした発想をこれまで明確に意識して持っていなかったと考えられる。ただ、自治体もこれまで本格的な実施を前に「効果を検証する」等として「モデル事業」という実施形態をとることもあったため、青柳･小林（2019）も指摘するように、この「モデル事業」は自治体におけるエビデンスをつくる場の候補になると考えられる。

次に、エビデンスをつくる場合の時間や場の確保に当たり、制約ともなりうる行政計画について言及しておきたい。自治体は、総合計画以外にも、多くの個別の政策分野で、主に国からの要請を受けて計画を策定し、これに基づき政策を実施している[11]。この計画の中に政策が並べられ、長期に固定化してしまうことになれば、結果として、その計画期間中にエビデンスをつくるインセンティブが失われることになる。計画策定の在り方について再考が必要であるし、新たに計画を策定（計画の更新も含む）する際は、エビデンスをつくる時間や場を確保できるよう柔軟性を持たせる工夫を検討すべきであろう。

⑦　意思決定におけるエビデンスの扱い方

EBPMでは、エビデンスを十分に意識して政策形成に当たっての意思決定を行うこととなる。とはいえ、「エビデンスを意識するように」と急に言われても戸惑うのが当然である。

あるエビデンスを得た場合、まず、結果の読み方がわからないという統計学的知識やエビデンスリテラシーの不足に起因する問題が生じるだろう。このとき、結果の読み取り、評価デザインを踏まえたエビデンスの質の理解が必須であるし、統計的に有意ではない場合、エビデンスの質が低い場合、外的妥当性等の理解と扱いの判断が必要になる。さらに、エビデンスの解釈には、統計学的知識に加えて個別の政策分野の知識も求められることから、個別政策分野に関する専門的知見の研鑽も必要となるだろう。

次に、担当の職員がそこまでできるとしても、その上司あるいは組織内で、エビデンスの考え方や扱い方が共有されていなければ、結局、組織の意思決定時に意識されないこととなる。

組織への定着に向けては、あらかじめ、組織内でエビデンスを何に使うのか、どう使っていくのかについての判断基準を組織の構成員の間で共有しておく必要があるだろう。職員レベル、組織レベルでエビデンスに対する一定の理解や認識が共有され、エビデンスが意思決定時に意識される土壌が求められる。ここからも、②で述べたトップやマネジメント層の認識、理解とその上でのリーダーシップが少なくとも必要になる。

⑧　政策形成に携わる職員

⑦とも関係するが、組織への定着という視点に立てば、政策形成に携わる職員に注目し、長い目で見て育成や人材の確保を進めることが重要になるだろう。例えば、広島県を見ると、前述のとおり、職員の育成が重視されていた。

④で述べたように、EBPMを推進する中で、職員には、従来の政策形成で行われてきた検討がこれまで以上に求められる。

また、全ての職員が同じ水準のスキルを有している必要はないが、実証分析に関する理解も求められる。政策介入の有効性の検証やデータ分析を行う際には、評価学や統計学、計量経済学に関する理論的な知識はもちろんのこと、実務上直面する時間やコストといった制約等を考慮しつつ必要とされるエビデンスの質に応える評価デザインを組み立てる能力が求められる。

しかし、そもそも自治体の評価の現状に関しては、2～3年毎の人事異動でやってきた評価の素人がにわか勉強の自己学習で評価業務をなんとかこなしていることや評価学の専門に習熟した公務員が多いとは思われないといった指摘もある（山谷2016）。

こう見ると自治体内部での育成はハードルが高いように思われるが、EBPMに取り組む場合、必要となるスキルの全体像を明らかにし、どのポジションにどのようなスキルを求めるのか、人材を内部で育成するのか、採用するのか、外部委託といった手段をとるのか、それは研究者かコンサルタントか、その実現可能性はどうか、といった点について議論を行う必要が出てくると考えられる。

なお、内部で育成せず、外部の専門家を頼る場合でも、自治体側が専門家の力量を見極めて、適切なパートナーを得ていく必要がある。そうなると、当然、自治体の少なくとも一部の部門には実証分析をはじめとしてEBPMを理解する職員を擁していることが求められるだろう。

(2) 自治体外部における課題等

組織への定着に向けては、自治体内部に留まらず外部にも課題等が存在すると思われる。

①「EBPM」という言葉・概念の定義

1,700を超える自治体がある中、「EBPM」という言葉が定義されていない状況は、EBPMを議論するための共通言語が存在しない状況であると言える。

何を以ってEBPMに取り組んでいると言えるのかがわからないままであるならば、前に踏み出そうとしている自治体に不必要な混乱を招き、結局踏み出せないといったことにもなりかねない。そればかりか、踏み出したとしても、易きところに流れ、そこで安住してしまう（例えば、ロジックモデルの作成のみを以ってEBPMに取り組んでいると言う等）ことにもなりかねない。

まず言葉・概念を定義・明確化することが必要であり、共通言語や共通の認識の下、実際に取組を進めるための知恵を自治体間で出し合うことが必要だろう。

なお、筆者は、第3節で述べたように、エビデンスを政策介入の有効性を検証する因果推論の実証結果という定義で考えた方が良いと考える。なぜなら、方法論として認識が共有されている実証分析を基本に、概念と対応させることで、EBPMという言葉の定義が可能となり、EBPMがもたらす新規性や付加価値を明確化できるからである。また、この明確化の結果、従来の政策形成プロセスが際立ち、そこで何をすべきかもより明確になると考えられる。

②「エビデンスをつかう」ための環境

EBPMを推進する中では、エビデンスをつくるだけでなく、エビデンスをつかうことに取り組む必要が出てくるだろう。

この点について、イギリスであればWhat Works Center、アメリカであればWhat Works Clearinghouseといった組織が政策分野毎に既存のエビデンスを整理し、わかりやすい形で提供しており、実務家でもエビデンスをつかうことのできる環境がインフラとして整備されている。

一方、日本では、現状、エビデンスをつかうための機会の提供が限られている。

こうした中で、例えば、広島県では、前述のとおり、2019年度に形成段階において、既存のエビデンスを参照するプロセスの整備を図ろう

としていた。

このように各自治体がエビデンスをつかう取組の必要性を認識して独自に取り組む場合も想定されるが、全自治体の共通基盤といった形でエビデンスをつかう、つかいやすくする環境が整備されていかなければ、将来、自治体のEBPMは構造的に行き詰まってしまう可能性がある。

エビデンスをつかうという点に関しては、研究者との連携・協働も論点の1つであろう。個別の政策分野に関するエビデンスについて詳しいのはその分野の研究者であり、自治体がEBPMを推進する上で、研究者との連携・協働は重要である[12]。

自治体においては、これまでも審議会や研究会といった形で、研究者との接点は存在してきたが、両者がエビデンスを意識して議論する機会は限られていたと考えられる。自治体がEBPMを推進するに当たっては、支援を得られるような連携・協働体制を構築することが望まれる。とりわけ物理的に近接した場所に位置する大学等の研究機関は候補であろう。

しかし、こうした意識を自治体が持つだけでは不十分で、研究者、研究機関の側の意識や取組も必要になる。

両者は地道に連携・協働を模索することとなるだろうが、将来的に、地域の高等教育機関と自治体等の恒常的な連携体制の構築が進められることも想定される（中央教育審議会2018）。例えば、この中で、EBPMの文脈から両者の連携・協働も議論の俎上に上り、その契機となることが望まれる。

7. 最後に

EBPMの組織への定着に向けて、EBPMが要請する行動様式を自治体の行政運営に適用しようとした場合、自治体のどこに課題が生じ、それを解決するためにどのようなことを検討する必要があるのか、また、どのような解決策が考えられるのかについて考察してきた。

これらからEBPMの組織への定着を図ることの難しさを改めて確認することができるが、難しいからこそ、今後易きところに流れて安住してしまうことのないよう、腰を据えて取り組む必要がある。また、取組を進めるための知恵を自治体間で出し合うことも必要だろう。

昨今の「EBPM」の流れの中では、EBPMの付加価値が自治体において継続的に発揮される状態を目指し、模索していくことが大切だと思われる。その過程で、本稿の内容が少しでも役に立てば幸いである。

なお、本稿の内容は、自治体の執行機関を対象としている。地方自治は、その統治構造上、二元代表制と言われており、政治との関係は自治体の政策形成を語る上で重要な論点である。この点は今後の検討としたい。

注記

1 本稿において示す考えは筆者の個人的見解であり、所属する組織の見解を示すものではない。

2 本稿の地方自治体は、都道府県・市町村・特別区を念頭に置いている（都道府県47、市792、町743、村183、特別区23の計1,788（2018年10月1日現在））。

3 一般的に、自治体の政策は、政策－施策－事業の3層構造で理解されるが、本稿の「政策」は、このうち、施策、事業レベルに近い。

4 自治体と関係する「EBPM」の動きが見られ始めたのは、2016年12月頃と考えられる（西畑2020）。

5 広域的自治体では、前述の自治体の他、岡山県、滋賀県、徳島県が筆者の把握している範囲で動きがある。

6 こうした自治体の状況の背景として、西畑（2020）は、「因果推論を踏まえた独自の流れ」、「IT、データ利活用の流れ」、「統計改革の流れ」、「地方創生の流れ」の4つを指摘する。なお、国を中心とした状況については、田辺（2018）の整理や青柳（2018a）の指摘が参考になる。

7 西畑（2020）が指摘する因果推論を踏まえた流れに該当する自治体は、広島県、葉山町、岡山県等が考えられ、少数であると思われる。このうち、組織への定着に焦点を当てている自治体として公表資料から確認できるのは広島県である。なお、広島県の取組は、「EBPM」の事例が整理されている神奈川県政策研究センターによる事例集の中で、「組織・人

材」に分類されている（神奈川県政策研究センター2018）。

8 当日の様子は、広島県知事ブログに掲載されている（https://www.pref.hiroshima.lg.jp/site/chijibulog/301130.html）。

9 例えば、小野（2016）では、広域的自治体の取組が取り上げられている。

10 専門家の間で、両者は相互補完的であることが言われている（田辺2014）。

11 自治体の行政計画策定については、松井（2019）に詳しい。

12 例えば、独立行政法人経済産業研究所のEBPMに関する研究プロジェクトでは、残された課題の1つに行政機関と研究者の連携を挙げる（https://www.rieti.go.jp/jp/projects/program_2016/pg-09/007.html）。

参考文献

青柳恵太郎（2018a）「経済学および国際開発の視点から」第2回エビデンスに基づく実践と政策セミナー（2018年6月30日）発表資料

青柳恵太郎（2018b）「EBPMが継承すべきEBMの思考法」、『行政&情報システム』、54（5）:28-32

青柳恵太郎（2019）「国際開発分野におけるEBPM実践と自治体への示唆」、『かながわ政策研究ジャーナル』、13:86-94

青柳恵太郎・小林庸平（2019）「EBPMの思考法　やってみようランダム化比較試験！ vol.4 ランダム化比較試験の実施工程(1)」、『経済セミナー』、710:75-84

秋吉貴雄・伊藤修一郎・北山俊哉（2010）『公共政策学の基礎』、有斐閣

磯崎初仁・金井利之・伊藤正次（2014）『ホーンブック地方自治』、北樹出版

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（1）:3-16

神奈川県政策研究センター（2018）「【事例集】『調査：「根拠に基づく政策運営（EBPM）」（中間報告）』」、2018年8月 https://www.pref.kanagawa.jp/docs/r5k/cnt/f7282/documents/3ebpmjirei.pdf(2019年9月13日閲覧)

柴田直子・松井望（2012）『地方自治論入門』、ミネルヴァ書房

田辺智子（2014）「業績測定を補完するプログラム評価の役割－米国GPRAMAの事例をもとに－」、『日本評価研究』、14（2）:1-16

田辺智子（2018）「エビデンスと政策の間－評価利用の研究が示すもの－」、第2回 エビデンスに基づく実践と政策セミナー（2018年6月30日）発表資料

中央教育審議会（2018）「2040年に向けた高等教育のグランドデザイン（答申）」、2018年11月26日 http://www.mext.go.jp/component/b_menu/shingi/toushin/__icsFiles/afieldfile/2018/12/20/1411360_1_1_1.pdf(2019年9月13日閲覧)

津富宏（2000）「EBP（エビデンス・ベイスト・プラクティス）への道」、『犯罪と非行』、124:67-99

西畑宏治（2020）「地方自治体におけるEBPM推進上の基礎的事項の研究－背景にある流れ、データ－」、『評価クォータリー』、52:17-30

広島県（2018a）「広島県におけるEBPM実装への挑戦」、広島県EBPMシンポジウム（2018年11月30日）発表資料

広島県（2018b）「広島県EBPM推進ワーキンググループの運営について」、広島県EBPMシンポジウム（2018年11月30日）発表資料

広島県（2019）「平成31年度根拠に基づく政策形成体制構築業務委託仕様書」、2019年3月6日 https://www.pref.hiroshima.lg.jp/uploaded/attachment/342530.pdf(2019年9月13日閲覧)

広島県マネジメント研究会（2016）『しごとの「強化」書－成果志向の行政経営』、ぎょうせい

松井望（2019）「分権改革以降の自治体計画策定」、『都市問題』、110（9）:48-61

三菱UFJリサーチ&コンサルティング自治体経営改革室（2019）「平成30年度 自治体経営改革に関する実態調査報告書」、2019年6月 https://www.murc.jp/wp-content/uploads/2019/06/seiken_190621.pdf(2019年9月13日閲覧)

山本清（2019）「EBPMと自治体企画」、『都市問題』、110（9）:92-99

山谷清志（2016）「地方分権改革と財政危機の自治体評価－20年のレビューから－」、『日本評価研究』、16（1）:31-45

湯﨑英彦（2017）「★成果志向のこれから」、時事通信社iJAMP記事（2017年6月15日）

（2020.3.12 受理）

Current status and issues for Institutionalizing Evidence-Based Policy Making (EBPM) in Japanese Local Governments

Koji Nishihata

Hiroshima Prefectural Government
k.j.nishihata@gmail.com

Abstract

In recent years, some local governments in Japan have begun to introduce Evidence-Based Policy Making (EBPM). It is predicted that more and more local governments in Japan will introduce EBPM in the near future, but the author thinks it is important that rather being diffused than each local government sporadically collecting their own evidence, the system should be rooted into an organization in each local government body. This article focuses on the current status and issues for institutionalizing the EBPM. The author will shed light on some points wherein what behavioral patterns the EBPM would require, and what management of conventional local governments will call for will differ, which is an issue applicable to most Japanese local governments. Then, the author will make clear possible solutions by referring to the case of Hiroshima Prefecture.

Keywords

Local Government, Evidence-Based Policy Making, Institutionalization in Organization,
Hiroshima Prefectural Government

【研究ノート】

教育分野におけるエビデンス仲介機関の特徴

森　俊郎
岐阜県養老町立笠郷小学校
m072132@gmail.com

岡崎　善弘
岡山大学大学院教育学研究科
okazakiys@okayama-u.ac.jp

要　約

エビデンスの活用は教育分野にも広がっており、2000年頃から教育のエビデンスを仲介する機関が諸外国で設立され始めた。本稿は、諸外国のエビデンス仲介機関の特徴について調査した。3つのエビデンス仲介機関（WWC、EEF、CUREE）の各特徴を調査した結果、WWCとEEFは研究を評価する基準を厳密に定めており、WWCはランダム化比較試験を主に集約している一方で、EEFはシステマティックレビューおよびメタアナリシスを主に集約していた。また、エビデンスの頑健性や介入効果を示す枠組みも異なっていた。さらに、WWCとEEFは研究の要約を主に作成しており、CUREEはエビデンスの活用に関する支援に主に取り組んでいた。最後に、教育分野のEBPMの観点からエビデンス仲介機関の在り方について考察した。

キーワード

エビデンス仲介機関、エビデンスに基づく教育、EBPM

1. はじめに

文部科学省はEBPMの考え方を取り入れ始めている（文部科学省 2019）。EBPMとはEvidence-Based Policy Makingの頭文字を取った略称であり、エビデンスを最大限に活用した政策的な意思決定を意味している（森田 2019）。

「エビデンス」はある介入効果を科学的に検証した研究結果を意味しており、エビデンスを活用する際、現時点で得られる研究の中で最も質の高い研究を参照することが求められる。しかし、蓄積され続けている研究の中から質の高いエビデンスを抽出することは容易ではなく、エビデンスの収集から活用に至るまでの間に多くのコストが生じる。そこで、諸外国では、エビデンスの集約や伝達を専門的に担う機関（以下、エビデンス仲介機関）を設立した。

エビデンス仲介機関は質の高い研究を厳選・要約して提供しているため、政策立案者や教育関係者は質の高いエビデンスを容易に参照できる。

日本の教育分野においてもエビデンスの活用を目指すのであれば、同等の機能を有する機関が必要になると思われる。そこで、本稿は教育分野のエビデンスを集約・伝達している諸外国のエビデンス仲介機関に焦点を当てる。そして、エビデンス仲介機関の特徴を示した上で、教育分野のEBPM推進の観点からエビデンス仲介機関の在り方について検討する。

日本評価学会『日本評価研究』第20巻第2号、2020年、pp.77-88

2. 目的

2-1. 教育分野におけるエビデンス

「エビデンスに基づく（Evidence-based）」という考え方は医療分野から始まっており、エビデンスに基づく医療とは、「現時点で得られる最良のエビデンスを参照しながら、医療者の専門性・経験と患者の環境・価値観を踏まえた上で意思決定すること」を意味している（井伊・五十嵐・中村 2019）。

「エビデンスに基づく」という考え方は教育分野にも広がり（Davies 1999; Hargreaves 1997）、日本でも議論が始まっている。例えば、国立教育政策研究所は2010年9月にエビデンスに関する国際シンポジウムを開催しており、2011年には「教育研究におけるエビデンス」というタイトルで紀要の特集も組んでいる。さらに、文部科学省はエビデンスに基づいた教育政策を推進する体制の構築に向けて諸外国の状況を調査しており、エビデンスに基づく教育政策を支えるデータベースの構築やエビデンスの生産（つくる）・伝達（つたえる）・活用（つかう）を主導する人材育成を課題として挙げている（家子・小林・近藤 2017）。

2-2. エビデンスをつたえる・つかう

ある教育的介入の因果関係を科学的かつ厳密に検証した研究は一般的に大学や研究機関が生産しており、膨大な数が国内外で蓄積され続けている。しかし、研究の質は様々であり、研究の収集や質の判断に求められるコストは小さくない。

エビデンスの活用を求めるのであれば、エビデンスを「つかう」政策立案者や教育関係者に対してわかりやすく「つたえる」機関を設置することが必要である。実際に、諸外国では質の高い研究を要約したレポート等を提供する機関が設立されている。

エビデンスを仲介する機関は、わかりやすく「つたえる」ために研究結果を要約したレポートを作成する等、利用しやすい形に作り直すことを主な業務としている。しかし、研究結果を要約・図示する枠組みはエビデンス仲介機関の間で異なっているため、各情報を参照する際には注意を要する。

2-3. 本研究の目的

エビデンス仲介機関を日本に設立する・しないに関わらず、諸外国で設立されているエビデンス仲介機関の特徴や枠組みを知ることは、日本の教育分野のEBPMを推進する上で役立つと思われる。そこで、本稿の第1の目的として、諸外国のエビデンス仲介機関の特徴を整理する。次に、本稿の第2の目的として、各エビデンス機関の特徴を踏まえた上で、教育分野のEBPM推進の観点からエビデンス仲介機関の在り方について検討する。

3. 方法

3-1. エビデンス仲介機関の選定

本稿で扱うエビデンス仲介機関は以下の3つの観点から選択した：(1) 質の高い研究を要約したレポートをウェブサイト上で公開していること、(2) 教育分野に特化した機関であること、(3) 利用者を教育関係者や政策立案者に設定していること。

上記の観点に基づいて各エビデンス仲介機関を調査した結果、アメリカのWhat Works Clearinghouse（WWC）、イギリスのEducation Endowment Foundation（EEF）、イギリスのThe Centre for Use of Research and Evidence in Education（CUREE）が該当した。

3-2. エビデンス仲介機関の調査項目

3つのエビデンス仲介機関を対象として、以下の3つの観点から調査した：(1) エビデンス仲介機関の基礎情報（設立背景・目的・役割）、(2) 提供している情報、(3) 提供する情報を作るプロセス。

上記3点について、各機関が公開している資料とインタビューに基づいて調査した。インタビューは第1著者が各現地に赴いて実施した。

4. 結果1: What Works Clearinghouse

4-1. 基礎情報

4-1-1. 設立背景

アメリカでは2000年頃から教育格差が大きな問題となっている。2002年に制定された「落ちこぼれ防止法（No child left behind act）（以下NCLB法）」は、科学的に実証された施策・指導方法の活用を政策や教育現場に求めた。また、「科学的に実証された教育実践の重視」がNCLB法の基本理念として挙げられており（吉良2009）、教育に関するエビデンスの需要を高めたことが仲介機関の設立と関係している。

アメリカ連邦政府は、エビデンスを教育現場に伝える組織としてWhat Works Clearinghouse（WWC）を教育省の教育科学研究所内に設立した。

4-1-2. WWCの目的と役割

教育効果の科学的情報を提供する中心的かつ信頼される情報源になることをミッションとして掲げている。教師や政策立案者が利用しやすいように、介入研究を要約したレポートや優先事項をまとめたガイドブック等を提供している。

NCLB法の制定等が背景にあるため、学習支援が主な目的として位置づけられている。アメリカ政府は2015年に「すべての生徒が成功する法（Every Student Succeeds Act）（以下ESSA法）」を新たに制定しているが、主な目的はESSA法でも引き継がれている（森 2019）。

4-2. 提供している情報（WWC 2020a）

WWCがレビュー対象としているトピックは12分野（2020年3月時点）であり、（1）介入レポート、（2）実践ガイド、（3）研究レビュー、（4）参考情報を提供している。

4-2-1. 介入レポート

介入研究の概要と効果の要約が記述されている。また、「エビデンスの頑健性」と「改善される程度」がマークを用いて視覚的に示されている。

エビデンスの頑健性は6段階に分類される。例えば、2件以上の研究で一貫してポジティブな介入効果が報告されており、ネガティブな結果を報告した研究が存在しない場合、エビデンスの頑健性は「＋＋」と表示される。一方で、2件以上の研究で一貫してネガティブな結果が報告されている場合、「－－」と表示される。

改善される程度はパーセンタイル順位で示されている。例えば、「15」と表示されている場合は、65番目に位置するスコアから50番目に位置するスコアに上昇することを意味している。

4-2-2. 実践ガイド

どこから何を始めたら良いのか等、学校や教室の課題に取り組む教師を支援するために作成されている。複数の研究結果を集約した推奨事項、想定される課題、課題に対応する際のアドバイスが記載されている。また、実践した教育関係者の意見や専門家の意見も含まれている。

各ガイドの巻末には、エビデンスとして扱った研究論文の詳細（介入方法・比較対照群・エフェクトサイズ等）が記載されている。現在24件の実践ガイドが公開されている。

4-2-3. 研究レビュー

WWCが要約した研究を検索することができる。また、各研究は評価基準（後述するWWCスタンダード）、研究デザイン、トピック、エビデンスの階層（ESSA法）にフィルタリングして検索することもできる。

4-2-4. 参考情報

WWCの概要・仕組み、実践ガイドの活用方法等、WWCに関する総合的な情報が動画等で公開されている。

4-3. 提供する情報を作るプロセス（WWC 2020a）

介入レポートを作る工程は5段階で構成されている：（1）レビュープロトコルの作成、（2）関連研究の収集、（3）スクリーニング、（4）レビュー、（5）レポートの作成。

4-3-1. レビュープロトコルの作成

介入方法、介入対象の年齢、トピックとして収める範囲等、データベースから研究を収集する基準を設定する。

4-3-2. 関連研究の収集

作成したプロトコルに基づいて、データベースを利用して関連する研究を収集する。

4-3-3. スクリーニング

作成したプロトコルの範囲に収まっているかどうか、介入の有効性は信頼できるか等の適合度の確認が行われる。

4-3-4. レビュー

ランダム化比較試験、準実験デザイン、回帰不連続デザイン、単一事例実験デザインがWWCのレビュー対象に含まれる。また、WWCはランダム化比較試験と準実験デザインをグループデザイン試験と総称している。

各研究はWWCスタンダードに基づいて3タイプのいずれかに分類される（WWC 2020b）。

「ランダムに参加者をグループに振り分けている」、「参加者の脱落が少ない」、「交絡している要因やアウトカムに対して懸念がない」という3つの条件が満たされた場合のみ、WWCスタンダードに適合する研究（Meets WWC Design Standards Without Reservations）として分類される。

「ランダムに参加者をグループに振り分けている」という条件を満たしていなくても、「介入する前のグループ間で違いが認められない」、「交絡している要因やアウトカムに対して懸念事項がない」という2つの条件が満たされていれば、WWCスタンダードに条件付きで適合する研究（Meets WWC Design Standards With Reservations）として分類される。

「ランダムに参加者をグループに分けていない」、「介入する前のグループ間で違いが認められる」という2つの条件に該当する場合、WWCスタンダードに適合しない研究（Does not Meet WWC Design Standards）として分類される。

4-3-5. レポートの作成

各研究結果を要約したレポートをWWCのウェブサイト上に公開する。レポートは、エビデンスの頑健性や改善される程度を示す指標と合わせて公開される。

5. 結果2：Education Endowment Foundation

5-1. 基礎情報

5-1-1. 設立背景

イギリスの教育格差は深刻であり、社会経済的に不利な立場に分類された子どもは約190万人と推定されている。また、富裕層の5歳児と貧困層の5歳児の間には約19ヵ月分の学力差が生じているとも言われている（EEF 2016）。

教育格差が拡大していることを解決するために、教育省や民間財団等から多額の設立助成金を受けて慈善事業団体（サットントラストとインペタス）が共同で2011年7月に設置した。助成金を受ける一方で、運営は政府や財団から独立しているため、政治的・資金的制約を受けていない。

5-1-2. EEFの目的と役割

社会経済的に不利な子どもに対する教育支援が主な目的である。親の収入と教育の関係を断ち切り、どのような環境であっても子どもたちが才能を最大限に発揮できるようにすることを目指している（EEF 2016）。

EEFはエビデンスを「つたえる」だけでなく、エビデンスを「つくる」役割も担っている。効果を測定したい教育プログラムの募集を行った上で、選ばれた教育プログラムに対してEEFが研究費を助成している。選ばれたプログラムは原則としてランダム化比較試験で効果を検証する（家子・小林・近藤 2017）。

5-2. 提供している情報（EEF 2018）

EEFはウェブサイト上で（1）Teaching and Learning Toolkit、（2）ガイダンスレポート、（3）

研究助成プロジェクトの情報を公開している。

5-2-1. Teaching and Learning Toolkit

研究を要約したレポートを無償で提供しており、Teaching and Learning Toolkit（以下、ツールキット）を通して現在36件のレポートが紹介されている（EEF 2016）。

ツールキットでは、エビデンスの頑健性、介入の効果、コストを視覚的に示している。

エビデンスの頑健性は、(1) 研究デザインのタイプと報告件数、(2) 測定されたアウトカム、(3) 因果関係の強さ、(4) 研究結果の一貫性、に基づいて5段階で評価される。例えば、(1) 少なくとも5件のメタアナリシスが報告されており、そのうち3件は過去3年以内である、(2) 5件とも標準化された指標を用いている、(3) ランダム化比較試験等の強い因果関係を示す研究デザインで検証されている、(4) メタアナリシスを通して結果が一貫している、という条件を満たす場合、エビデンスの頑健性は最も高く評価される。また、(1) 研究が少なく、観察データで相関関係を調べている、(2) 認知や学習に少なくとも関連している指標である、(3) 因果関係が弱い、(4) 結果が一貫していない、という条件を満たす場合、エビデンスの頑健性は最も低く評価される。

介入の効果は月数（12段階）で示されている。12か月（1年間）分の学習的向上をエフェクトサイズ（0 ～ 1）に置き換えて提示している。例えば、「+1」と表記されている場合、エフェクトサイズが約0.02から0.09（1/12）であることを意味している（表1）。

コストは25人の生徒を対象にした際の費用を示している。年間80ポンドから2000ポンドまでの費用が必要な場合、「£」のマークが1つ付けられる。「£」のマークは最大5つであり、年間30000ポンド以上の費用が必要な場合、「£」が5つ付けられる。

5-2-2. ガイダンスレポート

エビデンスを効率的・効果的に取り入れることができるように、実践的な視点で推奨事項が記述されている。優先度の観点から推奨される事項が記載されており、教師は何をどこから始めると良いのかを知ることができる。現在16件のガイダンスレポートが公開されている。

表1　学習的向上を示す表記（月数）とエフェクトサイズの対応

学習的向上（ヵ月）	エフェクトサイズ 以上	エフェクトサイズ 以下	評価
0	-0.01	0.01	とても小さい
1	0.02	0.09	小さい
2	0.10	0.18	小さい
3	0.19	0.26	中程度
4	0.27	0.35	中程度
5	0.36	0.44	中程度
6	0.45	0.52	高い
7	0.53	0.61	高い
8	0.62	0.69	高い
9	0.70	0.78	とても高い
10	0.79	0.87	とても高い
11	0.88	0.95	とても高い
12	0.96	1.00	とても高い

（出所）Education Endowment Foundation（2018）. Sutton Trust-Education Endowment Foundation Teaching and Learning Toolkitに基づいて作成

5-2-3. 研究助成プロジェクト

EEFが研究を助成したプロジェクトの結果のうち、効果を期待できるプロジェクトの一覧が掲載されている。参加した学校数、生徒数、費用、インパクト（月数）が掲載されている。

5-3. 提供する情報を作るプロセス

ツールキットで紹介する情報を作る工程は4段階で構成されている：(1) 特定・検索、(2) スクリーニング、(3) 適合度の確認、(4) 包括化。

5-3-1. 特定・検索

関心が高いトピックを定めた後、データベースを利用して関連する研究を収集する。網羅的かつ体系的に関連研究を探索することを目的と

しているが、定量的データに基づいた介入研究、システマティックレビューおよびメタアナリシスが対象になる。

レビュー対象になる主な研究論文はPRISMA（Preferred Reporting Items for Systematic Reviews and Meta-analyses）声明に基づいて検索される。PRISMA声明とは、メタアナリシス報告の質を向上させるために、医学系の国際研究グループが提案したガイドラインであり、システマティックレビューおよびメタアナリシスの報告において優先すべき項目が示されている（卓・吉田・大森 2011）。

PRISMA声明は1996年に作成されたQUOROM（Quality of Reporting of Meta-analyses）声明の更新版である。

QUOROM声明は27項目のチェックリストと4段階のフローチャートで構成されており、ランダム化比較試験のメタアナリシスによる報告の質の向上を目的に作成されていた。これらのチェックリストとフローチャートがPRISMA声明で改定されている。

5-3-2. スクリーニング

4つの観点からスクリーニングされる：トピックに関連する研究であるか（トピック）、システマティックレビューおよびメタアナリシスであるか（研究の種類）、標準化されている指標、認知能力検査、テスト成績であるか（アウトカムの種類）、早期（2歳－5歳）と学齢期（5歳－18歳）のデータが含まれているか（サンプル）。

システマティックレビューやメタアナリシスが存在しない場合、個別の研究がレビュー対象として含まれる。

5-3-3. 適合度の確認

最終的なスクリーニングが行われる。観察データに基づいたシステマティックレビューとメタアナリシスは除外される。ただし、実験研究や準実験に基づいたシステマティックレビューやメタアナリシスが存在しない場合、観察データに基づいたシステマティックレビューやメタアナリシスは除外されない。

最終的に残った研究がコード化された上で記録される。コードは16項目：研究名、エフェクトサイズ、FSMエフェクトサイズ、エフェクトサイズの種類、標準誤差、標準偏差、95%信頼区間、エフェクトサイズの範囲、研究数、効果を示した研究数、参加者数、参加者の年齢の範囲、媒介分析（有・無）、出版バイアス（有・無）、メタ分析（該当・該当しない）、研究の質（高い、高くない、不明瞭）。

5-3-4. 包括化

利用可能なデータを用いて、全体的なエフェクトサイズが算出される。

6. 結果3：The Centre for Use of Research and Evidence in Education

6-1. 基礎情報

6-1-1. 設立背景

教師を支援するために1998年に設立されたイギリスの民間会社である（Cordingley 2008）。教育プログラムの評価、エビデンス活用の支援、日本の教員研修に相当する継続的な教育能力の発達支援（Continuing Professional Development: CPD）に取り組んでいる。教師の教育能力の向上はイギリス全体の課題であるため、教育技能省（Department for Education and Skills: DfES）等の教員研修に関連する機関と協力関係を結んでいる。

6-1-2. CUREEの目的と役割

教育に関する有益な情報を提供しても活用するかどうかは教師次第であり、教師が自発的に情報を活用することが教育全体の水準向上につながると考えているため、CUREEはエビデンスの活用を具体的に支援するCPDに主に取り組んでいる。

CUREEのCEOであるCordingley（2015）はエビデンスにアクセスすることが必要な教師ほどエビデンスを利用しない傾向があることを指摘しており、利用したいと思うための支援が必要であると主張している。

6-2. 提供している情報

6-2-1. The Research Informed Practice Site

2006年頃まではエビデンス検索サイト（The Research Informed Practice Site: TRIPS）を公開していたが、現在は閉じている。TRIPSでは各教育方法の効果や具体的な実践方法を簡潔に要約したレポートが掲載されていた。

6-2-2. ルートマップ

CUREEが学校を支援する際、地下鉄の地図のようなルートマップと呼ばれるツールを用いる。研究を要約したレポートや解説動画等がルートマップ上に配置されており、学校の状況や教師のニーズに合わせてカスタマイズされている。レポートはTRIPSで提供していた情報と同一である。現在14種類のルートマップが用意されており、有償で利用できる。

6-3. 提供する情報を作るプロセス

ルートマップ上に配置されるレポート等の作成プロセスに関する情報は公開していなかった。CUREEはEEFやEPPIセンターと協力関係を結んでおり、EEFが公開しているレポートやEPPIセンターが有償で提供しているEPPI Reviewerを利用していたと思われる。

7. 総合考察

諸外国の教育分野のエビデンス仲介機関を概観し、諸外国のエビデンス仲介機関の特徴を調べた。各エビデンス仲介機関の特徴を表2に示した。以下では、日本の教育分野におけるEBPMの観点からエビデンス仲介機関の在り方について考察する。

7-1. 各エビデンス仲介機関の特徴

1つ目は「研究を収集する際の評価と基準」である。WWCスタンダードやPRISMA声明に基づく等、内的妥当性に強く焦点を当てて研究を収集している点はWWCとEEFで共通していた。しかし、評価の基準は異なっていた。WWCはランダム化比較試験や準実験デザインで実施された研究を主なレビューの対象としており、EEFはシステマティックレビューおよびメタアナリシスを主なレビューの対象としていた。どちらも内的妥当性の高さを担保できる研究であるが、研究を収集する際の基準が異なっていることはWWCとEEFを参照する際に留意すべき点である。

2つ目は「要約の視覚化」である。WWCとEEFのどちらも分かりやすさを重視しており、視覚的に分かるようにマークに置き換えて各情報を整理していた。しかし、エビデンスの頑健性を示す指標等、マークの意味が異なる点は注意が必要と言えるだろう。例えば、WWCは結果の一貫性に基づいて段階的にランクを示している一方で、EEFはエフェクトサイズに基づいて段階的にランクを示していた。視覚的な分かりやすさはエビデンスの活用の入り口として必要ではあるが、エビデンスの頑健性を評価する基準が異なっていることはエビデンスを参照する際に留意しなければならないだろう。

3つ目は「レビュー対象のテーマ」である。WWCとEEFが公開しているレポートとガイドブックを表3に示す。どちらもリテラシーに関するレポート・ガイドブックを多く作成しており、教育水準の向上を目的としている様子が伺える。特に、WWCは教科に関するレポート・ガイドブックや脱落を防止するレポート・ガイドブックを多く作成しており、NCLB法やEESA法が影響している様子が伺える。一方で、EEFは家庭環境と学力のつながりを断ち切ることを目的としているため、教科に関するレポートだけでなく、学習環境に関するレポート・ガイドブックを作成している。例えば、保護者と協力して子どもの学習を支援するためのレポート、芸術やスポーツに関するレポート、問題行動や大学進学に向けた支援に関するレポートが挙げられる。

4つ目は「レポートを作成するプロセスの開示の有無」である。WWCはレポートを作成するプロセスをProcedures Handbookで公開しており、EEFはToolkit Manual 2018で公開していた。一方で、CUREEはイギリスのEPPIセンターやEEFと協力関係を結んでいることを公表しているが、ルートマップ内で紹介しているエビデンスに関

表2　各エビデンス仲介機関の特徴

	WWC	EEF	CUREE
国	アメリカ	イギリス	イギリス
所属	教育省	慈善団体	民間企業
目標	学力格差の解消	学力格差の解消 教師の専門性の向上	エビデンス活用の支援 教師の専門性の向上
ミッション	教育効果に関する科学的情報を提供する情報源になる	世帯収入と教育成果のリンクを断つ	研究を利用可能な情報に作り直す 継続的な教育能力の発達支援
利用対象	学校関係者 教育政策立案者	学校関係者 教育政策立案者	学校関係者 教育政策立案者
提供情報	1.　介入レポート 2.　実践ガイド 3.　研究レビュー 4.　参考情報	1.　ツールキット 2.　ガイダンスレポート 3.　研究助成プロジェクト	1.　TRIPS 2.　ルートマップ
提供手段	ポータルサイト	ポータルサイト	ポータルサイト（現在閉鎖）
レポート数	586	34	－
ガイドブック数	24	16	－
適用対象	3歳－18歳	3歳－18歳	－
頑健性	6段階評価	5段階評価	－
改善指標	パーセンタイル順位	エフェクトサイズ	－
情報作成プロセス	1.　レビュープロトコルの作成 2.　関連研究の収集 3.　スクリーニング 4.　レビュー 5.　レポートの作成	1.　特定・検索 2.　スクリーニング 3.　適合度の確認 4.　包括化	－
スクリーニング対象研究	ランダム化比較試験 準実験デザイン 回帰不連続デザイン 単一事例実験デザイン	システマティックレビュー メタアナリシス	－

（出所）Education Endowment Foundation（2018）. Sutton Trust-Education Endowment Foundation Teaching and Learning Toolkit, What Works Clearinghouse（2020a）. What Works Clearinghouse Procedures Handbook, Version 4.1., CUREEのHPに基づいて作成

する指標や基準に関する情報は公表していなかった。

5つ目は「支援のスタイル」である。各エビデンス仲介機関の役割を表4に示した。EEFとWWCは研究を要約したレポートやガイドブックを提供している一方で、CUREEは教員研修を通してエビデンスの活用等に関する具体的な支援に取り組んでいた。EEFのCPDに関するレポートでも報告されているように、研究の結果を理解しても実践に適用することは難しい。エビデンスの適用まで支援するスタイルはCUREEの主な特徴と言える。

7-2. 日本の教育分野のEPBMとエビデンス仲介機関

EBPMの考え方で教育分野の施策を検討する際、最初にエビデンスの有無が調査されるため、要約されたエビデンスを豊富に紹介する機関の必要性は高い。日本にエビデンス仲介機関を設置・運営するためには少なくとも4つの課題を乗り越えなければならないだろう。

1つ目はエビデンス仲介機関を運営する組織である。政府や民間会社がエビデンスを仲介する場合、独立性が課題になるだろう。近年の例として、研究機関と製薬会社の間で治療薬の臨床結果を捏造していた事件が挙げられる。教育関

表3　WWCとEEFが公開しているレポート・ガイドブック

WWC		EEF	
レポート	件数	レポート	件数
義務教育（幼稚園から高校まで）	492	学校の構造（時間割・クラスサイズ・デジタルテクノロジー）	11
リテラシー（読み書きのスキル）	352	社会情動的スキルの育成	8
算数・数学	220	ポスト16（学習到達度が低い16歳以上の生徒の支援）	7
早期教育	85	特別支援教育	5
問題行動の抑制	56	生徒が自発的な学習者になるための支援	4
卒業までの道のり（途中退学防止）	47	学校体験を豊かにする（キャリア教育・芸術・スポーツ）	4
特別支援教育	36	算数・数学	4
英語学習者（英語が母国語ではない生徒）	32	行動（参加の促進・破壊行動の抑制）	3
サイエンス	12	言語とリテラシー	3
高等教育（大学進学の支援）	14	サイエンス	3
チャータースクール（民間運営の公立学校）	3	継続的な教育能力の発達支援	2
優れた教師	7	早期教育	1
		フィードバックと子どもの成長のモニタリング	1
		保護者の関与	1
ガイドブック	件数	ガイドブック	件数
義務教育（幼稚園から高校まで）	0	リテラシー	4
リテラシー（読み書きのスキル）	8	算数・数学	1
算数・数学	6	サイエンス	1
早期教育	1	社会情動的スキル	1
問題行動の抑制	1	メタ認知と自己調整学習	1
卒業までの道のり（途中退学防止）	3	問題行動の改善	1
特別支援教育	2	保護者と協力して子どもの学習を支援する	1
英語学習者（英語が母国語ではない生徒）	2		
サイエンス	1		
高等教育（大学進学の支援）	2		
チャータースクール（民間運営の公立学校）	0		
優れた教師	3		

（出所）Education Endowment FoundationとWhat Works Clearinghouseのサイトに基づいて作成

表4　エビデンスに関係する各機関の役割と取り組み

役割	つくる	つたえる		つかう
機関	大学　研究所	WWC　EEF	CUREE	教育行政　学校
取り組み	データベースの構築 実証研究（RCT等）	研究の要約 （研究の要約作成）	教員研修 （エビデンス活用支援）	エビデンスの活用 意思決定

（出所）筆者作成

連会社や他の団体の利益につながるエビデンスが捏造される可能性を含んでいることから、独立性の課題の克服はエビデンス仲介機関を設立する上で極めて重要と思われる。独立性の課題を乗り越えるためには、EEFのように、政治的背景や民間会社の利益から切り離された運営が最も望ましいだろう。

2つ目は評価基準の作成である。各エビデンス仲介機関は観察データに基づいた相関関係や単に収集されたデータはエビデンスとして扱っていない。また、経験に基づいた授業のテクニックや科学的に検証されていないベストプラクテ

ィスも対象外だった。EBPMの考え方が広がる一方でエビデンスに対する誤った理解も散見されるようになっているため（原田 2015）、誤解が生じないように、また、内的妥当性の高さを担保するためにもWWCスタンダードやPRISMA声明等を参照しながら各研究の評価基準を作成する必要があると思われる。

3つ目は人材育成である。人材育成は教育政策分野と学校教育分野に分けて議論する必要があるだろう。

教育政策分野の場合、教育政策の効果を検証する各自治体が今後増加すると思われるため、効果検証のデザインや統計学に関する知識等、エビデンスを「つくる」ことに重点を置いた人材育成が必要と思われる。一方で、学校教育分野の場合、「つくる」よりも「つかう」ことに重点を置いた人材育成が必要である。例えば、エビデンス仲介機関が作成したガイドブックを読んだ上で、推奨されている事項を学校や教室でどのように適用するかを検討することができなければならない。

どのような課題に対してどのような方法が選択できるのか等、ニーズに合わせてエビデンスを紹介するだけでなく、CUREEの取り組みのように、エビデンスの適用の仕方を支援する人材も必要と思われる。また、利用できるエビデンスが存在しない場合、エビデンスを作る（評価方法を設計した上で政策の効果を調べる）ことが検討される。その際も担当者は研究デザインと因果関係の強さの対応関係を理解しておかなければならない。

4つ目はエビデンスを利用する目的である。WWCやEEFは限られた財政の中で確実な教育効果を得るためにエビデンスの活用を試みている。WWCやEEFはより良い意思決定を支援するための道具の1つとしてエビデンスを提供している点に留意したい。日本では、経験に基づいた教育に対する批判としてエビデンスを持ち出している傾向が強い。経験だけに基づいて誤った介入が実施されることは避けられるべきだが、学校教育に対する批判の道具としてエビデンスを持ち出すのではなく、子どもたちの学力向上やウェルビーイングの向上に役立てる道具の1つとして認知されるように努めることもエビデンス仲介機関の義務と言える。

謝辞

本論文について有益なコメントを下さった京都産業大学の惣脇宏先生に深く感謝いたします。

参考文献

家子直幸・小林庸平・近藤碧（2017）「諸外国における客観的根拠に基づく教育政策の推進に関する状況調査報告書」、文部科学省

井伊雅子・五十嵐中・中村良太（2019）『新医療経済学：医療の費用と効果を考える』、日本評論社

吉良直（2009）「どの子も置き去りにしない（NCLB）法に関する研究：米国連邦教育法の制定背景と特殊性に着目して」、『教育総合研究：日本教育大学院大学紀要』、2：55-71

卓興鋼・吉田佳督・大森豊緑（2011）「エビデンスに基づく医療（EBM）の実践ガイドラインシステマティックレビューおよびメタアナリシスのための優先的報告項目（PRISMA声明）」、『情報管理』、54：254-266

原田隆之（2015）『心理職のためのエビデンス・ベイスト・プラクティス入門－エビデンスを「まなぶ」「つくる」「つかう」』、金剛出版

文部科学省（2019）「教育政策におけるEBPMの強化（令和 元 年12月3日 ）」、https://www5.cao.go.jp/keizai-shimon/kaigi/special/reform/wg7/20191203/shiryou1_1.pdf（参照日2020/03/06）

森俊郎（2019）「EBE実践で語ろう」、『「エビデンスに基づく教育」の閾を探る－教育学における規範と事実をめぐって』、春風社

森田正信（2019）「教育行政における EBPM の取組状況について」、『教育行財政研究』、46：25-30

Cordingley, P.（2008）. *Sauce for the Goose: learning entitlements that work for teachers as well as for their pupils.* London.

Cordingley, P.（2015）. *A World Class Teaching Profession*. London : Response to the DfE consultation.

Davies, P.（1999）. What is evidence-based education?. *British Journal of Educational Studies*, 47（2）, 108-121.

Education Endowment Foundation（2016）. *The EEFat 5*. London: Education Endowment Foundation. https://educationendowmentfoundation.org.uk/about/annual-reports/the-eef-at-5/（参照日2020/03/06）

Education Endowment Foundation（2018）. *Sutton Trust-Education Endowment Foundation Teaching and Learning Toolkit*. London: Education Endowment Foundation. https://educationendowmentfoundation.org.uk/resources/teaching-learning-toolkit（参照日2020/03/06）

Hargreaves, D. H.（1997）. In defence of research for evidence-based teaching: a rejoinder to Martyn Hammersley. *British Educational Research Journal*, 23（4）, 405-419.

What Works Clearinghouse（2020a）. *What Works Clearinghouse Procedures Handbook, Version 4.1.* Washington, DC.

What Works Clearinghouse（2020b）. *What Works Clearinghouse Standards Handbook, Version 4.1.* Washington, DC.

（2020.4.28 受理）

The characteristics of evidence institutions in education

Toshiro Mori
Kasago elementary school
m072132@gmail.com

Yoshihiro Okazaki
Okayama University
okazakiys@okayama-u.ac.jp

Abstract

The Evidence-informed decision making has spread to the education field, and the evidence institution that mediate the evidence in the education field have been established in foreign countries since about 2000. This paper investigated the features of evidence mediators in foreign countries. As a result of investigating each feature of three evidence intermediation organizations（WWC, EEF, CUREE）, WWC and the EEF rigorously made the standard which evaluates the research, and while the WWC mainly collects randomization comparative research, the EEF mainly collects systematic review and meta-analysis. Then, the frame which showed robustness of evidence and intervention effect was also different. In addition, the WWC and the EEF focused on the preparation of the report and the guidebook which summarized the research and the CUREE have supported on the utilization of evidence through teacher training. Finally, it was discussed that evidence institution from the viewpoint of EBPM in the education field.

Keywords

Evidence institutions, Evidence-based education, Evidence-based policy making

【研究論文】

国際開発分野におけるEBPMの現状

佐々木　亮
国際開発センター
sasaki.ryo@idcj.or.jp

要　約

国際開発分野における「エビデンスに基づく政策立案」（EBPM: Evidence-Based Policy Making）の実践として、インパクト評価（Impact Evaluation）が広く普及している。そこで使われているデザインは最も厳格なデザインであるランダム化比較試験（RCT）である。本稿ではインパクト評価を推進して、評価結果のデータベースを構築して公開している新興の3つの非営利の研究機関について比較して議論する。それらは、貧困アクションラボ（J-PAL：The Abdul Latif Jameel Poverty Action Lab）、IPA（Innovations for Poverty Action）、3ie （International Initiative for Impact Evaluation）である。さらに、伝統的な援助機関（世界銀行、アメリカ国際開発庁（USAID）、イギリス国際開発局（DFID）、日本の国際協力機構（JICA））におけるインパクト評価の現状について概観する。最後に今後の課題について記す。それらは次の通り。(1)50年の時を経て実現に向かう「実験する社会」と今後の課題、(2)「プロジェクト・アイランド」とインパクト評価の対象レベルに関する生来的な問題、(3)業績測定（パフォーマンス・メジャーメント）と評価の混同に起因する問題、(4)NGOsとの協働による介入実施とデータ収集の可能性、(5)IT業界のバックアップによるEBPMの普及と今後の課題である。

キーワード

インパクト評価、RCT、貧困アクションラボ、3ie、IPA

1．背景と既存文書の分析

国際開発分野において「エビデンスに基づく政策立案」（EBPM: Evidence-Based Policy Making）が世界的に大流行している。Shayda, M.S. and Brown, A.N.（2018）によると、インパクト評価の件数は、2000年代に入って急速に増加しており、近年は年間400-500件の規模で推移している（図1）。ただし、このグラフに含まれるインパクト評価のデザインは、ランダム化比較試験（RCT：Randomized Controlled Trial）だけではなく、マッチングデザインなどの準実験デザインも含まれることに留意して見る必要がある。このインパクト評価の増加の趨勢は、この論文で解説するアメリカの3つの研究機関の設立に端を発して世界的に拡大している。この動向をWhite H. & Masset, E（2018）は4つの波として解説しているが、筆者の専門の評価学から見ると、そのはるか以前にもう1つの波があったと言えるのでその波を加えて表にまとめた（表1）。

まず第0の波と言えるのが、1960-1970年代の波で、1969年にキャンベルが書いた論文で、実験

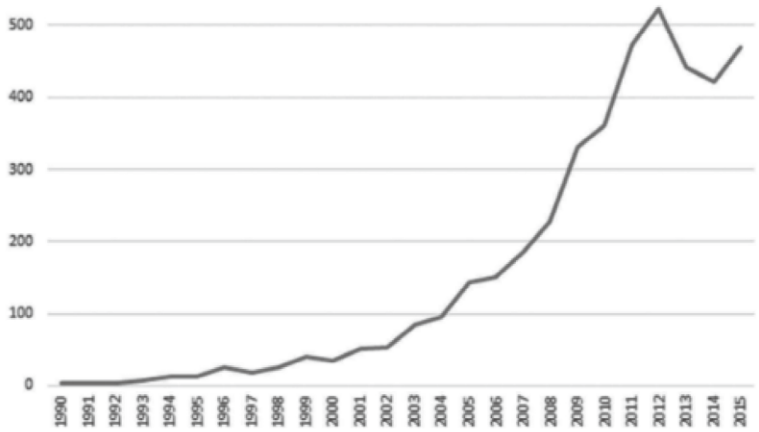
デザイン（Experimental Design）、現在で言うところのRCTを社会科学分野で適用しようという主張がなされた。次の通りである。「アメリカ合衆国とその他の現代国家は、社会の改善に向けて、実験的アプローチを用いる用意ができなければならない。ここで言う実験的アプローチとは、特定の社会問題を解決するためにデザインされた新しい施策を実施する際に用いるアプローチであり、このアプローチによって、不完全ながら複数の基準に照らしたときに明らかな効果があったのかどうかを確認し、その確認の結果に基づいて、施策を維持するか、改善するか、あるいは中止するかを決定することになる。」（Campbell, D.T.1969, p.409）。その後、1979年に、現在まで続くRossi, Lipsay, Freeman のEvaluation: A Systematic Approachの初版が発売されたが、そのはしがきによると、著者たちが個別にUNESCOの国際会議に招待されたことが書かれている。その会議に出席していたOECDの代表者の勧めで、その国際会議で出会った3人の著者が発表原稿を国際開発分野向けに編纂したのがEvaluating Social Projects in Developing Countries（1979）で、アメリカ国内向けに出版したのがEvaluation A Systematic Approach 1st editionということになる。その初版では、援助案件ではないが、Experimental Designの適用の実例としてメキシコにおける子供向け番組Sesami Streetの事例が掲載されている。その本でも解説されているとおり、キャンベルの論文発表後に、RCTに代表される科学的評価対実践的評価の論争があった。ただし、この第0の波は激しい論争を引き起こしただけで、論争の結果は、RCTに否定的な実践的アプローチおよび定性的アプローチの派閥が優勢のまま、その後10年以上の時が過ぎたことになる。

その後の第1の波は、1990年代のNew Public Management（NPM）の影響の下で起こった。具体的には、1992年のEffective implementation : key to development impact（いわゆるWapenhans report）において、それぞれのプロジェクトはアウトカム指標（Outcome Indicators）を設定すべきだという主張がなされた。その頃、アメリカではGovernment Performance and Results Acts（GPRA:1993）が成立して、全ての省庁は、民間企業に倣って、Strategic Planを作成してPerformance Measurementを実施してその結果を議会に報告するように義務付けられた。Strategic Planはもちろん、定量的な指標と数値目標を設定することを含む。国際開発の分野でも同じ発想に基づくMillennium Development Goals（MDGs）が国連主導で合意された。しかし、この時代にアメリカ会計検査院（GAO）がアメリカ国際開発庁（USAID）に対して指摘したように、「USAID以外の多くの要因が幅広く影響しているので、アウトカム指標は、ある特定の機関（USAID）の努力のみを測定することはできない」とその限界が認識された（White, H & Masset, E. 2018）。ただし、定量的にインパクトを測定することの重要性は広く認識されるようになった時期と言える。なお、1970年代から1990年代に国際開発分野においてRCTを適用した数少ない事例（7例）を解説した先駆的な論文がこの時期に発表されている（（Newman, Rawlings & Gertler 1994）。

第2の波は、2003年の貧困アクションラボ（J-PAL: Poverty Action Lab.）の設立と、Innovations for Poverty Action（IPA）の設立から始まる。貧困アクションラボは、インパクト評価の手法としてRCTしか使用しないと宣言しているし、IPAも同様にRCTを使用してインパクト評価を実施している。この実施の背景としては、Abt AssociatesのLarry Orr（1998）が指摘しているように、当時主流であった回帰分析をはじめとする統計分析（本人は“structural models”と呼称）では、客観的に介入のインパクトを特定できないということが広く経験されるようになったことが挙げられると思われる。貧困アクションラボとIPAは、驚嘆するしかない勢いでRCTを適用したインパクト評価を実施し続けて、設立からわずか数年で数百に及ぶインパクト評価を実施した。この時期に一気にインパクト評価の累積件数が増加したわけである（図1参照）。

第3の波は、RCTを適用したインパクト評価の継続的な増加と、こうした状況を踏まえて開始されたメタアナリシス（Meta-Analysis、別名Systematic Review）などのセクター別レビューの増加である。メタアナリシスの手法自体はすでに1990年代以前に確立されていたと言える。メ

図1　インパクト評価の件数の推移



（出所）Shayda M.S. & Brown, A.N.（2018）から転載

タアナリシスは、複数のインパクト評価（そのデザインは主にRCT）を統合して、より客観的なインパクトを明らかにする試みである。2006年に非営利の研究団体であるCentre for Global Developmentが発表した「When Will We Ever Learn?」という報告書で、エビデンスを集積する新しい機関の設立が提案された。その提案を受けて2008年に設立されたInternational Initiative for Impact Evaluation（3ie）が最初のメタアナリシス（Systematic Review）の報告書を公表したのが2008年である。それ以来、3ieはメタアナリシスの報告書を産出し続けている。3ieのSystematic Reviewのサイト（https://www.3ieimpact.org/evidence-hub/publications/systematic-reviews）で、「Systematic Reviews」を選択すると、44件の報告書が表示される（2019年11月時点）。また厳格なメタアナリシスより幅広い「レビュー」と「プロトコール」の合計では600件以上が登録されている。ここで言うプロトコールとは、レビューの理論的根拠、仮説、および計画された方法を説明したもので、レビューを開始する前に準備して公開されるものである。また、貧困アクシ

表1　国際開発におけるエビデンスの波（第0～第4の波）

波の段階	年代	説明
第 0 の波	1970-1980 年代	RCT の国際開発分野への紹介と、科学的評価対実践的評価の論争。
第 1 の波	1990 年代	New Public Management (NPM)の影響によるアウトカム重視の考え方の定着。
第 2 の波	2000 年代	RCT を適用したインパクト評価の実施機関の設立とエビデンスの蓄積の時代。貧困アクションラボ（J-PAL : Poverty Action Lab.）と Innovations for Poverty Action（IPA）の設立。
第 3 の波	2000 年代末～2010 年代中期	3ie の設立によるメタアナリシス(Meta-Analysis)とセクター別レビューの増加の時代。RCT を適用したインパクト評価の継続的増加。
第 4 の波	2010 年代後期	「エビデンスの洪水」と「エビデンスから政策への反映」の試行錯誤の時代。

（出所）White, H. & Masset, E.（2018）. 第0の波を筆者が追加したほか、一部区分を変更している。

ョンラボやIPAでも、厳格なメタアナリシスの手法を用いないまでも、セクター別レビューを産出するようになった時期である。

第4の波が、2010年代後半に入ってからの「エビデンスの洪水」（Flood of Evidences）とも呼べるような状況である。そしてそれに対応した「知識の伝達」（Knowledge translation）あるいは「エビデンスから政策へ」（Translating evidence to inform policy）の動きである。RCTを適用したインパクト評価およびそのメタアナリシスによってエビデンスが蓄積されてきたが、それが予想したとおりには政策に反映されていないという批判に対する動きである（浅岡2016、青柳2017）。つまりこの「エビデンスの洪水」という状況は、1970-1980年代の評価学での論争を起点として見ると、信じ難くそして歓喜に値する状況であるが、一方ですぐに問題点として語られるようになったわけである。これに対して各機関が専用のウェブサイトを用意して、産出したエビデンスが政策に活かされるように努力を続けているのが、現在の状況である。貧困アクションラボは「Evidence to Policy」、IPAは「Policy Engagement」といった専用のサイトを公表してその努力を続けている。また、3ieの「Evidence Gap Map（EGM）」は、本来はインパクト評価が足りない分野においてその実施を促すことが目的であると理解されるが、政策立案者がエビデンスの蓄積具合を参照するためにも活用できると思われる。

2. 新興の非営利の研究調査機関の現状（貧困アクションラボ(J-PAL)、IPA、3ie）

本節では、これらの動きをリードする3つの非営利組織である貧困アクションラボ（J-PAL：The Abdul Latif Jameel Poverty Action Lab）、3ie（International Initiative for Impact Evaluation）、IPA（Innovations for Poverty Action）について、その起源と最新の動向を概観する（表2、表3を参照）。

表2　新興の非営利の研究機関（貧困アクションラボ（J-PAL）、IPA、3ie）の基本情報1

略称	貧困アクションラボ（J-PAL）	IPA	3ie
正式名称	The Abdul Latif Jameel Poverty Action Lab	Innovations for Poverty Action	International Initiative for Impact Evaluation
創設年	2003	2002	2008
創設者	Abhijit Banerjee, Esther Duflo, & Sendhil Mullainathan https://www.povertyactionlab.org/node/12027	Dean Karlan https://www.poverty-action.org/about/history	Howard White https://www.cgdev.org/blog/3ie-enters-new-phase-and-thank-you-howard-white
所在地	本部：MIT, Cambridge, MA. 地域オフィス: Africa, Europe, Latin America & the Caribbean South Asia and Southeast Asia	本部：New Haven, CT: NY 支部：Washington D.C. East Africa (6), West Africa (6), Asia (3), Latin America (3)	本部：New Delhi, 支部：London and Washington, D.C.
紹介文	The Abdul Latif Jameel Poverty Action Lab (J-PAL) is a global research center working to reduce poverty by ensuring that policy is informed by scientific evidence. Anchored by a network of 171 affiliated professors at universities around the world, J-PAL conducts randomized impact evaluations to answer critical questions in the fight against poverty	Innovations for Poverty Action (IPA) is a research and policy nonprofit that discovers and promotes effective solutions to global poverty problems. IPA brings together researchers and decision-makers to design, rigorously evaluate, and refine these solutions and their applications, ensuring that the evidence created is used to improve the lives of the world's poor.	The International Initiative for Impact Evaluation (3ie) is a global leader in funding, producing, quality assuring and synthesising rigorous evidence. We support studies and reviews that examine what works, for whom, why and at what cost in low- and middle-income countries (L&MICs). We are also a global advocate for the generation and use of quality evidence in development decision-making.
パートナー（資金支援者など）	<Partners> - Community Jameel, - The Laura and John Arnold Foundation, - The Bill and Melinda Gates Foundation, - The William and Flora Hewlett Foundation, - The John D. and Catherine T. MacArthur Foundation, - The Alfred P. Sloan Foundation, - The Australian Dept. of Foreign Affairs and Trade - The UK's DFID	<Funders> 110 Funders and 150 donors. - Bill & Melinda Gates Foundation - Inter-American Development Bank (IDB) - UK's DFID, USAID, UN agencies, WB - University of California, each school - J-PAL, 3ie	The three main funders of 3ie are: - The Bill & Melinda Gates Foundation, - UKaid through the DFID - The William and Flora Hewlett Foundation
インパクト評価報告書の数と種類	**952 randomized evaluations** conducted by our affiliates in 81 countries as of 15 April 2019. Agriculture (82), Crime, Violence & Conflict (52), Education (234), Environment & Energy (41), Finance (277), Firms (22), Gender (145), Health (191), Labor Markets (142) and Political Economy & Governance (211).	**831 evaluations** in 51 countries as of 15 April 2019. IPA is the largest implementer of randomized evaluations in the development field. We have run over 830 studies in 51 countries in partnership with more than 600 researchers around the world.	**4,727 records** as of 15 April 2019. Only 283 were funded by 3ie. <Evaluation Approach> RCT (1974), RDD (90), PSM (499), IV (244), DID (506), Mixed Method (80). <The top 10 countries> India (390), China (281), Mexico (247), Kenya (233), Bangladesh (197), South Africa (194), Brazil (193), Uganda (173), Pakistan (105) and Peru (105).
メタ分析（ｼｽﾃﾏﾃｨｯｸﾚﾋﾞｭｰ）の数と種類	メタ分析は実施していない。かわりに、Policy Insight というサイトを設けて、複数のインパクト評価から得られたセクター別の Lessons Learned を掲載している。	メタ分析は実施していない。かわりに、Key Findings としてセクター別の主な Findings を掲載している。セクターは以下のとおり。 農業、教育、財務的包摂、ガバナンス、保健、平和・復興、SME、社会保護。	(1) メタ分析を実施している。厳格なメタアナリシスより幅広い「レビュー」と「プロトコール」の合計では 600 件以上が登録されている。 (2) Evidence Gap Maps (EGM)をセクター別に作成して、得られた教訓を公表している。

（出所）貧困アクションラボ（J-PAL）、IPA、3ieのウェブサイト。2019年4月時点の情報。

表3　新興の非営利の研究機関（貧困アクションラボ（J-PAL）、IPA、3ie）の基本情報2

略称	貧困アクションラボ（J-PAL）	IPA	3ie
データの公開	総件数のうち、198 件の評価でデータを入手可能。	一部の評価で、ハーバード大学のデータベースで公開している Randomized Controlled Trials in the Social Sciences Dataverse	総件数のうち、32 件の評価でデータを入手可能。
インパクト評価を実施するための Grant 提供	事業費を一部支援している。J-PAL に登録している研究者が自らインパクト評価を実施している Selected partners receive pro bono technical support from J-PAL North America staff, flexible project funding, and connections with J-PAL's network of leading academic researchers.	Competitive Funds として Proposal を募っている。現在 3 回目の募集中。 "Intimate Partner Violence Initiative: The IPV Initiative's third call for proposals is now open. Proposals are due May 17, 2019."	3ie 自らはインパクト評価を実施しない。公募して Grant を提供している。その総数は、304 grants (243 impact evaluations, 38 systematic reviews and 23 other studies) as of 15 April 2019.
政策への反映	Evidence to Policy というサイトを設けてスケールアップ（Scaling up）の事例を紹介している。 - Adapting and scaling a program - Scaling up an evaluated pilot - Scaling back an evaluated program - Shifting global thinking - Institutionalizing evidence use - Applying research insights	世界中の 19 の政府と Policy Engagement を締結している。また、28 の政府と何らかの関係を構築している。 IPA exists to solve these problems, building a world with more evidence and less poverty. Three pillars are (i) Create Stronger Evidence; (ii)Share Evidence Strategically: and (iii) Equip Decision-Makers to Use Evidence	以下の報告書を公表している。 - 30 Replication studies - 7 Unable to replicate studies
能力開発（キャパビル）	- Research Resources として各種のマニュアルを公開している。 - Evaluating Social Programs というエグゼクティブ向けのコースを世界各地で開催している。また、同名の Online Course も edX で運営している。	特段のセミナーやワークショップは実施していない。しかし、以下の Tool kit を公表している。 - Theory of Change and Program Design Resources - Resources and Tools for Impact Evaluation	セミナーやワークショップを開催している。 - Systematic review and meta-analysis seminar - Residential workshop on advances in social and behavior science research -Seminars
ニュースレターなど	Email による Newsletter を発行している。 また、Blog を掲載している。	Email による Newsletter は発行していないようである。ただし、News を掲載している。	Email による Newsletter を発行している。また、Blog を掲載している。
今後 5 年間の計画	公表していない。	IPA Strategic Plan 2013-2018	3ie strategy 2020
SDGs への言及	サイトの検索窓で検索すると 2 件のみヒットした。 - "SDGs: The people's agenda?" (2018). - "International Day of the Girl Child: How do we close the gender gap in school participation?" (2017)	サイトの検索窓で検索すると 2 件のみヒットした。 - The Importance of Management Support for Teacher-Led Targeted Instruction in Ghana. - Health, Water, and Sanitation Conference.	サイトの検索窓で検索すると 8 件のみヒットした。主なものは以下のとおり。 - Masset, E. & Jimenez, E., Executive Director, 3ie. (2016) "How synthesised evidence can help with meeting the Sustainable Development Goals" -White, H. CEO, Campbell Collaboration. (2015) "Toward evidence-informed policies for achieving the Sustainable Development Goals".

（出所）貧困アクションラボ（J-PAL）、IPA、3ieのウェブサイト。2019年4月時点の情報。

2-1. 新興の非営利の研究研究機関の概要とインパクト評価の概要

(1)　貧困アクションラボ（J-PAL）

貧困アクションラボ（J-PAL）は、2003年にマサチューセッツ工科大学（MIT: Massachusetts Institute of Technology）の経済学部の教授であるAbhijit Banerjee, Esther Duflo, Sendhil Mullainathanの3人のリードにより設立された非営利の研究センターである（J-PAL 2018）。当初は、ハーバード大学の経済学部の教授陣も名を連ねていたと記憶しているが、現在はこの3人がリードしたと公式に記載されている。センターの名前にあるAbdul Latif Jameelとは、MIT経済学部の卒業生のサウジアラビア人の名前で、貧困アクションラボは、同氏の寄付金を基に設立された。現在では、多数のパートナー（資金支援者）から支援を受けており、主な資金支援者にThe Bill and Melinda Gates Foundation、The William and Flora Hewlett Foundation、MacArthur Foundation、Sloan Foundation（MITのビジネススクールにもその名を冠している）などがおり、さらに、イギリスの国際開発局（DFID）が名を連ねている。

2019年4月時点で952件のRCTを適用したインパクト評価を実施している。81カ国の研究者（Affiliatesと呼称）がインパクト評価の報告書を執筆しているが、実際の介入は国際NGOsが実施する事業であることが多い。内訳は、農業（82）、犯罪・暴力・紛争（52）、教育（234）、環境・エネルギー（41）、ファイナンス（277）、企業（22）、ジェンダー（145）、保健（191）、労働市場（142）、政治経済・ガバナンス（211）となっている。一方、メタアナリシスは実施していない。かわりにPolicy Insightというサイトを設けて、複数のインパクト評価から得られたセクター別のLessons Learnedを掲載している。

また、図2に、マイクロファイナンスに関するPolicy Insightに掲載されている貧困アクションラボによるセクター別レビューの実例を掲載する。この実例から分かることは、貧困アクションラボは、複数のインパクト評価から得られたイン

図2　Policy Insightの実例（マイクロファイナンス）

Outcome	Bosnia and Herzegovina	Ethiopia	India	Mexico	Mongolia	Morocco	Philippines
Business ownership	↑	–	–	–	↑	–	–
Business revenue	–	–	–	↑	–	↑	–
Business inventory/assets	↑	no data	↑	no data	↑	↑	–
Business investment/costs	–	–	↑	↑	no data	↑	↓
Business profit	–	–	–	–	–	↑	–
Household income	–	–	–	–	–	–	–
Household spending/consumption	–	↓	–	↓	↑	–	–
Social well-being	–	–	–	↑	–	–	↓

Note: Green (red) arrows represent statistically significant positive (negative) differences in outcomes between the treatment and comparison groups at the 90 percent confidence level or higher, dashes represent no statistically significant difference. See below for more details.**

（出所）貧困アクションラボ（J-PAL）

パクトの幅を数学的に統合するメタアナリシスよりもこうしたレビューの手法をより信じていると推察されるということである。

なお、952件のうち、データを公表しているのは198件（21%）である。インパクト評価の適用は、貧困アクションラボの研究者（Affiliates）が国際NGOsと協力して実施するパターンが多く、NGOsの事業費の一部を支援する場合があるとしており、これがNGOsの協力のインセンティブになっていると理解される。

⑵　Innovations for Poverty Action（IPA）

IPAは、2002年にMIT経済学大学院の卒業生のDean Karlanによって設立された。彼のアドバイザーはAbhijit BanerjeeとEsther Dufloであったし、両氏はIPAの理事会にも名前を連ねている。したがって、貧困アクションラボとIPAは協力関係にあり、共同でインパクト評価を実施するケースも見られる。ただし、なぜ2つの組織に分かれて活動しているのかはIPAのウェブサイトにも記載がないので理由は不明であり、競争関係にあるとも言える。なお、同ウェブサイトでは、「IPAは、RCTを適用したインパクト評価の最大規模の実施者（Largest implementer）である」という記載があるが、インパクト評価の数では貧困アクションラボには及ばない。約110のFundersと約150のDonorsから資金支援を受けており、主な資金支援者にThe Bill and Melinda Gates FoundationなどのFoundationsの他、Inter-American Development Bank（IDB）、イギリスの国際開発局（DFID）、アメリカ国際開発庁（USAID）、そして国連機関や世界銀行などが並んでいる。貧困アクションラボが国際NGOsとの協力関係が深いのに比べ、IPAはこうしたドナー政府・ドナー機関との関係が深いようである。

IPAは、2019年4月の時点で51カ国において831件のRCTを用いたインパクト評価を実施している。メタアナリシスは実施していないが、Key Findingsとして、セクター別の主なFindingsを掲載している。掲載しているのは、農業、教育、ファイナンス、平和・復興、SME支援、社会包摂の各分野である。一部のインパクト評価でデータを公表している（ハーバード大学のサイトを利用している）が、件数は集計されていないので、総数に対する比率は計算できなかった。また、Competitive FundsとしてRCTを用いたインパクト評価の実施を資金支援している。

⑶ International Initiative for Impact Evaluation（3ie）

国際開発分野における学会誌であるJournal of Development Effectiveness（JDE）（無料で公開されている）を発行しているCenter for Global Development（CGD）という非営利の組織がある。設立は2001年で、最大の資金支援者は、The Bill and Melinda Gates FoundationとThe William and Flora Hewlett Foundationである。この組織は、研究者が集まっていくつかのワーキンググループに分かれて研究を行っており、そのひとつにEvaluation Gap Working Groupがある。同ワーキンググループには、貧困アクションラボの設立者の1人であるEsther Dufloがメンバーとして参加している。同ワーキンググループの目的は、「国際開発分野では、なぜ開発資金の総額に比べて極めてわずかの評価しか行われないのか」、つまり評価ギャップ（Evaluation Gap）の原因を議論して明らかにすることである。同ワーキンググループが発表した中で影響力の大きかった論文が「When Will We Ever Learn?: Improving Lives Through Impact Evaluation」（CGD 2006）である。その中で「The Evaluation Gap Working Groupは、新しいentity－それは委員会、ネットワーク、事務局、あるいは組織－が必要であるという合意に達した」と述べた。新しい組織の価値は、（貧困アクションラボをはじめとする）個別の機関が産出するインパクト評価の情報を整理してデータベース化するとともに、複数のインパクト評価をレビューして、国際開発に有用な知見を得て普及させることであった。具体的な機能として、①厳格な評価の基準を確立する、②評価デザインと評価結果のレビューを行う、③優先されるべきトピックを特定する、④インパクト評価のための資金支援（Grant）を提供する、⑤情報を整理して広報する、⑥知識を産出して解釈して活用する能力を構築する、などとした。また、同報告書に記載された重要な問いかけとして、「10年後に、我々は、極度の貧困と戦うために何が機能するかの知識が不足しているとして未だに嘆いているか、あるいは、社会開発の効果についてエビデンスを有して開発のための資源を有効に活用しているか、どちらだろうか」という問いを設定した。その報告書から10年以上が経過したが、現状はどちらであろうか。

この提言（CGD 2006）を受けて、2年後の2008年にInternational Initiative for Impact Evaluation（3ie）が設立された。本部はインドのニューデリーで、支部がロンドンとワシントンD.C.に所在する。3ieのサイトによると、主要な資金支援者のトップ3は、The Bill and Melinda Gates Foundation、イギリスの国際開発局（DFID）、The William and Flora Hewlett Foundationである。初代の事務局長（Executive Director）は、Campbell CollaborationのCEOであったHoward Whiteが就任した。

3ieでは、2019年4月現在で4,727件のインパクト評価の情報をデータベース化している。もちろん貧困アクションラボとIPAによるRCTを適用したインパクト評価の結果も含まれている。しかし3ieの特徴は、RCTに限らずいろいろなデザインのインパクト評価の結果を含めていることが挙げられ、3ieのデータベースでは評価デザインによる検索ができる。具体的には、RCT（1974）、RDD（回帰・分断デザイン、90件）、PSM（傾向スコアマッチング法、499件）、IV（操作変数法、244件）、DID（差の差法、506件）、Mixed Method（混合法、80件）が検索できる。もちろん、ひとつのインパクト評価で複数のデザインや手法を用いている場合も多数あると理解される。なお、データが入手可能なのはわずか32件（0.6%）に留まっている。

また、3ieではメタアナリシスを積極的に実施しており、すでに解説したとおり、現時点で44件のSystematic Reviewの報告書が公表されている。さらに、600件以上のレビューおよびプロトコールを公表している。さらに、メタアナリシスはそもそものインパクト評価が十分に数多くあることが必須の条件となることから、各セクターにおいて、どのサブセクターあるいはどのタイプの介入行為のインパクト評価が不足しているかを解説したEvaluation Gap Map（EGM）を作成して公表している。また、そのEGMの報告書にも整理された教訓が記載されている。

3ieは、自らインパクト評価を実施することはなく、インパクト評価の案件を公示して公募でGrantを提供している。その総数は、304件（イン

パクト評価243件、システマティックレビュー（メタアナリシスの別名）38件、その他の調査23件）である。

2-2. 貧困アクションラボ（J-PAL）、IPA、3ieのSDGsへの貢献に関する言及

近年、国際開発を語る上ではずせないのが、Sustainable Development Goals（SDGs）である。厳格なインパクト評価を主導してきたこれらの3つの機関のSDGsへの貢献はどの程度であろうか。それぞれの機関のウェブサイトにはSearch（検索窓）が設置されている。これはそれぞれのデータベースに登録された報告書の本文まで対象とした検索が行われると理解される。それによると、SDGsへの言及は、貧困アクションラボが2件のみ、IPAが2件のみ、3ieが8件のみであった。3ieの結果が比較的充実しており、3ieの現職の事務局長による「How synthesized evidence can help with meeting the Sustainable Development Goals」（2016）と、3ieの初代事務局長であるWhiteの「Toward evidence-informed policies for achieving the Sustainable Development Goals」という論文があった。総じて、SDGsの議論への貢献は極めて限られていると言わざるを得ないであろう。

2-3. 新興の非営利の研究研究機関のファイナンスの現状

以上のように大規模にインパクト評価を実施しているこれらの組織の活動資金はどこから出ているのであろうか。

図3は、それぞれの年次報告書に記載のある収入・支出のページに記載のある寄付を円グラフにしたものである。独立の非営利組織であるIPA、3ieの情報は得られたし、3ieに関しては具体的な金額一覧も情報も入手できた。一方で、大学の一部という形態を取る貧困アクションラボ（J-PAL）の情報は得られなかった。ただし、3ieの設立を提言した2001年設立の独立の研究機関であり、学会誌も発行しているCenter for Global Development（CGD）の年次報告書に寄付先の一覧の情報が得られたので同様に円グラフにして表示した。収入規模は、IPAが42.5億円、3ieが8.5億円、そしてCGDが66.4億円であり（1ドル=100円換算）、アメリカの研究機関の巨大な資金規模がわかる。なお、寄付者は、すでに解説した通り、Bill and Melinda Gates FoundationとWilliam and Flora Hewlett Foundationを筆頭に民間の財団が並んでいる。ただし、USAID（アメリカ）、DFID（イギリス）、国連機関も寄付者に名を連ねていることもわかる。

3. 伝統的な援助機関におけるインパクト評価の取組

インパクト評価に取り組んでいる国際機関である世界銀行、そして二国間援助機関であるアメリカ国際開発庁（USAID）、イギリス国際開発局（DFID）、日本の国際協力機構（JICA）におけるインパクト評価の実施状況はどうなっているだろうか。以下に解説する（表4参照）。

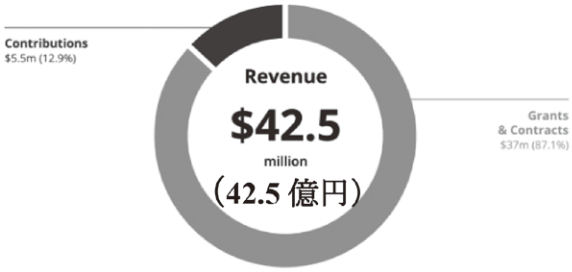
3-1. 世界銀行

世界銀行は、Development Impact Evaluation（DIME）というグループを立ち上げて、伝統的な援助機関の中では最も積極的にインパクト評価に取り組んでいる。2018年4月末時点で合計291件のPolicy Research and Working Papersを公表しており、200件以上がインパクト評価の結果を取りまとめた報告書である。データが得られた間近の過去3年間（2015-2017年）を見ると、2015年に40件、2016年に32件、2017年に33件のインパクト評価を実施している。デザインとしてはRCTが多いが、その他のデザインも用いているとしている。インパクト評価を60カ国で実施して、US＄180million（≒180億円）をインパクト評価の実施に投入して、US$180billion（≒18兆円）のレバレッジ（効果）を生み出したと報告している。政策に反映させるために、政策立案・実施の各フェーズで評価結果を活用しているとしている。

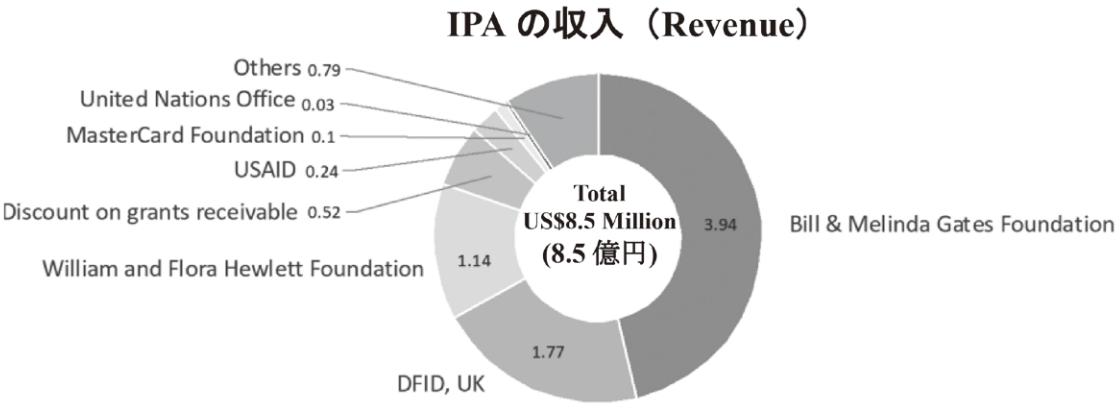
3-2. アメリカ国際開発庁（USAID）

USAIDでは、Development Experience Clearinghouseという部署を立ち上げて評価報告書を整理して提供している。データが得られた3年間

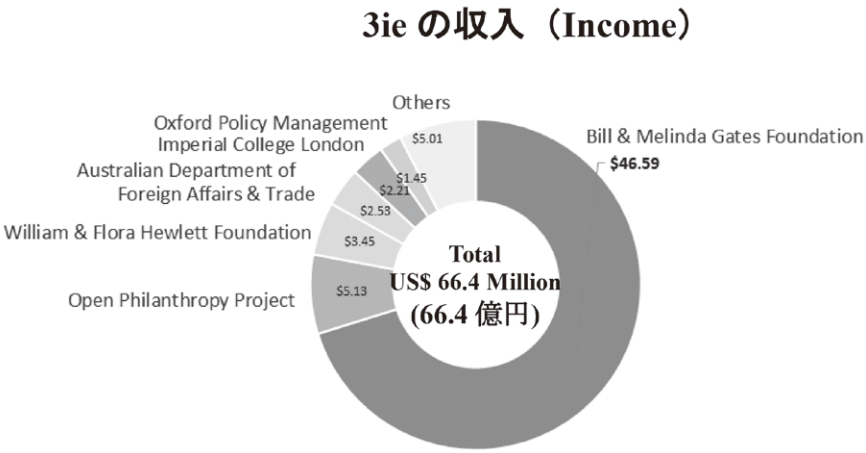
図3　3機関（IPA、3ie、CGD）の財源の内訳



（出所）IPA 2017 Annual Report



（出所）3ie Annual Report 2017



（出所）CGD Financial Transparency

3ieの設立を提言したCenter for Global Development(CGD)の収入（複数年のプレッジ）
（Funding Agreements Active in 2019）

表4　伝統的な援助機関（世界銀行、USAID、DFID、JICA）のインパクト評価の概要

機関名	世界銀行	USAID（アメリカ国際開発庁）	DFID（イギリス国際開発局）	JICA（国際協力機構（日本））
部署	Development Impact Evaluation (DIME)	Development Experience Clearinghouse	The Evaluation Unit (EvU) & The Centre of Excellence for Development Impact and Learning (CEDIL)	Evaluation Department, JICA Research Institute, Human Development Department, etc.
インパクト評価の概要	DIME generates high-quality and operationally relevant and research data to transform development policy, help reduce extreme poverty, and secure shared prosperity. The group conducts research in 60 countries with 200 agencies leveraging a $180M research budget against $18B in development finance. It also provides advisory services to 30 multilateral and bilateral development agencies in the world.	The study found that USAID was doing markedly fewer evaluations and a very limited number of in-depth, program evaluations. As part of a 2009 study, a survey of 296 recent evaluations submitted to the Development Experience Clearinghouse (DEC) found that while all reported that some data was collected from individual interviews, only 39 percent used a survey, only 9 percent reported on a comparison group, and only one used an experimental design involving randomized assignment.	In 2017/18 a new Evidence Department was formed within RED. The Evaluation Unit (EvU) is part of the Evidence Department. EvU has proposed a new initiative, to support the strategic use of randomised control trials, and other rigorous impact evaluations of DFID programmes. The aim is to collect in a more strategic way, highly rigorous evidence on the effectiveness of selected DFID programmes in areas where this evidence will provide major global good, and also to help improve the design of future programmes.	JICA conducts "Impact evaluation" in certain projects, which rigorously measures the changes brought in the target society by policies and projects. Using statistics and analyzing the gathered data, impact evaluation makes it possible to more accurately grasp the changes brought about by projects. To further enhance project effectiveness and quality, JICA has been promoting Evidence-Based Practice (EBP) and emphasizing the application of impact evaluation as its major tool.
インパクト評価の実施件数	More than 200 engagements in 60 countries. 291 Policy Research and Working Papers. $180 Million research program, leveraging $18 Billion in lending Type / Year: # of evaluation reports(IEG) / # of Impact Evaluation (DIME) 2015: 165 / 40 2016: 207 / 32 2017: 152 / 33	Number of evaluations at USAID and the evaluations including the word "Randomized Controlled Trial" are as follows[1]. Type / Year: # of evaluation reports published / # including the word "Randomized Controlled Trial" 2015: 188 / 23 2016: 145 / 20 2017: 161 / 17	DFID evaluations are published on DFID's open data site and on the UK government external website. For 2017/18, a total of 13 completed evaluation reports were published, with management responses published for 10. Type / FY: # of evaluation reports published / # including the word "Randomized Controlled Trial" 2015/16: 29 / 2 2016/17: 25 / 4 2017/18: 13 / 3	JICA publishes Impact Evaluation Reports in English with short Japanese summary (total 29 issues as of April 2019), in addition that it publishes external ex-post evaluation reports[2]. Type / FY: # of external ex-post evaluation / # of impact evaluation (IE) 2015: 90 / 6 2016: 99 / 5 2017: 45 / 3
インパクト評価のデザイン	Randomized Controlled Trials (RCT), non- or quasi-experimental methods such as regression discontinuity designs, difference-in-differences, propensity score matching, as well as utilization of instrumental variables	Randomized Controlled Trials (RCT) would not be well known among USAID staff. CLA's Evaluation Toolkit does not include the concept or explanation of RCT.	EvU has proposed a new initiative, to support the strategic use of Randomised Control Trials (RCT), and other rigorous impact evaluations of DFID programmes.	According to Impact Evaluation, most of them are the application of Randomized Controlled Trial (RC) and some are quasi-experimental designs.
政策への反映	DIME adopts a programmatic approach to Evidence-Based Policy-Making to increase economies of scale in research and learning	A 2016 study found that 90 percent of the evaluations reviewed were reported by USAID staff to have influenced some kind of programmatic action.	The Evaluation Use and Influence Strategy and Action Plan was developed, circulated and operationalised internally. Both documents will be incorporated in the DFID Evaluation Policy (2013) and Strategy (2014-19) refresh due to take place in 2018/19.	These evaluation results can then be used as reliable evidence by JICA for project management and by partner countries for policy-making.

（出所）各機関のAnnual Report 2015-2018; Rose, S. & Glassman, A.（2017）.

1 "The extremely small proportion of USAID impact evaluations (less than 1 percent, according to two recent studies) suggests that the agency is overlooking opportunities to answer questions about what works. (2009-2012)". (Rose, S. & Glassman, A. 2017)

2 In addition to external evaluation, significant number of internal evaluations are conducted every year.

（2015-2017年）の評価報告書のうち、Randomized Controlled Trialという単語に言及した報告書は、2015年 は188件 中23件、2016年 は145件 中10件、2017年は161件中17件であった。ただし、その言及が、RCTを適用したことの証左なのかどうかはそれぞれの報告書を深く読んでみないと分からない。一方、別の報告書によると、最近提出された296件の評価結果のうち、わずか39%がサーベイを実施した評価で、たった9%が比較グループを設定した評価であった。そして、わずかに1件がRCTデザインを適用したインパクト評価であったという報告があった（Rose, S. & Glassman, A. 2017）。USAIDの研修教材であるCLA（Collaborating, Learning, and Adapting）'s Evaluation ToolkitにもRCTの説明は出てこないようである。こうした状況を受けて、Center for Global Developmentでは、なぜUSAIDはEvidence-Based Practiceへの取組がなされていないかの議論がおこなわれ、その議論が掲載された（Rose,S. & Gassman, A. 2017）。それによると、(i) Strategic PlanとPerformance Measurementの組み合わせというマネジメントツールが普及しており、設定された指標の指標値を収集するモニタリングに忙しい、(ii) 人事ローテーションが頻繁に行われるため、Performance Measurementによる定期的な指標値の報告に重点が置かれ、結果が出るまで長期間を要するインパクト評価に取り組むインセンティブがない、などを理由としてあげている。そして、解決策として、(i) Evidence, Evaluation, and Learning（EEL）という新しい部局を創設すること、(ii) 各種のインセンティブを設定することなどが提案されている。この状況は、民間のマネジメント手法の導入を定めた画期的なGovernment Performance and Result Act（GPRA:1993）が、アメリカの政府機関に広く定着した結果であり、皮肉にもその定着が、USAIDにおけるEvidence-

Based Practiceの普及を遅らせてしまったというわけである。

3-3. イギリス国際開発局（DFID）

DFID内のResearch and Evidence Division（RED）の中にEvidence Departmentを設置して、さらにその中にEvaluation Unit（EvU）を設置して、Evidence-Based Practiceに取り組んでいる。EvUは、厳格なエビデンスの収集を行い、それをDFIDの援助プログラムに反映させることを業務としている。EvUでは、RCTをはじめとして、その他の厳格なデザインの適用を奨励している。DFIDの評価報告書の数はそもそも限られているが、2015/16年には29冊の報告書を公表し、そのうち2件でRandomized Controlled Trialに言及している。2016/17年度は、評価報告書25冊の公表のうち、4件で言及があった。2017/18では、評価報告書13冊の公表のうち、3件で言及があった。政策への反映のために2ページのサマリー（two-page, digestable summary）を用意することに力を入れている。また、EvUとは別に、The Centre of Excellence for Development Impact and Learning（CEDIL）を設置して、インパクト評価とメタアナリシス（Synthesisと呼称）に取り組むとしており、これは、イギリス版3ieといった役割を期待されていると理解される。なお、DFIDは、3ieなど、この論文で紹介した非営利の研究機関の寄付者に名を連ねているのも特徴的である。

3-4. 国際協力機構（JICA）

最後は日本のJICAであるが、日本の行政機関の中では最も積極的にインパクト評価を推進しており、国内では他の追随を許さないほどである。2019年4月の時点で合計29件のインパクト評価報告書を公表しており、用いられているデザインのほとんどはRCTである。これは国際開発分野の援助機関として、国内の横並びでは全く足りずに、国際的な横並びを求められることに起因していると考えられる。インパクト評価の公表件数について、データが得られた過去3年（2015-2017年度）を見ると2015年度が6件、2016年度が5件、2017年度が3件となっている。また、2018年度は2件の公表があった。この傾向に関して、RCTがインパクト評価という形を取らずに、個別の技術協力プロジェクトにおける実証フェーズで広く使われるようになったという説明がなされることがある。また、JICA評価部だけではなく、JICA研究所の他、課題部でもインパクト評価を実施している。なお、インパクト評価を実施して、その結果がどのように日本の援助の政策選択あるいは事業選択に役立ったのかの研究とその報告はまだなされていないと認識しており、研究の実施が待たれるところである。

4. 考察と今後の課題

以上、新興の非営利の研究機関と伝統的な援助機関におけるインパクト評価の取組の現状を解説したが、以下の点が考察と今後の研究の課題として挙げられる。

4-1. 50年の時を経て実現に向かう「実験する社会」と今後の課題

冒頭でも述べたが、1969年にキャンベルが「実験する社会」（Experimental Society）を実現しようと宣言した。今の言葉で言うなら、RCT（キャンベルはExperimental Designと呼称した）を用いてエビデンスを産出して、合理的な意思決定を行っていこうという宣言である（Campbell, D.T. 1969, 1998）。このキャンベルの宣言は、それから50年を経て、先進国に限らず世界中で実現しつつあると言えるであろう。国際開発分野でのRCTを用いたインパクト評価の普及がその証左である。キャンベルの先見の明に感服するのみである。この「実験する社会」の実践の普及は今後も推進されていかねばならない。

4-2.「プロジェクト・アイランド」とインパクト評価の対象レベルに関する生来的な問題

国際開発の分野では、RCTを適用したインパクト評価の件数が爆発的に増加したし、その後、メタアナリシスやセクター別レビューも実施され、整理された知見が公表されてきた。それにも関わらず、それらの知見が思ったよりも用いられていないことが指摘された。これは、いわ

ば2000年代に議論された「プロジェクト・アイランド」（Project Island）あるいは「アイランド・エフェクト」（Island Effect）の状況が続いていることを示唆しているのであろう。つまり、貧困アクションラボ、IPA、3ieなどの研究機関や、それぞれの援助ドナーがプロジェクト・レベル（個別事業レベル）で「成功！成功！」と言っているが、国全体あるいはセクター全体としての状況は変わっていないという状況である。この反省から、2000年代にはプロジェクトを離れてSector-Wide Approach（SWAPs）（別名セクタープログラム（Sector Program））の導入が進んだはずであった。その過程で、「プロジェクト」という考え方に対する激しい批判があり、当時、日本の援助機関は防戦一方だった（国際開発センター 2003）。北欧ドナーなどは、プロジェクトという用語を使うのをやめて、プログラム・コンポーネントという用語を使い出していた。しかし現在、インパクト評価が対象としているのは、まさにその「プロジェクト」である。つまり、RCTが適用できるのは、プロジェクト・レベル（個別事業レベル）であり、それ以上の政策－戦略－プログラムのレベルを対象とできないという生来的な問題があるのではないだろうか。RCTを使用したインパクト評価の試みが、この問題を乗り越えられるのかどうかは定かではない。また、この指摘はEvidence Gap Mapをはじめとする政策への反映の実践に関しても言えることであり、本来は、ある国あるいはセクター全体に関する政策/戦略が策定され、続いてプログラムが策定され、そして最後にプログラム・コンポーネント（あるいはプロジェクト）が配置されるという流れが推奨されるのに対して、例えばEvidence Gap Mapは単なるプロジェクトの類型の羅列であると言わざるを得ない。「プロジェクトを束ねてもプログラムにはならない」とは当時頻繁になされた批判である。現在の実践は、この2000年代の「プロジェクト」に対する批判がそのまま当てはまることになり、その組み立て方に関して再考を要する。

4-3. 業績測定（パフォーマンス・メジャーメント）と評価の混同に起因する問題

USAIDの評価の中で、RCTを用いたインパクト評価の件数が少ないことが指摘されていて、その状況の改善が議論されていた。ただし、USAIDが導入したのは、戦略プランと業績測定の組み合わせ（Strategic Planning and Performance Measurement）であり、民間企業が用いているマネジメント手法であることが指摘された。それに対して、評価は、Rossiらが定義するように、社会科学的な手法を適用した営みであり、①セオリー評価、②プロセス評価、③インパクト評価、④効率性評価で構成されると定義されている（Rossi et al 2003）。つまり、業績測定（パフォーマンス・メジャーメント）と、評価はそもそも別の行為でありながら、混同されて議論されているのが現状である。もし、USAIDの実践を非難するのであれば、同じ非難が、業績測定の実践を導入している国際機関および国連が主導するSDGsにも当てはまるだろうし、同様にGPRAによって業績測定の実践が導入されているアメリカの行政機関すべてに当てはまることになる。業績測定と評価の関係については多数の議論があり（龍・佐々木 2004）、それらの結論は、両者は相互補完的であるので両方を組み合わせて使うべきということである。したがって、RCTを用いたインパクト評価を推進する勢力も、業績測定の導入はRCTとは違った意義があることを理解して、上手に利用していくことを考える必要があるだろう。

4-4. NGOsとの協働による介入実施とデータ収集の可能性

この論文で解説した新興の非営利の研究機関は、実際の介入行為とデータ収集を国際NGOsと協働して実施していることが分かる。対して、日本では研究機関とNGOsの協力は概して希薄である。村民レベルまで入っていってデータ収集できるのはNGOsの強みである。NGOsの側はごく限られた対象地域で成果を上げて満足するのではなく、その限られた地域を対象とした厳格なインパクト評価の実施に協力することを通じて、世界的な政策変更に影響を与えることを考

えるべきである。また、研究機関の側も、事業資金の支援や、論文の共著者として迎えることを含めたNGOsとの協働の可能性を探るべきである。

4-5. IT業界のバックアップによるEBPMの普及と今後の課題

本論文で解説した新興の3つの非営利の研究機関は、いずれの機関にも共通する有力な資金支援者がいることがわかる。それはThe Bill and Melinda Gates Foundation、The William and Flora Hewlett Foundationなどである。これらのIT事業の成功に由来する資金支援者が、既存の政府系の援助機関ではなく、独立かつ非営利の研究機関を支援しており、それが昨今のEvidence-Based Policy Making（EBPM）を盛り上げていることがわかる。これに対して、伝統的な援助機関では、世銀が健闘しているが、USAIDなどはこの動きに追いついていないことがわかる。国際開発における研究に関して主役交代が起きているのかもしれない。これは国際開発全体に関しても言える傾向で、国際機関・二国間援助機関の援助量よりも、民間の資金流入がその国の開発を左右し始めた最近の状況と同じ傾向が見て取れるのかも知れない。この傾向が続くのかどうか今後の動きを注視する必要がある。なお、日本ではこうした規模の大きい資金支援者がそもそもいないことから、個々の研究者の細々とした活動に留まらざるを得ない状況であり、たいへん残念な状況である。あるいは、日本においてもEBPMの推進のために、ITで成功した資金余力のある企業経営者に対する働きかけが必要であることが示唆されているのかもしれない。

参考文献・ウェブサイト

青柳恵太郎（2017）『途上国開発分野におけるエビデンスの活用』、RIETI 主催EBPM シンポジウム「エビデンスに基づく政策立案を推進するために」Available at: https://www.rieti.go.jp/jp/events/17121901/handout.html

浅岡浩章（2016）「国際開発分野におけるエビデンス活用の現状と課題」、『日本評価研究』、17（1）Available at: http://evaluationjp.org/files/Vol17_No1.pdf

国際開発センター（2003）『IDCJフォーラム：特集/援助協調を超えて』、Available at: http://dl.ndl.go.jp/view/download/digidepo_2392454_po_forum23.pdf?itemId=info%3Andljp%2Fpid%2F2392454&contentNo=1&alternativeNo=&_lang=en

龍慶昭・佐々木亮（2004）『政策評価の理論と技法（増補改訂版）』、多賀出版

Abdul Latif Jameel Poverty Action Lab.（2018）. *Annual Report 2018 and Look Ahead to 2019*. Available at: https://www.povertyactionlab.org/annual-report-2018

Abdul Latif Jameel Poverty Action Lab.（2018）. *Policy Insight（Microfinance）*. Available at: https://www.povertyactionlab.org/policy-insight/microcredit-impacts-and-limitations

Cameron, D.B., Mishra, A. & Brown, A.N.（2015）. The growth of impact evaluation for international development: how much have we learned?, *Journal of Development Effectiveness,* 8-1, 1-21. Available at: https://doi.org/10.1080/19439342.2015.1034156

Campbell, D.T.（1969）. Reform as Experiment, *American Psychologist*, 24（4）, pp.409-429.

Campbell, D.T. and Russo, M.J.（1998）. *Social Experimentation*. Sage Publication. CA: Thousand Oaks.

Center for Global Development（2006）. *When Will We Ever Learn? Improving Lives Through Impact Evaluation.* Available at: https://www.cgdev.org/publication/when-will-we-ever-learn-improving-lives-through-impact-evaluation

Center for Global Development（2008）. *Q&A with Howard White, Head of New Impact Evaluation Entity, 3IE*. Available at: https://www.cgdev.org/article/qa-howard-white-head-new-impact-evaluation-entity-3ie

Center for Global Development（2019）. *Financial Transparency - Funding Agreements Active in 2019*. Available at: https://www.cgdev.org/section/funding

Chliova, M, Brinckmann, J. & Rosenbusch, N.（2013）. Is microcredit a blessing for the poor? : Meta-analysis, *Frontiers of Entrepreneurship Research:* 33（18）, Article 2. Available at: http://digitalknowledge.babson.edu/fer/vol33/iss18/2

International Initiative for Impact Evaluation（2017）. *Annual report 2017: Evidence - Action - Impact.*

Available at: http://www.3ieimpact.org/sites/default/files/2019-02/3ie-annual-report-2017.pdf

Freeman, H.E., Rossi, P.H., & Wright, S.R.(1979). *Evaluating Social Projects in Developing Countries*. Development Centre, OECD.

Innovations for Poverty Action(2018). *2017 Annual Report.* Available at: https://www.poverty-action.org/publication/2017-ipa-annual-report

Masset, E. & Jimenez, E.(2016). *How synthesized evidence can help with meeting the Sustainable Development Goals.* Available at: http://www.3ieimpact.org/blogs/how-synthesised-evidence-can-help-meeting-sustainable-development-goals

Newman, J., Rawlings, L., & Gertler, P(1994). Using randomized control designs in evaluating social sector programs in developing countries. *The World Bank research observer*. 9(2). Available at: http://documents.worldbank.org/curated/en/228191468774882721/pdf/multi-page.pdf

Orr, L.L.(1998). *Social Experiments: Evaluating Public Programs With Experimental Methods 1st Edition.* Sage Publication. CA: Thousand Oaks.

Rose, S. & Glassman, A.(2017). *CGD Notes: Advancing the Evidence Agenda at USAID.* Available at: https://www.cgdev.org/publication/advancing-evidence-agenda-usaid

Rossi, P.H., Freeman, H. & Wright, S.R.(1979). *Evaluation: A Systematic Approach.* Sage Publication. CA: Thousand Oaks.

Rossi, P.H., Lipsey, M.W. &Freeman, H.(2003). *Evaluation: A Systematic Approach, 7th ed.* Sage Publication. CA: Thousand Oaks.

Shayda M.S. & Brown, A.N.(2018). Is impact evaluation still on the rise? The new trends in 2010–2015, *Journal of Development Effectiveness,* 10(3). Available at: https://doi.org/10.1080/19439342.2018.1483414

White, H. & Masset, E.(2018). The rise of impact evaluations and challenges which CEDIL is to address, *Journal of Development Effectiveness. 2018*, 10(4), pp.393-399.

White, H.(2018). *Toward evidence-informed policies for achieving the Sustainable Development Goals.* Available at: http://3ieimpact.org/blogs/toward-evidence-informed-policies-achieving-sustainable-development-goals

World Bank(1992). *Effective implementation : key to development impact (Wapenhans Report)* (English). Washington, DC: World Bank. Available at: http://documents.worldbank.org/curated/en/596521467145356248/pdf/multi0page.pdf

The Abdul Latif Jameel Poverty Action Lab
https://www.povertyactionlab.org/
International Initiative for Impact Evaluation(3ie)
http://www.3ieimpact.org/
Innovations for Poverty Action(IPA)
https://www.poverty-action.org/

(2020.1.31 受理)

The current status of Evidence-Based Policy Making (EBPM) at international development field

Ryo Sasaki

International Development Center of Japan Inc.
sasaki.ryo@idcj.or.jp

Abstract

As a practice of EBPM: Evidence-Based Policy Making (EBPM), impact evaluation has been widely practiced in international development field. The design widely used in the field is Randomized Controlled Trial (RCT). This article compares and discusses the three emerging not-for-profit research institutes that disseminate the database of evaluation reports, J-PAL (The Abdul Latif Jameel Poverty Action Lab), IPA (Innovations for Poverty Action) and 3ie (International Initiative for Impact Evaluation). In addition, the traditional aid agencies, such as the World Bank, USAID of the United States, DFID of the United Kingdom and Japanese JICA will be compared and analyzed.

Based on those analysis, the emergent agenda will be proposed. They are: (1) the idea of "experimenting society" proposed by Campbell is now in reality and this current practice of impact evaluation should be disseminated more widely; (2) The reason why application of impact evaluation results is limited would be the old discussion is still continuing, that is "Island Effect". All aid agencies claimed their success of individual project but the vast ocean (i.e., a whole country or a whole sector) remains unchanged; (3) The confusion between the practice of strategic planning/performance measurement and the practice of program evaluation which root is social science is widely observed, and they are actually complement each other, thus the camp for pushing impact evaluation using RCT design should understand there is another thought of strategic planning/performance measurement and there is possibility of collaboration. (4) Possibility of collaboration between research institutes and international NGOs should be pursued in the process of intervention and data collection at field; (5) Major private foundations which roots are IT industries provide vast financial support to the emerging not-for-profit research institutes and this might be the same trend observed at the whole international development field, that is "from government to private sector".

Keywords

Impact Evaluation, RCT, J-PAL, 3ie, IPA

【研究論文】

地方創生アンケートから評価する市町村の合併と連携

村上　裕一
北海道大学
yuichim@juris.hokudai.ac.jp

要　約

本稿では、筆者らが2016年から2017年にかけて実施した地方創生アンケートの結果を、「平成の大合併」における合併・非合併自治体の異同に注目して分析した。これは、実地調査の成果を踏まえ、フランスとの比較も織り交ぜながら、「平成の大合併」と地方創生の政策手段・プロセスとしての市町村合併・連携を評価し、その条件や効果、自治体の思考様式の理論構築を試みるものである。その結果、非合併自治体は、地方創生によっても広域連携を促されることはあまりなかったのに対し、近隣自治体よりも域内外産業との連携を志向しがちであることが明らかになった。合併自治体は、合併により自治体職員の専門性を十分向上させたとは言い切れず、意思決定が役所内で完結しがちであることから住民参加にも消極的になっている可能性がある。とはいえ非合併自治体と比べると、合併により隣接自治体との調整なく独自で域外、さらには国に対し積極的な行動を起こせることが明らかになった。

キーワード

地方創生、市町村合併、自治体間連携、政策評価、フランス

1．はじめに

(1) 本稿の目的

村上・小磯・関口（2017）と村上・小磯・関口（2018）では、2016年から2017年にかけて、一般財団法人北海道開発協会の協力も得ながら北海道、愛媛県、香川県の市町村に対して実施した、地方創生に関するアンケートの結果を紹介している。

アンケートは全50問に及び、回答は3道県で190にも上った（回答率：88.0%）。それは地方創生に直接関わった自治体職員の生の声を反映していることから、地方の行政実態を知る上で資料的価値が極めて高いと言える。アンケート結果についてはこれまでにも様々な切り口で分析し、成果を発表してきた（自治体の多様なインバウンド対応のカテゴリーについて（嶋・児矢野・野田・村上・津田・樽本・落合2017、pp.144-146）、政策現場と内閣主導の「距離」について（村上2018a）、北海道と四国の比較について（村上2018b）、国土政策や自治政策の歴史と地方創生との関係について（Murakami 2019））。

本稿では、そのアンケート結果をまた別の新たな視点、すなわち市町村合併（もしくは「平成の大合併」）との関係という視角で分析し、その結果を整理してみたい。

2014年からの地方創生（第1弾）では、広域連携を交付金申請の条件にするなど、水平的な自

日本評価学会『日本評価研究』第20巻第2号、2020年、pp.105-119

表1　市町村の合併・連携の政策評価と地方創生との関係

	過去		「平成の大合併」		地方創生		将来の政策
年			1999～2010 年		2014 年～		
目的			・人口減少・少子高齢化等の社会経済情勢の変化や地方分権の担い手となる基礎自治体にふさわしい行財政基盤の確立（総務省『「平成の合併」について』（平成 22 年 3 月））		・人口急減・超高齢化という我が国が直面する大きな課題に対し、政府一体となって取り組み、各地域がそれぞれの特徴を活かした自律的で持続的な社会を創生（まち・ひと・しごと創生『長期ビジョン』ほか）		・人口急減・超高齢社会に適した持続可能な仕組 ・「社会構造の変化に強靭な自治体行政」（※1）
手段			・合併特例債等により市町村合併を促進（**2**．を参照）。		・各自治体が策定する『総合戦略』を通じて、自治体間連携へと誘導		・市町村合併～圏域論～柔軟な自治体間連携？
結果		n 期	(1) 6 割強の自治体が合併を決定。平均人口は 68,947人（2010 年 3 月）に増。「目的」の行財政基盤の成否（**2**．(1) も参照）。 (2) 相対的には合併に否定的評価。プラス効果として、財政支出の削減、職員の能力向上。マイナス効果として、行政と住民相互の連帯の弱まり、財政計画との乖離、周辺部の衰退（全国町村会『平成の合併をめぐる実態と評価』(2008 年 10 月)）。	n+1 期	(1) 実際にどれだけの自治体間連携が見られたか。KPI（重要業績評価指標）に対して、例えば、各自治体の人口動態。「目的」に書かれた「持続的な社会」「創生」の成否。 (2) 自治体の地方創生担当者を対象としたアンケートによれば、「地方創生は各自治体に水平的連携を促進しなかった」、「地方創生が各自治体の意思決定や問題意識に変化をもたらすことはなかった」、「企業版ふるさと納税へのスタンスは消極的である」など（**2**．(3) を参照）。	n+2 期	

（注）※1では総務省「自治体戦略2040構想研究会」の文言を参照した。計12本の細い矢印は、政策評価の実施・活用の方向性を指す。
　　山崎（2018）は、この歴史を自治政策と国土政策の中に位置付ける。
（出所）表中に記した各種資料を基に、筆者作成。

治体間連携への誘導策が埋め込まれた。国が推進しようとしたこの水平的な自治体間連携の成否と「平成の大合併」とは、一体いかなる関係に立つのだろうか。本稿は、地方創生という「実験」についてのアンケート結果に、新たな切り口で分析を加える試みである。

(2)「評価」の対象と手法

・対象と位置付け

本稿では、地方創生を直接担当した自治体職員を対象としたアンケート結果を基に、市町村合併や水平的な自治体間連携の「評価」を行う。このことの意味を整理したのが、表1である。

まず、「平成の大合併」や地方創生を1つのまとまりを持った政策だとすると、それにはそれぞれ上記の目的・手段・結果がある。両政策はいずれも①自治体（市町村）を対象とし、②既存の市町村の境界を超えた何らかの活動（合併や連携）による自治体行政の効率化を「手段」とし、③人口問題対応とそれに耐え得る行財政基盤の確立を政策の「目的」に掲げているという点で共通している。こうしたことから、両政策は表1のように、1つの時間軸の上に載せて捉えることができる。さらに、これらの前後に同種の政策が存在する（し得る）ことも想定できる。

こうした時間軸の中では、表1の12本の細い矢印が示すように、過去の政策の「手段」や「結果」を評価して得られた教訓を、将来の政策の「目的」や「手段」に活かしていくことが期待される。これは、政策評価実務において、繰り返し行われてきたことである。

「平成の大合併」での合併の有無と地方創生アンケートの結果とをクロスさせることにより、市町村の合併・連携という政策手段・プロセスを評価しようという本稿の手法には、確かに多くの問題点がある（後述）。が、表1の時間軸の中では、①いかなる特性を持った市町村が「平成の大合併」で合併した（しなかった）のか（n期の評価）、②合併・非合併自治体が、地方創生に対していかなる態度をとったか（そうした態

度をとったのはなぜか）（n+1期の評価）、③地方創生でいかなる態度をとった自治体が、今後市町村合併・連携に向かい得るか（n+2期の評価）という3つの問い（のいずれか）に、手掛かりを与え得るものである。特に①や②に該当する場合は、本稿が約20年前の「平成の大合併」そのものを政策評価していることになる。

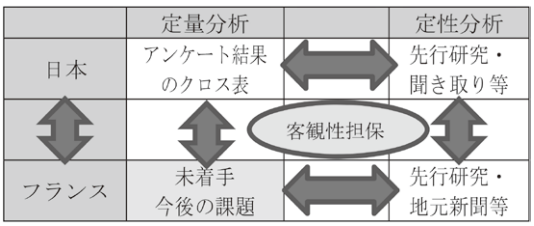
ここで留意すべきは、本稿のクロス表で見えるのが、あくまで合併経験の有無と地方創生アンケート結果との相関関係に過ぎないということである。例えば、仮にクロス表から「非合併自治体には地方創生に期待しない自治体が多かった」という関係が見出せたとして、実際のところ、地方創生のような国策に期待しない自治体が「平成の大合併」で合併しなかったのか、「平成の大合併」で合併しなかった自治体が（何らかの理由により）地方創生に期待しなかったのかは即断できない。ただ言えるのはそこに相関関係があるということだけであって、本稿の成果が上記の①～③のどれに当たるのか（すなわち、因果関係や一般性）については、追加的な検討が必要である。

ただし、いずれにしても、市町村合併や自治体間連携という政策手段についての評価をしていることには違いない。すなわち、市町村の合併・連携がいかなる特性を持った自治体に有効であり得るか、あるいは、それらの手段が自治体にどのような変化をもたらし得るかという、いわば市町村合併・連携の条件や効果に関する一般理論に繋げられる可能性はある。

・実地調査の成果の反映

本稿では、ひとまず、自治体の合併の有無と地方創生を直接担当した当該自治体職員を対象にしたアンケートの結果とをクロスさせるという、定量的な分析手法を採用している。その一方で、その分析に当たっては、関連する各種先行研究や個別聞き取り・実地調査の成果を踏まえるとともに、フランスの市町村合併や自治体間連携の制度・運用に目を向けるという定性的な分析手法も用いることで、実態をより客観的かつ立体的に捉えようとしている。本稿では、表2のように実態を定量・定性の両面から分析しリンクさせることで、各手法の限界を補おうとしている。

表2　本稿における定量・定性分析



（出所）筆者作成。矢印は「相互参照」を意味する。

前述の通り、クロス表の相関関係を基にした分析の結果は、必ずしも因果関係を明らかにするものではない。すなわち、「平成の大合併」そのもの、もしくは、一般的特性としての市町村合併の有無を、地方創生の手段や結果に対する「原因」と見るのか、あるいは「結果」と見るのかは、時系列等に沿って改めて検証せざるを得ない。

なお、表2の整理に基づき本研究を通して得られる成果は、現段階で未着手となっているフランスの自治体に対するアンケートの研究設計にも資するものである。

・評価の主体について

政策評価には、まず、担当者による内省的評価か、第三者による客観的評価かという分類がある。本稿の評価自体は筆者自身によるものにほかならないが、その基となるデータが客観的・外形的なもの（例：合併自治体数、人口動態、行財政基盤確立の有無）か、主観的・内心的なもの（例：自治体住民の満足度や自治体職員の認識といったアンケート結果）かという点も、評価主体に関わる重要問題である。政策担当者が自ら政策を振り返ることにも意義はあるが、担当者の認識を超えた、客観的指標を用いた岡目八目の実効的な評価が望ましいとする立場からすれば、担当者の認識内容を基にした評価には限界がある、ということになる。これについては、表1の「結果」欄において、(1) 客観と (2) 主観とに大まかに区別して記してある。

本稿では、地方創生を直接担当した自治体職

員の認識を聞いたアンケートの結果を1つの重要データとしていることから、上記のような客観性において限界を抱えていることは否定できない。すなわち、これはあくまで自治体職員の認識の、筆者による評価・分析に過ぎない。とはいえ、村上・小磯・関口（2017）にある通り、①アンケートの問いは役所外の第三者的な調査者が手順を踏んで設定していること、②集計作業を通して回答の個性が失われ、全体的な傾向が捉えられるようになっていること、③合併・非合併という客観的指標とクロス分析をしていることなどから、（担当自治体職員が実際にどう感じたかという意味での主観性は残っているものの）分析を妨げる過度な主観性は排除されている。さらに本稿では、制度的・歴史的前提条件の違いを考慮に入れた上で、フランスの状況との比較を行っている。異なる国々で単純比較はできないが、互いの制度と運用を客観視する上で、これは有意義だと考えている。その意味で、本稿における評価結果のフィードバック先は、大きな傾向を見出しているという点において、データ提供者かつ合併・連携の当事者である自治体（職員）でもあり得るし、自治体行財政基盤を確立するべく市町村合併や自治体間連携の仕組み作りに携わる政府（政治家や府省行政官）でもあり得る。

(3) 本稿の構成

本稿では、まず「平成の大合併」を簡単に振り返った上で、合併・非合併自治体の地方創生へのスタンスの相違（2.（3））を見る。続いて、市町村合併が自治体の能力を高めるのか（3.（1））、自治体の対外関係をどう変えるのか（同（2））、自治体の意思決定（地方政府の民主主義）にいかなるインパクトを与えるのか（同（3））といった観点で、アンケート結果に分析を加えたい。これは、市町村合併・連携の進む条件や効果、自治体の思考様式を導くことを目指したものである。

日本の市町村合併の特徴や本稿における評価の限界については、フランスにおける市町村合併等の経緯（4.（1））、現況（同（2））、個別事例（同（3））と比較することによっても、浮かび上がらせることができる。

2.「平成の大合併」について

(1) 概容

日本では、合併特例法（1965年制定）が1999年に地方分権一括法で改正され、①町村から市へ（5万人→3万人）や市から政令指定都市へ（100万人→70万人）の移行の人口要件緩和、②合併後の議員定数や交付税額変更の猶予措置、③有利な合併特例債による財政支援措置の拡充といった誘導策により、「駆け込み」も含め、市町村合併が進んだ（曽我2019、p.163）。

日本では、この「平成の大合併」により町村数が3分の1程度になり、2019年8月現在、市町村数は1,718にまで減っている。表3・表4から明らかなように、人口1万人前後の町村の減少率が最も高く、それらが合併した結果、人口3万人未満の市の増加率が高くなった。

「平成の大合併」は、2000年代をピークとする。それに伴う市町村減少率は、大阪府・東京都・神奈川県といった大都市圏の低さが目立ち、北海道も15.6％（212→179）に止まる一方、長崎県が73.4％（79→21）、広島県が73.3％（86→23）、

表3　日本の人口規模別自治体人口

	1990年		2015年	
	数（千人）	%	数（千人）	%
市				
～3万人	1,561	1.3	2,124	1.7
～10万人	21,731	17.6	25,274	19.9
～30万人	23,825	19.3	30,790	24.2
～100万人	23,232	18.8	28,446	22.4
100万人～	25,296	20.5	29,503	23.2
市総数	95,644	77.4	116,137	91.4
町　村				
～5000人	1,960	1.6	725	0.6
～1万人	6,528	5.3	1,772	1.4
～3万人	15,593	12.6	6,020	4.7
3万人～	3,887	3.1	2,441	1.9
町村総数	27,968	22.6	10,958	8.6
市町村総数	123,611	100.0	127,095	100.0

（出所）国立社会保障・人口問題研究所（2020、p.157）のデータを基に、筆者作成。

表4　日本の人口規模別自治体数

	1990年		2015年	
	数	%	数	%
市				
～3万人	63	1.9	91	5.3
～10万人	384	11.8	439	25.6
～30万人	144	4.4	189	11.0
～100万人	54	1.7	60	3.5
100万人～	10	0.3	11	0.6
市総数	655	20.2	790	46.0
町　村				
～5000人	629	19.4	267	15.5
～1万人	897	27.6	242	14.1
～3万人	961	29.6	354	20.6
3万人～	103	3.2	65	3.8
町村総数	2,590	79.8	928	54.0
市町村総数	3,245	100.0	1,718	100.0

（出所）国立社会保障・人口問題研究所（2020、p.157）のデータを基に、筆者作成。

愛媛県が71.4％（70→20）、香川県が60.5％（43→17）と、西日本で高い傾向が見られる。

地方創生に関するアンケートでは、愛媛・香川両県の回答数がそれぞれ20、17とあまり多くないため、抽出した各道県の特徴と地方創生のアンケート結果とを直接結び付けることは難しい。他方、3道県合わせた190市町村中、「平成の大合併」でどことも合併しなかった「非合併自治体」が146（76.8％）、どこかと合併した「合併自治体」が44（23.2％）だった。このことから、辛うじてではあるものの、合併の有無とその回答との間に有意な関係性を見出せるのではないかと考えた。本稿では、それらのクロス表を基に考察を行う。

（2）合併・非合併指標の検証

果たして、合併の有無そのものが有意な変数か（合併の有無と対応関係にある変数で、それを代替し得る、より直接的なものがあるのではないか）という問題がある。これについて、例えば、全国の市町村を対象とした「平成の大合併」の計量分析では、人口が小さいほど、人口減少率が大きいほど、財政力が低いほど、合併する傾向が見られたという。確かにそれは直感にも合う。しかし他方で、人口が小さかったり人口減少率が大きかったりしても、何らかの理由で非合併のままの自治体や、財政状況が良いからこそ周囲を寄せ付けず非合併のままの自治体など、例外もないわけではない（曽我2019、pp.167-168）。

そこで本稿では、まず今般のアンケート結果のみを対象に、合併の有無と他のいくつかの要素との関係性を検証した。そうしたところ、ダミー変数化した合併の有無と各自治体の①人口、②財政力指数、及び、③若年女性人口減少率との相関係数は、表5のようになった。

表5　合併の有無と自治体の特性との相関

	相関係数	関連性の有無
①人口	0.105	無い
②財政力指数	0.263	ほとんど無い
③若年女性人口減少率	0.279	ほとんど無い

（出所）総務省等のデータを基に、筆者作成。

すなわち、今般の地方創生に係るアンケート結果において、これらの相関性はいずれも低く、合併の有無に代わり得る変数ではない。したがって、今般のアンケート結果に限っては、基本的には合併の有無と各回答との直接的な関連性を推測することができそうである。一定のサンプル数を備えたアンケート結果から巨視的に傾向を見出すことには、それなりの理論的意義が認められよう。ただし、合併・非合併という単純な二分法に問題があるということについては、後で検討する。

それでは、以下、クロス表に基づき分析を進め、11個のテーゼを示したい（なお、表でパーセンテージを四捨五入した結果、合計が100.0になっていないものがある）。

（3）地方創生と市町村合併

・非合併自治体への連携促進効果は乏しかった

「今般の地方創生を受けて、あなたの自治体と周辺の自治体との連携は促進されたか」という問いに対し、「促進された」と「促進されず（以前とあまり変わらなかった・促進されなかった）」の回答状況は表6のようになった。

全体的に地方創生は各自治体に水平的連携を

表6　合併の有無と水平的自治体間連携促進の有無

	促進された	促進されず	合計
非合併	32.2	67.8	100.0
	47	99	146
合　併	44.2	55.8	100.0
	19	25	44
合　計	34.7	65.3	100.0
	66	124	190

（注）上段は行の中での割合、下段は市町村数。
（出所）アンケート結果を基に、筆者作成。

促進しなかった（村上2018b）が、中でも、その誘導策のターゲットだったとも言うべき非合併自治体への効果が乏しかった。

総務省資料「未合併要因に関する調査」は、非合併自治体が「平成の大合併」の潮流に乗らなかった理由として、①合併について意見集約ができなかった（33.7％）、②合併せずに単独で運営していこうと考えた（30.8％）、③合併を望んだが、相手がそれに消極的・否定的だった（26.4％）を挙げている。

そうした非合併自治体は通常単独で、もしくは必要に応じてせいぜい一部事務組合等による機能的な連携をすでにしているとみられる。また、地方創生が自治体にあまり時間的余裕を与えない政策でもあったことから、やはり今般の地方創生が非合併自治体にさらなる連携を促進する効果はあまりなかったと考えられる。

・非合併自治体に地方創生はあまり響かず

「地方創生を受けて、あなたの自治体における意思決定のあり方や問題意識は変化したか」という問いに対し、「変化した」と「変化しなかった・その他」の回答状況は表7のようになった。

表7　合併の有無と地方創生による変化

	変化あり	変化なし・他	合計
非合併	27.6	72.4	100.0
	40	106	146
合　併	43.2	56.8	100.0
	19	25	44
合　計	31.1	68.9	100.0
	59	131	190

（注）上段は行の中での割合、下段は市町村数。
（出所）アンケート結果を基に、筆者作成。

全体的に、地方創生が各自治体の意思決定や問題意識に変化をもたらすことはなかった（村上2018b）が、その傾向は、非合併自治体により強く表れている。

可能性として考えられるのは、①そもそも非合併自治体の方が人口減少など自分自身の将来に対する問題意識が強く（したがって、すでにそのための具体的な行動をとっており）、地方創生はただその後を追う「国策」に過ぎなかった（したがって、自治体では変化がなかった）ということ、②非合併自治体は「平成の大合併」の潮流にも乗らないほど単独行動志向であり、そもそも国の政策によって影響されにくい自治体だった（合併自治体はその逆だった）ということである。あるいは、より現実的には、③地方創生に充分対処するだけのリソース（時間、専門性、人員など）が足りていなかった、ということもあり得よう。可能性の①では非合併自治体の問題対応能力が高いことが前提になっているが、②ではその能力が高い場合も低い場合もあり得ることに注意が必要である。

・非合併自治体の方がふるさと納税に対し貪欲

企業版ふるさと納税は、企業が自治体に寄付をすると税負担が軽減される制度で、2016年4月に実際に運用が始まった。これへの取り組みについて、「導入に向けて積極的に検討中」と「その他・消極的」の回答状況は表8のようになった。

表8　合併の有無と企業版ふるさと納税へのスタンス

	積極的	消極的	合計
非合併	22.8	77.2	100.0
	33	113	146
合　併	15.9	84.1	100.0
	7	37	44
合　計	21.1	78.9	100.0
	40	150	190

（注）上段は行の中での割合、下段は市町村数。
（出所）アンケート結果を基に、筆者作成。

前述のように、地方創生そのものに対して非合併自治体は薄い反応を示していたにもかかわらず、企業版ふるさと納税に関しては非合併自治体の回答の方がやや積極的だという結果であ

る。全体的には企業版ふるさと納税へのスタンスが消極的だという結果となっているが、どちらかと言うと、非合併自治体の方がそれに貪欲だった。

そうなった理由として考えられるのは、①非合併自治体がそもそも国や道県を含む他自治体に頼るのではなく、「自主財源」を獲得することの方に積極的であること、さらには、②そのために普段から納税を含め様々な形で「組む」ことのできる民間企業を想定していることなどである。

3. 基礎集計とその分析

本章では、引き続きクロス表を基に、市町村合併と自治体行政との関係を分析する。

(1) 市町村合併と自治体の能力

・合併自治体の方が国にモノが言い易かった

「各自治体から国に対する声・要望は届きやすくなっているか」という問いに対し、「届きやすくなっている」と「届きにくくなっている」の回答状況は表9のようになった。

表9　合併の有無と国への声の届きやすさ

	届きやすくなっている	届きにくくなっている	合計
非合併	53.5	46.6	100.0
	78	68	146
合　併	60.5	39.5	100.0
	27	17	44
合　計	55.3	44.7	100.0
	105	85	190

(注) 上段は行の中での割合、下段は市町村数。
(出所) アンケート結果を基に、筆者作成。

地方創生の全体的傾向として、各自治体は国に対する声・要望が届きやすくなったと認識していると結論した（村上2018b)。この傾向は、わずかながら合併自治体においてより強く表れている。

これはまず、そもそも地方創生、ひいては国主導の政策に対する合併・非合併自治体の認識の差を表している可能性がある。そういった認識が定着していくと、「平成の大合併」という国の比較的大きな誘導策にも乗らなかった自治体と国との距離は、（その良し悪しはともかく）ますます広がっていく恐れがある。

あるいは、「声が届きやすくなっている」というのが実態としてもそうであるとすれば、合併自治体が国に対する交渉力を得ているという理解が可能である。この観点からも、それが行き着くのは、「非合併自治体と国との隔たりが今後ますます大きくなる」という結末である。

・合併が職員の能力を高めたとは言い切れない

「「人口ビジョン」、「総合戦略」の策定に当たって、コンサルタント等を利用した（利用しなかった）理由は何か」という問いに対し、「自治体職員の専門的知識」と「それ以外」の回答状況は表10のようになった。

表10　合併の有無と自治体職員の専門知識

	専門知識	それ以外	合計
非合併	32.9	67.1	100.0
	48	98	146
合　併	56.8	43.2	100.0
	25	19	44
合　計	38.4	61.6	100.0
	73	117	190

(注) 上段は行の中での割合、下段は市町村数。
(出所) アンケート結果を基に、筆者作成。

「それ以外」には「国からの交付金があったから」や「自治体のマンパワーを活用できたから」などが含まれる。それらと「自治体職員の専門的知識」と対比した場合に有意な差が見られたため、上記のように整理した。

ここから推察されるのは、職員の専門知識が不足しているとの認識が、合併自治体でより強いということである。もちろん、合併自治体において専門知識がより必要だったということもあり得る。逆に非合併自治体は、地方創生に専門知識が必要だということを理由にコンサルタントを利用しようとはあまり考えなかった（実際、地方創生対応を職員の自助努力にかなり依存した小規模自治体もあった)。上記の推察が正しいとすれば、非合併自治体がこう考えるとし

ても不思議はない。ただしこれは、合併すれば自治体職員の専門性が高まるだろうという私たちの直感に必ずしも沿うものではない。これは、合併・非合併自治体にも、例えば、吸収合併か対等合併かや、吸収する側かされる側かといった、様々な自治体があるということを反映しているのかもしれない。

・合併自治体の積極的な姿勢が目立つ

「あなたやあなたの自治体にとって、今般の地方創生策は一言で言えば何か」という問いに対し、「地方創生のチャンス」、「学ぶところがあった」、「その他」の回答状況は表11のようになった。

「その他」には、「事務作業に忙殺された」、「住民にとって有益だった」、「自治体の政策や方針が改善した」などが含まれる。それらをまとめた上で、「地方創生のチャンス」や「学ぶところがあった」と対比した場合に有意な差が見られたため、表11のように整理した。

表11　合併の有無と地方創生の意義

	チャンス	学ぶところ	その他	合計
非合併	33.6	28.8	37.6	100.0
	49	42	55	146
合　併	40.9	20.5	38.6	100.0
	18	9	17	44
合　計	35.3	26.8	37.9	100.0
	67	51	72	190

（注）上段は行の中での割合、下段は市町村数。
（出所）アンケート結果を基に、筆者作成。

結果は、「地方創生のチャンス」と回答した自治体がやや多かったが、合併自治体においてその割合が高く、非合併自治体に関してはそれと「学ぶところがあった」との差が小さかった。これは、「平成の大合併」における国の誘導策を彷彿とさせる、合併自治体の国主導の政策への向き合い方を示唆している可能性がある。すなわち、国策をチャンスと捉える合併自治体に対し、それを幾分冷めた目で見る非合併自治体の姿が現れている。

（2）市町村合併と対外関係

・合併自治体は周囲との調整を要しない傾向

「総合戦略の策定に当たって、周辺地域の自治体との連携、調整はあったか」という問いに対し、「特に連携・調整の機会はなく、独自に進めた」と「その他」の回答状況は表12のようになった。「その他」は、他自治体の担当者個人との間、あるいは組織との間の何らかの連携・調整があったというものである。

表12　合併の有無と周辺との連携の有無

	独自で	その他	合計
非合併	39.7	60.3	100.0
	58	88	146
合　併	45.5	54.5	100.0
	20	24	44
合　計	41.1	58.9	100.0
	78	112	190

（注）上段は行の中での割合、下段は市町村数。
（出所）アンケート結果を基に、筆者作成。

この結果によると、合併・非合併いずれも独自に進めたところは多くないが、個人や組織レベルでの連携・調整が盛んに行われたのはむしろ非合併自治体の方だった。この点は、市町村が合併するという形を採らずとも、担当者個人や組織レベルでの機能的な連携・調整がそれなりにあり得ることを示唆している。実態としては、自治体間の連携・調整の必要があるにもかかわらず何らかの理由で合併できなかった（もしくはそのコストがより大きくなると判断した）場合に、担当者個人や組織レベルでの機能的連携・調整で対応するということだろう（後述の通り、フランスでもこうした対応が多いとみられる）。

なお、「その他」の中で「北海道庁の仲介で調整を行った」の回答がパーセンテージにおいて非合併は合併の2倍だった（村上・小磯・関口2017）ことも、それを示唆しているのかもしれない。自治体間の機能的連携に、広域自治体（都道府県）による仲介が功を奏する可能性が高い。

・合併自治体は道県外との繋がりを求める

「人口ビジョン」、「総合戦略」の策定に当たっ

てのコンサルタントへの業務委託先について、「道県内コンサルタント」、「道県外コンサルタント」、「委託せず」の回答状況は表13のようになった。

表13　合併の有無と委託コンサルタント

	道県内	道県外	委託せず	合計
非合併	43.8	37.7	18.5	100.0
	64	55	27	146
合　併	34.1	45.5	20.5	100.0
	15	20	9	44
合　計	41.6	39.5	18.9	100.0
	79	75	36	190

（注）上段は行の中での割合、下段は市町村数。
（出所）アンケート結果を基に、筆者作成。

非合併自治体は同じ道県内に、合併自治体は道県外に、それぞれ委託しがちだったことが明らかになった。非合併自治体は、割と道県内向きという意味でドメスティックである可能性がある。逆に、自治体が合併したことによって道県外に飛び出していく力を得たとすれば、私たちの直感にも合うところがある。

コンサルタントは都市圏、さらには関東・関西圏に多く集積していると考えられる。そのため、総じて合併自治体は関東・関西圏のコンサルタントに委託しがちだったと言える。なお、国から配分されたコンサルタント委託費がこのような形で都市圏に「逆流」したことは、地方創生の政策を評価する上で、省みられるべきかもしれない。

・非合併自治体は国策から影響を受けにくい

「総合戦略」策定のための庁内体制を「総合計画」のそれと比較した場合、「拡充した」、「変わらず」、「縮小した」の回答状況は表14の通り。

表14　合併の有無と地方創生への庁内体制

	拡充	変わらず	縮小	合計
非合併	18.5	69.2	12.3	100.0
	27	101	18	146
合　併	20.5	59.1	20.5	100.0
	9	26	9	44
合　計	18.9	66.8	14.2	100.0
	36	127	27	190

（注）上段は行の中での割合、下段は市町村数。
（出所）アンケート結果を基に、筆者作成。

全体的には「拡充」が20％程度で、「変わらず」が最も多いという結果である。ところが、非合併自治体に「変わらず」が多いのに対し、「縮小した」とする合併自治体の割合がやや高いことをどう捉えるべきか。

1つには、概して領域（面積）がより広い（であろう）合併自治体では、それなりに大きくなり得る「総合計画」への注力に比して、急を要した地方創生への注力は小さくなったところがあったことが推測される。逆に非合併自治体は、この点でも国主導の政策から影響を受けることがあまりなかったと捉えることができる。

(3) 市町村合併と民主主義

・非合併自治体は近隣より域内産業界と組む

「あなたの地域において、通常の自治体の政策決定に対して最も大きな影響力を及ぼす人の属性は」という問いに対し、「産業界」、「官界」、「その他（学界・金融機関・労働組合・マスメディア）」の回答状況は表15のようになった。

この設問は、地方創生にとどまらず通常時の地域内権力関係を問うものであるが、非合併自治体では産業界が、合併自治体では官界（役所）が、それぞれ影響力を持ちがちであることが分かる。

表15　合併の有無と地域内権力関係

	産	官	その他	合計
非合併	50.7	34.9	14.4	100.0
	74	51	21	146
合　併	31.8	47.7	20.5	100.0
	14	21	9	44
合　計	46.3	37.9	15.8	100.0
	88	72	30	190

（注）上段は行の中での割合、下段は市町村数。
（出所）アンケート結果を基に、筆者作成。

市町村合併は複数の役所（市町村役場）を1か所に集約し機能強化するものであるため、合併自治体で官界が影響力を持つようになるという結果は自然である。また、非合併自治体においては、企業版ふるさと納税の項で考察したよう

に、産業界と手を携えて統治を行っている場合が多いと捉えることができる（ただし、「産業界」が具体的に何を指すかには注意する必要がある）。あるいは、非合併自治体は域内の産業界が強いから合併せず、合併自治体は産業界の領域的な意味における桎梏が小さかったから合併できたという解釈ができるかもしれない（もっとも、市町村の境界を超えて活動する産業界の要望で合併が進み、合併に伴って役所が合理化されて力を失うという因果関係もあり得ないではない）。いずれにしても、合併・非合併と地域内権力関係（とりわけ産業界との関係）とは、何らかのロジックで繋がっているようである。

・合併自治体は役所で完結する

「総合戦略」策定のための住民参加を「総合計画」の場合と比較して、「拡大・変わらず」と「縮小」の回答状況は表16のようになった。

表16　合併の有無と住民参加の度合い

	拡大・不変	縮小	合計
非合併	76.3	23.7	100.0
	111	35	146
合　併	36.4	63.6	100.0
	16	28	44
合　計	66.8	33.2	100.0
	127	63	190

（注）上段は行の中での割合、下段は市町村数。
（出所）アンケート結果を基に、筆者作成。

非合併自治体には「拡大・不変」が多く、合併自治体には「縮小」が多い（全体的には「拡大・不変」がダブルスコア）という結果になった。

前項で考察した「合併自治体では官界（役所）が強い」という点と併せて考えると、合併自治体はあまり住民を顧みないという解釈も可能である。確かに地方創生では矢継ぎ早に「人口ビジョン」と「総合戦略」の策定が求められ、役所（市町村役場）に十分な対応リソース（時間・人員など）を与えられなかったこともあって、とりわけ領域（面積）の広い合併自治体では住民参加を縮小せざるを得なかった、と理解できないではない。とはいえ、市町村合併により役所（市町村役場）と住民との距離が遠くなるという懸念も指摘される中で、地方創生でその問題が顕在化した可能性がある。他方で、非合併自治体においては、やはり上記の通り、民間と手を携えた統治をしているという意味において、本稿の解釈と大きく矛盾しない結果が出ている。すなわち、合併自治体に比べれば非合併自治体で、住民参加の程度が大きかった。

4. フランスの自治体の合併と連携

本章では、筆者が2018年9月から在外研究をしているフランスの国策としての自治体合併（fusion、groupement）等の進捗を概観し、前章までに分析した日本の状況を踏まえて、今後いかなる視角で日仏比較が可能かを整理しておきたい。

（1）経緯

フランスでは、36,000を超えるコミューン（基礎自治体）に、大小に関わらず同一の法的地位が与えられている。

前掲表2によれば、日本の人口5,000人未満の町村数は0.6％だが、フランスでは人口1,000人未満のものが全体の76％を占め、国民の約半分が人口10,000人未満のコミューンに住んでいる（表17）。これが日仏比較の最大のポイントと言える。日本だと市は100,000人まで、町村は10,000人から30,000人までが平均的だが、フランスだと5,000人前後のところが「最頻」となっている。

表17　フランスの人口規模別コミューン数・人口（1999年）

	コミューン		人　口	
	数	%	数	%
～1,000 人	27,794	76.0	9,393,701	15.7
～5,000 人	6,922	18.9	14,298,005	23.9
～10,000 人	975	2.7	6,729,792	11.2
～30,000 人	633	1.7	10,647,377	17.8
～100,000 人	205	0.6	9,844,714	16.4
～300,000 人	31	0.1	4,885,416	8.1
300,000 人～	5	0.01	4,152,430	6.9
合　計	36,565	100.0	59,951,435	100.0

（出所）Pinson（2010、p.70）。

表18　1コミューン当たりの平均人口・面積の比較

	フランス	日　本
全人口(a(人))	66,990,000	126,230,000
全面積(b(km^2))	551,500	377,972
自治体数(c)	36,565	1,718
平均人口(a/c)	1,832.1	73,475.0
平均面積(b/c)	15.1	220.0
人口密度(a/b)	121.3	334.0

（出所）総務省･外務省ホームページのデータを基に､筆者作成。

1基礎自治体（コミューン）当たりの平均人口・面積の比較は、表18の通りである。1自治体当たりの面積は日本が約14.6倍、人口は約40.1倍であり、この点は制度・運用比較の際に留意しておくべきである。そのフランスでも、事務の効率的処理のために、政府が主導して自治体合併を推進してきた。しかし、長い歴史の中で見れば、その成果はあまり表れていない。

フランスでは、1837年以来の単一目的事務組合が、1890年3月22日法により法人格を得て制度化した。1959年1月5日オルドナンスで、多目的事務組合も規定されるようになるとともに、消防と住宅に関して義務的権限を持つ広域コミューン区の創設が可能になったことにより、コミューン間広域行政が発展した（財団法人自治体国際化協会2009. p.40）。

1959年7月22日デクレでは350件の合併があり、746コミューンが統合したが、実際には上記のコミューン間の広域行政組織が大いに活用されている。コミューン・レベルの広域行政組織としては、「組合型」（2008年1月現在、単一目的11,739、多目的1,451）と、広域行政組織自体が独自の税源を持ち法令で規定された一定の必須的権限を持つ「連合型」がある（財団法人自治体国際化協会2009、p.38）。共和国の地方行政に関する1992年2月6日法は、多様な形態のコミューン間広域行政を奨励し、コミューン共同体（2008年1月現在、2,393）、広域都市共同体、コミューン間広域行政の促進を図る提言を行う県委員会の設置を決めた。その結果、1,500近いコミューン共同体が創設された（同制度はその後簡素化）。コミューン間の相互協力の促進と簡素化に関する1999年7月12日法（「シュベヌマン法」）は、固有税源を持つ広域行政組織を再編した。その後、大都市共同体、新都市組合、都市圏共同体などにより運用された（財団法人自治体国際化協会2009、pp.12-13、pp.38-39）。

コミューンの合併と広域化に関する1971年7月16日法（「マルスラン法」）は、各県に合併促進協議会を設ける手続の整備や合併コミューンへの特定補助金50％増といった財政的優遇措置を設けて合併を促そうとした（財団法人自治体国際化協会2009、pp.37-38）。これによりコミューン数が劇的に減少することはなかったが、1960年代に整備・拡充された「独自税源を有するコミューン間協力型広域行政組織（EPCI）」の枠組みを活用した新コミューン制度が2010年に設けられ、さらに2015年法（「ペリサール法」）で合併コミューンの議員身分保障や財政的優遇措置がなされると、コミューン数が1,243減るという成果を出した（中田2019）。それでも、フランスのコミューン数は欧州連合全体のそれの3分の1を占める[1]。

（2）現況

フランス北部・ソンム県の2コミューンの合併（2019年1月）は、いずれのコミューンも今後単独では存続できないと考えたことによって実現したようである。同時期に首都パリに近いウール・エ・ロワール県内の合併で誕生した新コミューンでも、やはり財政と人的資源面の強化が課題とされた[2]。フランス南東部・オーヴェルニュ・ローヌ・アルプ州のオート・サボア県では、当事者が教育など域内行財政面のハーモナイゼーション（調和）の困難を感じながらではあるものの、最小人口4,258人、最大人口54,087人の6コミューンの合併が実現した[3]。

とはいえ、全体的には「フランス革命以来大きな変化は見られない」と言われる。フランス南西部・ヌーベル・アキテーヌ州のシャラント県で2018年11月、4コミューンの合併話が「破談」になり、その不透明なプロセスに抗議した地方議員が辞職するという事件もあった[4]。2018年12月11日法により、各県に域内コミューンの合併を進める圧力がかかるようになったが、特に小コミューンの反対により合併は進んでいない。

(3) 個別事例

ヌーベル・アキテーヌ州ロット・エ・ガロンヌ県のあるコミューン（人口1,500人）の首長は、3キロメートルの範囲内に収まる周囲のコミューンを巻き込み、人口2,500人の新コミューンを立ち上げようと試みたが、やはり合意が得られなかった。彼曰く、「彼を支持する28のアソシアシオン（非営利団体）の利益が、域外に流れていっている」。「他のコミューンから来た150人の子供が、域内のサッカーチームに参加している」。こうしてコミューンの境界を越えた負担と便益が存在しているため、負担に対してより大きな便益を得ているコミューンからすれば、現状を変えようということにはならない。

さらにフランスでは、こうした財政問題のみならず、当事者のエゴも合併の障害になっている。すなわち、各首長は「合併後の人口2,000人の新コミューンの一役人としてではなく、せいぜい人口130人のコミューンの「大将」であり続けたいと思っている」。フランスで合併を進めるには、この部分を突破していかなければならない。もっとも、これは基本的には日仏問わず、普遍的な問題だろう。

フランスでは前述の通り、複数コミューンによる（合併を伴わない）広域行政組織化がすでに進んでおり、当事者も、合併を急がなければならないとは必ずしも認識していない[5]。合併に反対する人々は、合併により非効率な複層的行政（「ミルフィーユ」）が出現する恐れがあり、個々のコミューンのアイデンティティや役所と住民との近接性、自然災害対応などの機動性を失う可能性があると、考えているようである。

5. おわりに

(1) 市町村合併・連携の理論

本稿では、2016年から2017年にかけて実施した地方創生アンケートの結果を、「平成の大合併」における合併・非合併自治体の異同という観点で分析した。その際、実地調査の成果を踏まえ、フランスとの比較も織り交ぜながら、「平成の大合併」と地方創生における市町村合併・連携という政策手段を評価し、それによりその条件や効果、自治体の思考様式に関する理論構築を試みた。

クロス表を用いた分析の結果を再掲すると、次の通りである。

・非合併自治体への連携促進効果は乏しかった
・非合併自治体に地方創生はあまり響かず
・非合併自治体の方がふるさと納税に対し貪欲
・合併自治体の方が国にモノが言い易かった
・合併が職員の能力を高めたとは言い切れない
・合併自治体の積極的な姿勢が目立つ
・合併自治体は周囲との調整を要しない傾向
・合併自治体は道県外との繋がりを求める
・非合併自治体は国策から影響を受けにくい
・非合併自治体は近隣より域内産業界と組む
・合併自治体の意思決定は役所で完結する

これらを整理し直すと、次のようになる。

地方創生は、非合併自治体の心にはあまり響かず、それへの広域連携促進効果は乏しかったが、非合併自治体は企業版ふるさと納税に対して割と貪欲であるなど、近隣自治体政府とよりもむしろ域内外の産業界との連携・調整を志向しがちであることが明らかになった。他方合併自治体は、合併によって自治体職員の専門能力が向上したとは言い切れず、また意思決定が役所内で完結しがちであることから住民参加（民主性）の度合いが低い可能性がある。とはいえ非合併自治体との比較では、合併したがゆえに他自治体との連携・調整なしに独自で、域外（同道県外）、さらには国に対して積極的な行動を起こすことができる。

(2) 本稿の「評価」をめぐって

本稿では、自治体をあえて機械的に合併・非合併に区別して捉えたことから、多くの要素を捨象せざるを得なかった。つまり、市町村合併・連携を評価するに当たり、大きな限界を抱えている。

第1に、合併自治体は多様である。例えば、大都市に隣接する町や村が大都市に吸収合併され

る場合、隣接する町や村がほぼ対等な立場で合併（連携）する場合、あるいは、吸収合併で吸収する自治体か吸収される自治体かは、本稿の評価では区別せずにカウントしている。また、同じ合併や連携でも、それに至る手続の違い（政治プロセスによるか、行政プロセスによるかなど）も、クロス表では一切加味できていない。確かにこれらを区別し始めると作業が膨大になり得るが、これらが自治体の属性や能力、対外関係にとって決定的である可能性は充分にある。その点、本稿で取り上げた比較的対等に近いものから大都市共同体・メトロポールに至るまで、多様な手続に基づく多様な連携形態が用意されているフランスの事例は、カテゴリー整理に当たっても有用である。

第2に、本稿で用いた地方創生アンケート結果は、全50問中の5分の1程度に過ぎず、その選別にかなり恣意的なところがないとも言い切れない。もちろん、本稿では全問についてクロス分析をした上で、合併・非合併に有意な差が見られたものを取り上げた。とはいえ、サンプル数が増えれば取り上げるべき問いは今より増えるかもしれない。極論すれば、合併・非合併で有意な差が現れなかったものについてその原因・理由を探ることの方が、より有意義かもしれない。これについてはフランス調査も含め今後の課題とする。

第3に、本稿のような評価に際して、国際比較をどう活用するかという問題がある。本稿では、日仏の自治体行政の前提条件の違い、政府による合併誘導策の推移、合併が進まない要因の共通性などを整理した。フランスでも同様のアンケートが実施できるならば、評価に当たっての国際比較の有用性を今以上に示すことができよう。本稿からは、日仏比較では合併・連携自治体への権限・税財源配分のあり方や、合併によって生じる複層的行政（「ミルフィーユ」）への住民の懸念（地方自治観）とそれをいかに克服するかが重要なファクターとなることが窺われる。住民の利益や地方自治の観点からは、市町村が合併してしまうことよりも、多様な連携形態を模索しその手続を整えていく方が現実的であるようにも思われる。

(3) 今後の課題

今後の課題としては、評価に関する上記のポイントに留意しながら、フランスでも同様のアンケートを実施するなどして、本稿の結論の真偽をより多くのサンプルからマクロに検証すること。それとともに、個別事例を追跡してミクロにも検証することにより、「市町村合併・連携の理論」を精緻化するということがある。その際、実態を規定している因果関係をアンケート結果から明らかにする分析手法を、洗練していく必要もある。

その上で、合併から機能的な自治体間連携に至る多様な選択肢の中から、地方行財政基盤の強化に向けて、また、住民にとっての様々な価値尺度を踏まえて、欧州連合のような地域連合、国、州・都道府県、市町村の中での権限配分のあり方を考慮に入れつつ、また国際比較もしつつ最適解を導き出す。まさに表1に言う「結果」の政策評価にもなるこれが、究極的な課題だと考えている。

謝辞

本稿は、2018 ～ 2020年度 科学研究費補助金「日・仏・欧比較による多層的規制モデルの構築」、2020年度 同「マルチレベル・ガバナンスの研究：人や組織のdiscretionによる事例分析」、及び、2017 ～ 2019年度 同「規制機関の多層性と規制の公益性」の成果の一部をまとめたものである。本研究にご指導・ご協力くださった査読者の先生方、及び、Sciences Po BordeauxのProfessor Gilles Pinson、Professor Andy Smith、Professor Olivier Costa、Professor Xavier Itçainaをはじめとする皆様に、深く御礼申し上げる。

注記

1. SudOuest ‘Fusion de communes en Lot-et-Garonne: pourquoi ça coince’（https://www.sudouest.fr/2019/01/09/y-a-t-il-trop-de-communes-en-lot-et-garonne-5720525-3603.php）.
2. France 2 ‘Avec la multiplication des fusions, la France compte moins de communes’（https://www.francetvinfo.

fr/france/auvergne-rhone-alpes/drome/avec-la-multiplication-des-fusions-la-france-compte-moins-de-communes_3126303.html).

3. Le Monde ‘Fusion des communes: à Annecy,un parcours du combattant mais un pas de géant pour la ville’ (https://www.lemonde.fr/politique/article/2019/01/05/a-annecy-un-parcours-du-combattant-mais-un-pas-de-geant-pour-la-ville_5405405_823448.html).
4. SudOuest ‘Pourquoi quatre élus ont démissionné’ (https://www.sudouest.fr/2018/11/23/pourquoi-quatre-elus-ont-demissionne-5591689-815.php).
5. 前掲（注1）記事。
※ウェブへの最終アクセスは2020年4月6日。

参考文献

国立社会保障・人口問題研究所（2020）「人口統計資料集2020」、『人口問題研究資料（第342号）』

財団法人自治体国際化協会（2009）『フランスの地方自治』（http://www.clair.or.jp/j/forum/series/pdf/j30.pdf）

嶋拓哉・児矢野マリ・野田耕志・村上裕一・津田智成・樽本英樹・落合研一（2017）「道内のインバウンド現象を巡る法的問題の把握とその実証研究」、『（一財）北海道開発協会開発調査総合研究所 研究助成の実績と研究成果』、125-149

総務省資料「未合併要因に関する調査」（http://www.soumu.go.jp/main_sosiki/singi/chihou_seido/singi/pdf/No29_senmon_2_s6.pdf）

曽我謙悟（2019）『日本の地方政府：1700自治体の実態と課題』、中公新書

中田晋自（2019）「フランスの地方自治体改革（2010年）における新しい市町村合併政策：「新コミューン（commune nouvelle）制度」の創設とその現況」、『愛知県立大学外国語学部紀要（第51号）地域研究・国際学編』、63-87

村上裕一（2018a）「政策現場と内閣主導：「地方創生」を通して見るそれらの関係」、『季刊 行政管理研究（第161号）』、4-18

村上裕一（2018b）「自治体担当者は地方創生をどう受け止めたか」、小磯修二・村上裕一・山崎幹根編著『地方創生を超えて：これからの地域政策』、岩波書店、35-64

村上裕一・小磯修二・関口麻奈美（2017）「「地方創生」は北海道に何をもたらしたか：道内自治体調査の結果とその分析を通して」、『年報 公共政策学（第11号）』、119-137

村上裕一・小磯修二・関口麻奈美（2018）「「地方創生」は地方に何をもたらしたか：愛媛県・香川県内自治体調査の基礎集計と予備的考察」、『年報 公共政策学（第12号）』、49-72

山崎幹根（2018）「自治政策・国土政策から見た国と地方：広域連携から考える」、小磯修二・村上裕一・山崎幹根編著『地方創生を超えて：これからの地域政策』、岩波書店、97-123

Murakami, Y.(2019). Improving the current administration's Local Revitalisation Policy: Promoting a recently co-authored publication. *Annals, Public Policy Studies*, [13], 99-116.

Pinson, G.(2010). France. In Goldsmith, M.J., & Page, E.C.(eds.), *Changing government relations in Europe: From localism to intergovernmentalism*. Routledge, 68-87.

（2020.6.4 受理）

Evaluation of Municipality Mergers and Cooperation as Policy Tools Based on the Survey Results Regarding the Local Revitalisation Policy in Japan

Yuichi Murakami

Hokkaido University
yuichim@juris.hokudai.ac.jp

Abstract

This paper analyses survey results on the Local Revitalisation Policy（LRP）enacted in Japan in 2014, by focusing on differences in answers provided by the merger and non-merger municipalities of the governmental campaign that took place around 2000（*Heisei-no-Dai-Gappei*）. To evaluate municipality mergers as a policy tool, a case study on Japan was taken into consideration and compared with（non-）merger cases in France, aiming to construct the general theory of its conditions and effects. It was found that ‘non-mergers’ were not strongly influenced by LRP and did not feel obliged to cooperate with neighbours as the LRP desired. On the other hand, they were likely to cooperate with industries, according to their actual and potential partnerships. It cannot be determined whether ‘mergers’ have developed their own competencies as was desired, although they seem to have decreased the extent of democratic participation because their decision-making takes place inside their office. However, they can take relatively more aggressive actions towards other local and central governments without onerous coordination with neighbours.

Keywords

The Local Revitalisation Policy, municipality merger, inter-communal cooperation, policy evaluation, France

【実践・調査報告】

参加型・質的評価手法 MSC（Most Significant Change）の国内３事例を考察する －学習・改善目的中心の評価－

田中　博
一般社団法人参加型評価センター
tanaka.pecenter@gmail.com

要　約

MSC（Most Significant Change）は、欧米のNGOが活用する参加型・質的評価手法であり、学習・改善目的の評価に適しているが、日本国内での活用例は少ない。本稿では、震災復興、教育、福祉と国内社会セクターにおいて実施されたMSC評価3事例を比較検討し、国内でMSCを有効活用するための参考情報を提示する。事例研究の結果、トレーナー、コーチとしての評価専門家の存在、評価目的を手法の特色に合わせて設定する、現場視点のデータ収集、データ分析の妥当性を高める工夫の必要性など、活用上の重要項目が明らかになった。また比較的容易に内在化が可能であるという、他の参加型評価手法との違いが認められた。今後国内社会セクター事業でのMSC活用が期待されるが、指導する専門家の不足や、適切なフィードバックのあり方を模索するべきという課題がある。

キーワード

参加型評価、質的評価、学習・改善、社会セクター、NGO/NPO

1．はじめに

Most Significant Change（MSC）は、欧米の国際協力NGOが多用している、参加型・質的（定性的）評価手法である。評価目的において、説明責任よりも学習に焦点が当てられており、プログラムの改善を促進するといわれる（Davies & Dart 2005）。

評価を行う目的は、大きく分けて学習と説明責任の2つである（三好 2008）。学習とは、評価を通じて得られた教訓を将来の活動に生かしていくことである。これは改善とつながっているため、本稿では両者をまとめて「学習・改善」目的の評価と呼ぶ。そして説明責任は評価の結果を、利害関係者に報告することである。説明責任も、組織上部や資金提供者への報告である「上向き（Upward）説明責任」と、受益者への報告である「下向き（Downward）説明責任」に大別される（Humentum 2019）。下向き説明責任は、評価で得られた教訓や提言を、評価結果の利用者である受益者に還元・共有する意味で、学習効果があるといえる。

筆者はMSCが国際協力のみならず、NPO[1]や行政が日本国内で実施する、教育や福祉をはじめとした社会セクター事業への学習・改善目的の評価に有効であると考えている。社会課題の解

日本評価学会『日本評価研究』第20巻第2号、2020年、pp.121-136

決を目指す社会セクター事業にとって、プログラム目標を効果的・効率的に達成するためには学習・改善目的の評価は重要である。しかしながら、MSCの国内事業への活用はまだまだ少ない。またどのような手法でも、実施の際は基本的な手順を追いながらも、個々のプログラムの性格に合わせ詳細を柔軟に応用する必要がある。

本稿では、国内事業の評価にMSCを適用し、筆者が外部専門家として関与した、学習・改善目的を中心とした震災復興、教育、福祉と性質の異なる3事例の評価プロセスを報告する。それぞれの事例では、MSCの基本的進め方に沿いながらも、細部にわたっては柔軟に調整が行われた。本稿の目的は、これから主に学習・改善目的の評価で、MSCを活用する、もしくは活用を始めたが経験の少ない実践者や研究者を対象に、上記3事例でどのようにMSC手法を現場で活用したか実例で示すことで、より効果的かつ効率的にMSCを実施できるようになるための参考情報を提供することである。

本稿の構成について述べる。2章では、MSC手法の概要、実施方法、4つの特色について述べる。3章では、事例研究の分析の枠組みを説明する。4章では、MSCを用いた国内事業の評価プロセスの詳細を3事例紹介する。5章では、事例研究を踏まえた考察を行い、MSCを国内事業で有効に活用するための要点をまとめ、今後の展望と課題を述べる。

本稿で用いる用語を説明する。社会課題を解決するための、開始時点と終結時点のある介入をプログラムと事業を同義語で用いる。また物語とストーリーを、質的分析でいうところのNarrative、vignettes、anecdotesの概念として、出典での記述に従い、同義語で用いる。

2. MSC（Most Significant Change）とは何か

（1）手法の概要

MSCは、リック・デイビーズ博士が1990年代にバングラデシュのNGO、CCDB（Christian Commission for Development in Bangladesh）の事業のために考案した、参加型モニタリング・評価（M/E）手法であり、質的評価手法でもある。活動現場で起こった「重大な変化」を物語の形で複数収集し、それらの物語から利害関係者（ステークホルダー）の参加で「最も重大な変化」を1遍選択することが基本的な進め方である。そのプロセスを通じて、質的な変化、想定外の変化の詳細を把握したり、プログラムを改善するための教訓を学習したりできる。また参加を通じてプログラム・スタッフや受益者など利害関係者の相互理解が進み、エンパワーメントされる手法である。これまでMSCは、欧米のNGOを中心に開発途上国の国際協力プログラムのM/Eに使われてきたが（Davies & Dart 2005）、日本国内の実践事例はまだ少ない[2]。

（2）MSCの基本的な実施方法

MSCの基本的な手順は、MSCのガイドブックである「Most Significant Change（MSC）Technique: A Guide to Its Use」（Davies & Dart 2005）[3]に詳しく示された10のステップを繰り返し実施することである（表1）。

表1　MSCの基本10ステップ

1. 導入と注意喚起
2. 変化の領域を決める
3. 時間の範囲を決める
4. 重大な変化の物語を集める
5. 最も重大な変化の物語を選ぶ
6. 選択過程をフィードバックする
7. 物語が事実であるか検証（事実確認）
8. 定量化
9. 二次分析及びメタ・モニタリング
10. システムの改定

（出所）Davies & Dart 2005

しかしながら、この10のステップは必ずしも必要ではない項目も含まれており、デイビース自身「最も重要なステップは4、5、6である」と述べている。筆者もMSCの理解と実施にはこの3ステップで十分と考えている。3ステップとはステップ4「重大な変化の物語を集める」、ステップ5「最も重大な変化の物語を選ぶ」とステップ

6「選択過程をフィードバックする」である。以下、順番に解説したい。

① 重大な変化の物語を集める～データ収集

MSCの基本設問の設計と、プログラム・スタッフなどが受益者などを情報提供者として、変化の物語を集める質的データ収集の段階である。基本は以下1～5まで1文で連なるオープン・クエスチョンを行うことである。括弧内の例を参照のこと（Davies & Dart 2005）。

1. 特定の期間を限定する（例：先月を思い出して）
2. 相手自身の意見を聞く（例：あなた＝情報提供者の考えでは）
3. 変化の起こる範囲である「領域」を絞る（例：人々の生活の質に関して）
4. 物語を1つに絞る（例：最も重大な）
5. 漠然とした状況ではなく、変化を聞く（例：変化は何ですか）

まず、「過去6カ月の間に」など、変化を調べる期間（＝時間の範囲）を指定する。また変化の起こる範囲・分野である「領域」を事前に決めておく。例えば、「A村の女性の意識について」という感覚で、知りたい情報に合わせて自由に変えて良いとされる。そして、情報提供者の考えるいくつかの変化の中から、重大な変化を1つ選んでもらう。この場合の「重大（Significant）」とは、「意味深い」「意義がある」という意味に捉える。また、漠然とした状況ではなく具体的な「変化」の詳細を物語の形で収集する。この設問に続いて、「なぜ、その変化が重要だと思うか」という、情報提供者にとっての重要性の理由を問う設問が続く。

物語の集め方であるが、いくつかの選択肢がある。プログラムのスタッフが、現場で受益者に直接インタビューする、グループ討論を通じて集める、受益者自身に文章を書いてもらうなどから、事業や組織の実情に合わせて決める。受益者から直接収集できない場合は、受益者に近く、彼らと共に行動している人物から集めても良い、とされている（Davies & Dart 2005）。

② 最も重大な変化の物語を選ぶ～データ分析

複数の重大な変化の物語とそう思う理由が揃ったら、プログラム・スタッフなど利害関係者のグループで、物語の選択作業をとおしての質的データ分析を行う。基本は、選択者の裁量に従って以下の討論を行うことである（Davies & Dart 2005）。領域が複数ある場合、領域ごとに選択を行う。

1. グループ皆で物語を読み込む。
2. グループで、どの物語を選ぶのか、深く議論する。
3. どの物語が最も重大に感じたかを決める。
4. それを選んだ理由を書き留める。

通常、組織やプログラムには、活動現場から意志決定者まで、いくつかの階層がある。下位のレベルで選択された物語は、上位の階層に提出、同様に選択される（図1）。各階層の選択を終える毎に、選んだ理由が記録され、他の利害関係者に共有される。即ち、上位の選択過程は下位の選択過程における判断理由を踏まえて実施されることになる。選択を繰り返した後、最上位の階層の「最も重大な変化」と選択理由を、領域ごとの文書にまとめる（Davies & Dart 2005）。

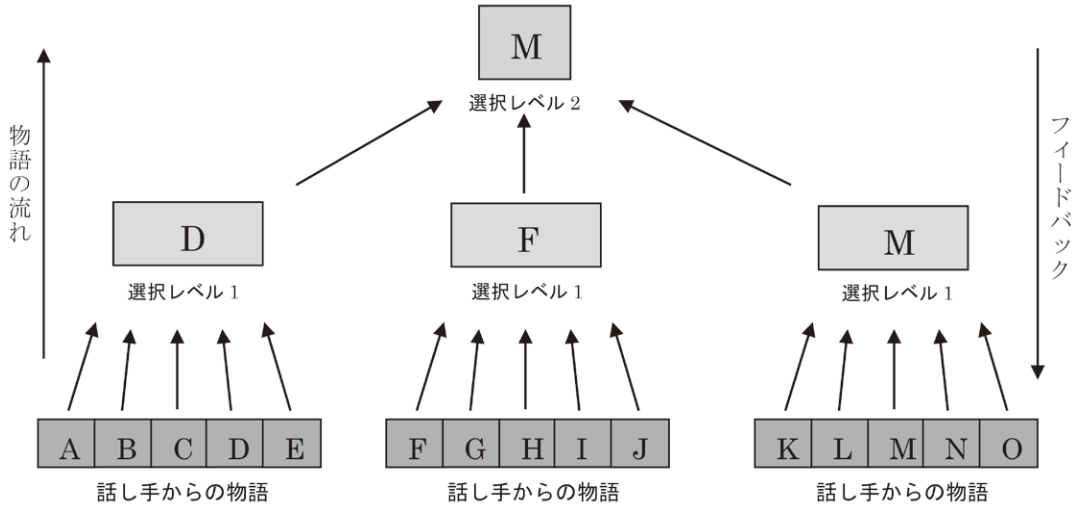
③ 選択過程をフィードバックする

全体の「最も重大な変化」が1つ決まったら、選択結果である「最も重大な変化」の物語と、それを「選んだ理由」を、物語を集めた現場の情報提供者や、選択を行った各階層の関係者に文書や口頭で伝える。デイビースら（2005）によると、このフィードバックは選択過程で導き出された評価結果（教訓や提言）を受益者などと共有し、利害関係者間の対話を深め改善につなげる「下向き説明責任」の効果があるとして重要視している。

（3）MSCの4つの特色

筆者はデイビースらの文献を参考に、バングラデシュでMSCを実践して手法の4つの特色を以下のようにまとめた（田中 2014）。

図1　選択とフィードバックの流れ



（出所）　Davies & Dart 2005：29

① モニタリング手法及び評価手法である

MSCは基本ステップを定期的に繰り返し実施する中で、モニタリングと同様にプログラム管理のための進捗状況を提供する。さらに、「重大な変化」を集め、「最も重大な変化」を選ぶ過程では、アウトカムとインパクトに着目するため、評価の性質も持っている（Davies & Dart 2005）。

② 参加型評価手法としてのMSC

参加型評価は、評価専門家が行う従来型評価と異なり、プログラム・スタッフや受益者など、評価対象の利害関係者が評価プロセスに参加する。それが学習過程として機能し、彼等のプログラムに対する当事者意識の促進、行動の変容、評価能力の向上など、能力開発に貢献する。これを通じて評価結果の活用度合いが高まり、プログラムの改善に貢献する（源 2016、2008；三好・田中2001;JICA 2001）。源（2016）は、参加型評価の特性として以下の4つをまとめている。

1. 利害関係者本位の評価：評価の権限と責任は、評価の専門家集団ではなく、プログラムの実施者や関係者にも委ねられる。
2. 評価の過程（プロセス）への参加をとおして、関係者の当事者意識（オーナーシップ）の向上が期待できる。
3. 評価プロセスへの参加は、参加する個人や所属する組織・コミュニティの能力強化につながる。
4. 評価専門家は、多様な経験と異なる立場の人々の討議や対話、合意形成を促進するファシリテーター、コーチといった役割が必要となる。

また評価結果は関係者と共有し、次の行動へつなげていくことが目指され、下向き説明責任が重視される（JICA 2001）。

MSCのデータ収集は、受益者や現場スタッフなど利害関係者が、物語を集める形式で行う。収集データの分析も、利害関係者が、物語を選択する過程を通じて行う。この過程を通じて、利害関係者によるデータ収集及びデータ分析を促進し、その結果、データ分析やインパクトを表現する能力を強化する人材育成に有効になる。また最も重大な変化を現場にフィードバック（下向き説明責任）することで、組織内コミュニケーションを通じて、継続的な対話を作ることが可能になる（図1）。MSCは、利害関係者が、評価プロセスへの参加を通じて「学習」し、意識啓発・能力開発・相互理解につながる参加型評価の特色を備えている。

一方、参加型評価の課題として、専門家によらない関係者の評価であるため、評価の質（妥当性）が低い、また必要な参加者が参加しているかどうかという代表性の問題などの指摘がある（源 2008）。そのため「多様なデータ・情報間の整合性に注意することによって、評価の偏りを回避する」必要があるという（三好・田中 2001）。

③ 質的分析手法としてのMSC

質的分析とは、数量化されない質的データから、どのような意味が見出せるか、帰納的な分析を行うことである。対象のひとつひとつを事例として、存在する文脈に関してできるだけ再現可能でかつ妥当な推論をする。質的分析では、対象の変化を深く理解して、達成に影響を与えた要因を把握できる特色がある（長尾 2009）。MSCは、変化を数値・簡素化するのではなく、複雑な質的変化を物語の形でいきいきと描写できる、質的分析手法の特色がある（Davies & Dart 2005）。一方、質的分析は手続きの一般化が難しく調査の妥当性が担保されにくい、という課題がある（古賀 2008）。

④ 組織学習の手段としてのMSC

組織学習とは、「組織が、変革の必要性を見出し、より一層の成功をもたらすと信ずる変革に着手しうる能力を取得し、発展させること（白石 2009）」である。参加型評価は、事業を行う組織の価値観や力量を問い直すことによる組織学習の効果がある（Estrella & Gaventa 1998）。MSCも、物語を選ぶプロセスを通じて組織の価値観が明らかになり、それを議論したり変化させたりすることができ、プログラムの具体的な活動とパフォーマンスの領域を越えて組織全体が成長することが可能である、という（Davies & Dart 2005）。

3、4章では、3つの国内社会セクター事業のMSC評価を事例研究として取り上げる。

3. 事例研究の比較分析の枠組み

事例研究にあたりMSCなど参加型評価や質的分析の特性を考慮して5つの視点による枠組みを設定する。それらは、①評価目的（MSCは学習・改善に焦点を当てている）、②評価専門家の役割（参加型評価ではコーチ・ファシリテーター）、③データ収集～情報提供者・収集者は誰か・収集方法は何か（必要な関係者から適切に収集しているか）、④データ分析：重大な変化の選択者と選択過程（参加型評価や質的分析では、分析結果の妥当性が問われる）、⑤フィードバックとその後の状況（参加型評価では下向き説明責任を重視）。これら視点から各事例の評価プロセスを確認したい。

4. 国内社会セクター評価3事例

（1）事例1：震災復興事業

「Johnson & Johnson 東北被災地での市民・コミュニティエンパワメントプログラム」とそのMSC評価（以下、震災復興と称する）を、（特活）日本NPOセンターによる報告書（日本NPOセンター 2017）に基づいて報告する。

このプログラムは、ジョンソン・エンド・ジョンソンによる東北復興支援を目的とした助成を受け、日本NPOセンターが2014年2月から2017年3月まで3年間実施した。プログラム目的は「住民が互いに支え合う関係性のある『居場所』と、自らの力と役割を発揮できる『出番』の機会が生み出されることで、住民一人ひとりとそのコミュニティがエンパワーされること」であった。

具体的な活動は岩手県と宮城県で3連携団体が実施した[4]。各団体は受益者参加でバス旅行の実施やお茶会、親子で参加できるイベントの開催などを行ってきた。日本NPOセンターは連携団体と協力しながら、学び・経験・意見共有の機会づくりをサポートする伴走支援活動を行った。

① 評価目的

プログラムの評価を行うことになり、2つの目的でMSCを採用した。まず「ロジックモデル[5]や数的指標では把握しにくい、質的な変化を把握し、記録すること」、次に「『地域と住民のエンパワーメント』、『地域が元気になること』に関する学びを明らかにすること」である。前者はどちらかといえば説明責任に、後者は学習・改

善に力点を置いている。また評価プロセスに、日本NPOセンター、連携団体、受益者、協力者など、利害関係者が参加することで、意識啓発、能力向上、相互理解促進をすることを目指した。

② 評価専門家の役割

連携団体にとって初めてのMSC評価であったため、筆者が外部専門家としてスタッフに対して、MSCの概要・特色・進め方に関する研修を実施した。またデータ収集においてのインタビュー手法を指導した。加えて時間の範囲や領域の設定、データ収集で誰が情報提供者や収集者になるべきか、物語の選択は誰がどう行うのが良いか、フィードバックはどうするかといったコンサルテーションを実施した。その後実際のMSC実施は各連携団体スタッフが進行した。

③ データ収集

訊く時間の範囲は、連携団体と筆者で相談し、関係者の記憶が比較的鮮明であると思われる「遡って1年」とした。領域は2つ設定した。プログラム目標にある関係者の行動変容に着目し、領域Ⅰ:「あなたにとって、この1年で個人（被災者）の役割に起こった最も重大な変化は何ですか」を決めた。また波及的な変化を幅広く知るため、領域Ⅱ:「あなたにとって、この1年で、生活環境に起こった重大な変化は何ですか」も決めた。それぞれに、情報提供者にとって「そう思う理由は何ですか」も訊いた。

これらの設問を、現場の生の声を聴くために各連携団体は受益者である仮設住宅の住民や転居した元住民、自宅再建者や支援者・協力者であるボランティア、子育て中の母親、その家族などプログラムの受益者とその周りの人々から、スタッフが聞き手となってインタビュー調査を実施した。最終的に連携3団体が2つの領域に対して、10 ～ 12ずつ計32のストーリーを収集した。

④ データ分析

各連携団体は集めたストーリーから、1編を選ぶ作業を行った。スタッフと関係者[6]を含め選択チームを結成し、最も重大な変化を1遍選んだ。そしてそのストーリーを選んだ理由を記録した。その結果、各連携団体にとっての「最も重大な変化」が、領域ごとに1編、合計2編ずつ選択された。

・「個人の役割における変化」領域に関しては、話し手が自宅再建者（被災者）で、「疎外感を感じつつも一歩を踏み出せないでいる被災者の友人を、イベントに一緒に参加しようと誘っている」ストーリーや、イベント参加者の家族による「子どもを連れて行けるイベントが増えてきた。妻やまわりのママ友の意識も、そのイベントに提案を出すなど積極的に変わってきた。イベントは参加するだけでなく、一緒になにかをつくるきっかけになっている」というストーリーが、最も重大な変化に選ばれた。被災者・住民がプログラムへの参加を通じて、受動的な受益者ではなく、支え合う、提案をする、作り手側になるなど、役割が積極的なものに変化、主体的に行動してきていることが選択理由である。

・「生活環境についての変化」の領域であるが、以下のようなストーリーが収集・選択された。イベント参加者のストーリーで、「同じ地域内でも被災した人と、していない人の間では心の壁がある。それでもこのプログラムではみんながその壁をなくして1つになって楽しめる」というものであった。被災状況や、補償に対する違いにより、被災者同士で「壁」ができてしまっている。それにも関わらず、連携団体主催のイベントの場では、お互い会話を楽しみ、地域の絆が広がる実感を得ていることがわかる。別のイベント参加者からは「イベントを通じて被災地域と隣の内陸地域が繋がれるのが有り難い。（中略）イベントによって訪れる機会が増えて、人同士が知り合えるきっかけになった」ことがわかった。イベントを契機に異なる地域にまたがって人が交流し、生活環境に影響を生みだしている、プログラム想定外の効果を知ることができた。

⑤ フィードバックとその後の状況

「最も重大な変化」が決定したら、情報提供者に、ストーリーから何を学ぶことができたのかを紙面や口頭で伝えた。フィードバックが受益者に与えた直接的な影響に関しての情報は得られなかったが、連携団体スタッフにMSC実施の感想を聞いた。まずMSCで活動を振り返ることで、質的な変化を把握し、改善していく契機になったという。「人に関する難しいテーマを目的

として暗中模索で実践する中、どのような変化をもってエンパワーメントが発現したと考えるべきか参考になった」「またプログラムの目標設定や復興支援のあり方を、（中略）数値などの指標だけに頼らず、より意味のある、より人の役に立つ活動に近づける上で、大いに助けになった」「活動のアドバイスをもらって今後に活かせる」とある。

また、利害関係者間の理解の促進・共有を助ける効果があったという。「現場では当たり前となっている出来事や価値観が、文字に落とされ、外部者を含めてみんな（利害関係者）が共に理解し、価値を共有すること」ができた。また連携団体スタッフが「団体の活動の目的を再確認できた（活動者自身も、第3者も）」という組織文化を確認する作用があった。さらに連携団体スタッフが、自分たちの活動への反応を聞けることで、自信を持てたり、取り組むべき行動の指針としたりすることができた。

(2) 事例2：イングリッシュキャンプ

岡山県立和気閑谷高等学校は、江戸時代に創立された閑谷学校をルーツに持つ伝統ある学校である。英語研究部は文化系の部活動であるが、2016年当時14名の生徒が、週2回英検の勉強や英語出前授業などの活動をしていた。イングリッシュキャンプ（以下、キャンプと称する）は、活動の1つで、和気町との共催で開催されている。紹介するものは1泊2日で8月に実施され、参加生徒は英語研究部部員7名及び部員外生徒計17名に加え、小・中学生、在日外国人を合わせて70名の参加があった。内容は、英語で遊ぼう、異文化体験ワークショップ、外国人の方にインタビュー、など多岐に及んでいる。キャンプに指導的に関わっていた大野浩志教諭と向敦史支援職員の2名（以下、指導者）が執筆した報告書（大野・向 2016）に基づき報告する。

① 評価目的

指導者によりキャンプの評価にMSCを活用することが発案され、キャンプ終了後2時間のMSCワークショップが開催された。参加者は英語研究部員7名である（キャンプ参加6名）。評価目的は「既に身につけていることと新しい学習内容・経験を自身の中で葛藤させ、新しい知識を獲得していくこと。自分たちの大切にしたい価値観や変化について考え、活動を通して得た学びを言語化したものを自分の言葉として獲得できること」だった。事業を改善するための学習というよりは、MSCの特色である、参加者への学習効果や能力強化、意識啓発を期待したものであった。

② 評価専門家の役割

筆者は同年8月に外部専門家として和気閑谷高校から招聘され、同校を中心に指導者を含む高校教員約30名に対して、MSC手法の概要・特色・進め方に対する研修を実施した。実際のMSCワークショップは指導者の2人が進行した。

③ データ収集

基本設問は指導者が設定した。「キャンプ期間中に自分の考えや行動に起こった一番の変化は何か（変化）」と「なぜそれが一番大きいと感じたのか（理由）」である。これに「それを今後どう生かしていくか」が加えられた。上記について参加生徒全員に、キャンプ終了時に各自が執筆する作文として課せられた。インタビューではなく自筆にしたのは、生徒が書くことに慣れており、その方が多様なストーリーが収集できるという指導者の判断があった。提出された作文の中には「変化」と「理由」が混在、もしくは欠落しているものがあり、両方が記してある6編を、指導者が選出した。

④ データ分析

2段階のストーリーの選択作業が行われた。1段階目は個人で参加生徒7名が6編のストーリーを読み、1番重大と思うストーリーを1編選んだ。2段階目は7名を3名と4名の2グループに分け、各自が選んだ物語から各グループで1編のストーリーを「最も重大な変化」として選出した。1班では3年生が中心となり、議論し自力で選出した。その一方で2班では、決めるための判断軸を自分たちで確立することができず、指導者が介入し「自分たちが大切にしたい視点は何か」という問いかけを行い、「変化後の状態が良いもの」を判断軸にすると決めて1編を選出した。2班が選んだストーリーと選択理由を表2に引用する。

表2　キャンプ2班による「最も重大な変化」と「選択理由」

ストーリーのタイトル
誰かがしてくれるという考えではなく、自らが行動し、その行動の全てに責任を持つこと
最も重大な変化とそう思う理由
（前略）私は、今回のEnglish Camp［原文ママ］を通して『誰かがしてくれるという考えではなく、自らが行動し、その行動の全てに責任を持つこと』を学びました。決して多くない人数と短い準備期間、それに加えて楷楓祭や受験などで忙しい中での準備でした。そんな中で今回のEnglish Campが成功したのは、ひとりひとりが頑張ったからだと思います。自分の役割を全うするのはもちろんのこと、全体を見て自分のできることを見つけ、臨機応変に対応することもできました。小中学生や外国人とのコミュニケーションもよくとれていたと思います。そして私は、今回学んだことを将来“公認会計士”になったときに役立てたいと考えています。そのために大学に行き、留学をして外国の経済学や経営学を学びたいと考えています（後略）。
最も重大な変化に選んだ理由
（前略）この変化は、将来に役立つ可能性が高く、変化の後の状態「自らの行動に責任を持つ」ということが一番良く、また、これからずっと活かしていける変化だという点で、このストーリーを選びました。

（出所）大野・向 2016

⑤ フィードバックとその後の状況

ワークショップ後に両班が、「最も重大な変化」と「選んだ理由」を発表・共有した。自己評価であり、選択とフィードバックが同時に行われたため、フィードバック単独の影響を判別することは困難であった。

全体を通じて、「生徒は良い雰囲気の中、高い興味関心を示し、集中してワークショップに参加」した。生徒の学びに関しては「時間をかけて自分たちの大切にしたい価値観や変化について考えるため、活動を通して得た学びや、起こった変化を言語化したものが、より自分の言葉になる。そして、学びに対する納得感が高まる」という彼らの学習と能力強化が発現した。結論として、生徒が「自発性と積極性と協力性を持つことがどんなことでも一番大事、キャンプで学んだことを積極的に普段の生活にフィードバックする視点」を得ることができ、「自分の意思を持って行動する生徒が増えた」という。またMSCは「様々なストーリーの中から、『1つだけ選ぶ』ということを通して、自分たちの価値観を掘り起こしながら『自分ってこんなことを大切にしたいと思っているんだ』を再認識するという、組織の文化（大切にしたいこと）を浸透させるのに有効である」という感想を得た。

課題としては「選択の際に、共通の判断軸を作るのが難しい」ことがあった。選択時に、各生徒が自分の価値観・経験に基づいて選ぶことはできても、グループで選ぶとなると、抽象度の高い自分たちの判断軸を構築しなくてはならず、議論が止まる傾向があった。実際2班では指導者が介入し、判断軸を作るサポートをすることで1編を選ぶことができた。次に、「自分または知人のストーリーがあることで客観的に判断できなくなる」傾向が見られた。生徒が書いたストーリーを同じ顔ぶれの生徒で選択作業を行ったため、対象のストーリーに自分のストーリーが混じる。「自分の変化が一番重大と捉える傾向がどうしてもあるので、各自選択のし辛さがあったり、またストーリーによっては、誰が書いたのかわかるものがあり、友人関係からくるその人への配慮が客観的な判断を鈍らす可能性がある」とのことだ。

（3）事例3：所沢市社会福祉協議会

所沢市社会福祉協議会（以下、所沢社協と称する）は、「地域住民が抱える生活課題を地域全体の課題としてとらえ、（中略）誰もが安心して暮らせるように、住民、行政、社会福祉関係者、ボランティア、NPO、企業等がお互いに協力していく（所沢社協 2018a）」という地域福祉の考え方に基づき、コミュニティソーシャルワーカー（CSW）の配置[7]、地域福祉サポーター養成講座、生活困窮者自立支援事業などの事業に取り組んでいる。第3次発展計画（2018年～2020年）では、「市民が一緒に作る！健やかに　自分らしく暮らせる　支え合う　心やさしいまち」を使命にしている（所沢社協 2018b）。所沢社協のK職

員を通じて入手した情報をもとに報告する。

① 評価目的

所沢社協ではここ数年、年次評価にロジックモデルに基づく評価を行っていたが、職員の負担が大きい反面、次の展開につながらない、評価結果が活用されないなど、評価が形骸化してきた。そこで平成30（2018）年から、職員のエンパワーメント[8]や事業の改善に適しているMSCを採用すると決定した。そして評価目的は「MSC評価を活用して、重点項目の取り組みの改善を図る」という学習・改善志向のものであった。

② 評価専門家の役割

筆者が外部専門家として社協職員に対してMSCの概要・特色・進め方の研修、及びインタビュー手法の指導を行った。また領域の設定やデータ収集の対象を誰にするか、重大な変化の選択方法への助言、フィードバックについてコンサルテーションを実施した。その後実際のMSC実施は、K職員を中心に所沢社協の関係者で進行された。

③ データ収集

2019年2月下旬から3月末に実施された。情報提供者は常勤職員及び任期付き常勤職員60名であった。基本設問は「平成30年度に起こった重大な変化」と「その変化をなぜ重大と思うか」だが、領域は第3次発展計画の重点項目3つからそれぞれ3つずつ、合計9つ設定された。例えば、重点項目1は「コミュニティソーシャルワークの推進」であるが、「蓄積事例の整理」「支援プロセス」「CSWとしての意識」に関する変化、と3つに分けて領域が設定された。

60名の職員全員から、9つの領域に対する変化全てを収集するのは、人数、情報量ともに多く整理に混乱が予想されるため、職員1名から領域1つのみに「重大な変化」を収集するように、ランダムに組み合わせが調整された。データ収集方法はインタビューが採用され、7名の職員（現場担当）が行った。1名あたり約9人の職員から物語を収集し、計60編が収集された。

④ データ分析

物語の選択作業は6月の所属長（課長補佐含む）・現場担当者の話し合い「MSC評価のワークショップ」によって行われた。それ以前の5月の段階で、9つの領域の60の物語から、所属長（課長補佐含む）9名が1編の「最も重大な変化」をあらかじめ選んでおくという準備作業が行われた。

ワークショップで2つのグループに別れ、所属長（課長補佐含む）の選んだ物語を持ち寄り（4編＋5編）、グループ討論でその中から「最も重大な変化」を、それぞれ1編選んだ。次に各グループで選んだ物語を発表し、全員での意見交換の後、多数決で「全領域で最も重大な変化」を最終的に1編選出した。その物語と、選択理由を表3に引用する。「蓄積事例の整理」の領域のものである。

表3　最も重大な変化と選択理由

ストーリーのタイトル
CSWの見える化
最も重大な変化
私にとって平成30年度に起こった、最も重大な変化は、これまで行ってきたCSWの取り組みを「見える化」したことで、市民の理解を得ることができたことです。
そう思う理由
ここ数年、地域福祉推進課では、CSWを配置して様々な取り組みを行ってきましたが、その記録の作成や周知方法など、不十分な点が多くありました。市民や市職員などから「CSWって何？」「名前は知っているが何をやっているのかは知らない」という声を聞くことも多く、市民はもちろん、将来の財政確保の観点から市職員に対しても、CSWの取り組みの「見える化」を図り、その必要性を伝えることが重要であると感じていました。 そこで平成30年度は、CSW相談システムにて、日頃の取り組み内容をしっかり記録に残すという共通認識を持つこととし、同時に市民に対して何を「見える化」するべきかの検討を行った結果、「CSW活動報告」を作成・発行することに至りました。 「見える化」するツールができたことで、各地区民児協での説明の機会や「CSW★ラボ（実践報告会）」の開催につながることになりました。今年度初めて開催した「CSW★ラボ」では、自治会（まちづくり協議会）、市職員、地域包括支援センター職員などを招いての活動報告（報告書説明、事例発表）を行い、「CSWの取り組みが必要であることを理解できた」「期待しています」「他の職員にも聞かせたかった」という感想をいただくことができました（後略）。

最も重大な変化に選ばれた理由
● CSW の理解がしっかりできていることが重要であり、市民と行政に協力を得るには「見える化」が必要であるため（何を行ったか伝えること）。（中略） ● プレゼン力が必要であり、行政・市民に伝えていく。法人全体で共有して取り組むことが求められる。

（出所）評価プロセスに従い筆者作成

その後ワークショップで、領域ごとの「重大な変化」の中から、今後の活動を改善させる教訓を読み取り、具体的な提言にまとめる話し合いが行われた。上述の「CSWの見える化」に関しては、「CSWについてより多くの市民に知ってもらえるようにする」課題が明らかになったので、CSWと管理職員のミーティングを活用して「見える化」を目指す、業務手順書（マニュアル）の整備、ホームページの見直し、新年度もCSW活動報告を9月～10月に発行する、などの改善策が提案された。一連のワークショップでは、職員による活発な意見交換が行われた。

⑤ フィードバックとその後の状況

「最も重大な変化」と「選択理由」は、所沢社協全体に朝礼、課内会議、口頭、回覧を通じて周知された。選ばれなかったそれ以外の物語は、各職員の所属する管理職と現場担当者でインタビュー結果を共有し、課題をピックアップして職員会議などの場で課題解決に向けた取り組みを協議した。

選択結果を一般職員へフィードバックする際も、マイナスに捉える意見はなく、順調に共有が行われた。特筆できるのは、MSCで活動現場の職員からの情報に基づき、そこから学んだ教訓を具体的な改善の取り組みへと結びつけている点である。「CSWの見える化」をはじめ、他の領域においても、複数の改善案が議論、提案された。その後、これらの改善策は概ね実現したという（2020年4月情報）。組織学習に関してはK氏によれば、「CSWの見える化」が選ばれる議論は、関係者が組織の使命にある「市民が一緒に作る」という価値観を再確認する過程であったという。

所沢社協では今後も継続的にMSC評価を実施する意向である。初回の反省から次回以降は、領域の数をもっと少なく絞る、インタビューの技術をより高めるなどの改善を行うこととなった。

5. 考察と結論

（1）事例の比較から示唆されること

本章ではこれら3事業を相互比較することで、MSC評価を効果的に行うための要点を整理する。またMSC国内活用にむけて今後の展望と課題についても論考したい。

① 評価目的

どの事例でも、MSCの特色に合わせて明確に評価目的が設定されていた。3事例のうち2事例で活動への「学習・改善」を評価目的にしており、1事例は「参加者の学習の深化」であった。またそれぞれの評価目的はMSCの特色に合致したものであった。

震災復興では「質的な変化を把握する」、「活動を振り返ることで教訓を学ぶ」という評価目的であった。被災者の主体的な変化が質的な情報として把握され、今後に活かせるアドバイスを入手することができた。キャンプでは、「参加生徒が大切にしたい価値観や変化を考え、自分の言葉として獲得する」が目的であった。これは参加型評価の特性である、関係者の能力強化に着目している。結果として「自発性や積極性を普段の生活に生かし、自分の意思を持って行動する」という学びが生徒の中で深まった。所沢社協では「重点項目の改善を図る」という目的であったが、「CSWの見える化」を筆頭に、他の事業も改善に向けての具体的な取り組みが明確になり、その後実際に改善が実現した。手法の特色に合致した評価目的を設定することが肝要なようだ。

② 評価専門家の役割

全事例において筆者が評価専門家として、関係者の参加を促すとともに、評価の妥当性を高めるための指導を行った。具体的には事前にMSCの概要・特性・進め方に関する研修を実施した。震災復興と所沢社協においては事実を聞

き取るためのインタビュー手法を指導した。これらは評価知識を参加者に教えるトレーナーとしての役割を果たしたといえる。同様に震災復興と所沢社協において設問や領域の設定に関する助言や、データ収集の提供・収集、最も重大な変化の選択者は誰が参加するのが望ましいか、選択の基準はどうするか（現場の視点を重視すること）、フィードバックはどうするか、などのコンサルテーションを実施した。これらは評価を関係者主体で進める援助をするコーチの役割といえる。実際のMSCプロセスの進行は、全事例において関係者内部によって行われ、導入時に指導を受ければ、MSCは関係者内部で進行（内在化）できることがわかった。しかし評価専門家が導入時に適切な指導を行ったため、関係者の進行でも「お手盛り評価」ではなく「多様なデータ・情報間の整合性に注意し、偏りを回避」する評価に近づけることができたと考える。

その一方、専門家なしでMSCを進行するためには、関係者内部に合意形成を促進するファシリテーターの役割が必要な場合があるようだ。キャンプにおいて参加生徒のみで最も重大な変化の選択基準を設定できなかったため、指導者が介入し援助したのが1例である。他の事例では関係者のみでも、選択基準設定に関する混乱はなかった。

③ データ収集

震災復興では、過去1年間の「個人の役割の変化」と「生活環境の変化」2領域が設定された。収集は連携団体スタッフによるインタビューで、情報提供者は震災の被災者を中心に、加えて被災者をよく知る人やボランティアの声も聞くなど、現場の実情に合わせて範囲を広げた。キャンプでは、キャンプ期間中、「自分の考えや行動」領域に起こった変化を、キャンプ参加者である生徒にとってやり易い、自分で文章に書く形で行われた。所沢社協は、領域が9つと数が多く、情報提供者も職員60名と大人数である特徴があった。混乱や情報の偏りが無いように1人が1領域のみ担当するというランダムな調整を行った。聞き手は、現場担当の職員であり、インタビュー手法で重大な変化を収集した。

データ収集において3事例に共通することは、MSCの定石に準じて「重大な変化」と「重大に思う理由」の2設問を設定してはいるが、時間の範囲や領域は、それぞれ評価目的や活動の性質・規模に合わせて柔軟に設定されたことである。情報提供者に関しては、プログラム受益者やキャンプ参加生徒、社協職員など、活動現場に深く関わる利害関係者からデータを収集している。このことが現場の視点を生かした評価結果を導き出すことに貢献したようだ。データ収集方法は、1対1のインタビューの場合と、参加者が直接記入する方法とがあり、現場の実情に合わせて選択された。

④ データ分析

震災復興では連携団体スタッフと関係者が「選択チーム」を結成して選択した。これには当事者ではない第3者的な関係者を選択チームに入れることで、選択（分析）の妥当性を高める期待があった。キャンプでは各生徒が1編を選択し、2グループに別れて「最も重大な変化」を1編選択、最終的に2編を選ぶ2段階の選択であった。

指導者によれば「自分自身の体験を振り返って考える段階、次に他人の価値観に触れ、自分と共通するものを思案する段階を経て、生徒の内省を深める」狙いがあった。

所沢社協でも2段階の選択が行われた。最初は領域ごとの物語から所属長（課長補佐含む）が1編を選択しワークショップに持ち寄り、2グループでさらに1編ずつ選出、その2編から「全領域で最も重大な変化」を1編選択した。これは選択過程に内部人材とはいえ多様な視点を含めることで、選択結果（分析）の妥当性を高める工夫であった。

情報提供者と選択者との関係に対して、2通りのやり方が観察される。1番目は、震災復興における「選択チーム」や所沢社協の管理職・現場担当による「MSC評価のワークショップ」でみられた話し手とは別の集団が選択するという情報提供者と選択者が異なる場合である。2番目は、キャンプが該当する話し手（書き手）による物語集の中から、参加生徒が「最も重大な変化」を自ら選択するという情報提供者と選択者が同一の場合である。2種類の方法は、それぞれ長所と短所がある。

前者の提供者と選択者が別人物の場合のメリットとして、選択者であるスタッフなど関係者が、現場の受益者などから直接情報を収集できる点である。現場の視点が他の利害関係者に共有され、評価結果に反映されるという参加型評価の長所が活かせる。加えて、提供者が自ら選択する方法と比べて、「お手盛り評価」を避け、第3者的な視点で妥当性の高い選択をすることが可能になると考える。その一方、情報提供者（受益者）が選択過程に直接関与するわけではないので、データ分析への参加による受益者のオーナーシップや能力向上などの効果は期待できなくなるデメリットがある。

後者の情報提供者と選択者が同じである場合、提供者（受益者）が自らデータ分析にも参加することになる。このことは、参加型評価の長所である「評価の過程への参加をとおして、関係者のオーナーシップの向上」「参加する個人や所属する組織・コミュニティの能力強化」が、受益者へ直接発現するメリットが期待できる。キャンプの「自分の意思をもって行動する生徒が増えた」はその好例だろう。その一方、自らの物語を自ら選択することになるため、キャンプでは「自分または知人の物語があることで、客観的に判断できなくなる」問題が生じた。デイビースら（2005）も、受益者が物語の選択に参加することについて「自分自身で書いた物語の選択に本人が含まれると、具合が良くないことがある。別の人々によって物語が描かれた場合、選ぶのは簡単である」とリスクを表明している。どちらが良いかは、プログラムの性質や評価目的にそって決定すべき、と考える。

結論として現場を知る人材で選択を行うが、必要に応じて当事者でない関係者を選択者に含めることで妥当性の高い選択が可能になる。また選択過程を複数段階に分けることで、参加者の内省を深めたり（キャンプ）、多くの視点を動員することで選択の妥当性を高めたりする効果が期待できる（所沢社協）。選択主体において情報提供者が同時に同一の場合と、異なる場合がある。それぞれ長所と短所があり、評価目的やプログラムの性質を考慮して選択するとよい。

⑤ フィードバックとその後の状況

3事例全てにおいて選択結果が情報提供者にフィードバックされた。震災復興では「何を学ぶことができたか」が伝達されたが、フィードバックが受益者に与えた影響は情報入手できなかった。キャンプでは生徒が選択に参加することで結果は自動的に共有できたが、フィードバック自体の影響は把握できなかった。一方、所沢社協では職員全員に「最も重大な変化」と「その理由」が組織的に、朝礼や回覧などで報告された。評価結果は事業の「見える化」のさらなる推進など具体的な改善策に昇華され、その後の改善につながった。フィードバックの際、職員が自然に受け入れたことから考慮して、フィードバックの結果共有が、改善に一定程度寄与したと推察できる。

デイビースら（2005）のいう「下向き説明責任」として評価結果を関係者間で共有し、相互の対話を促進する効果は所沢社協で一定程度観察された。しかし、他事例では十分に把握できなかった。

(2) 3事例で観察されたMSCの特性

3事例の中で第2章の「MSCの4つの特性」に該当する状況が発現していれば、それらはMSCが、プログラムや利害関係者に与えた影響だと考えられる。順次照らし合わせて考察する。

① モニタリング手法及び評価手法である：

今回3事例は全てプログラム終盤もしくは終了後に行う「評価」として実施しており、モニタリングの形では実施していないため、この視点の考察は省略する。

② 参加型評価手法である

「利害関係者本位の評価である」点に関しては、3事例全てに関して評価専門家ではなく、利害関係者の積極的な関与が見られた。震災復興では、主に受益者が物語を提供、連携団体スタッフが収集した。スタッフを中心とした選択チームによって「最も重大な変化」が選ばれ、学びが提供者に報告された。キャンプでは参加者が自ら物語を執筆し、選択と共有も参加者自身で行った。選択過程に指導者が介入したが、査定者ではなく選択を側面支援する内部ファシリ

テーター的な役割であった。所沢社協では、社協職員60名全員が情報提供者であり、インタビュアーも職員であった。物語の選択は、管理職・現場担当の職員で構成される「MSC評価のワークショップ」で行われ、結果が全ての職員に還元された。「参加者のオーナーシップや能力向上」につながる点においては、以下のようである。震災復興では、スタッフが自分たちの活動への反応を聞くことで、自信を持つことができた。キャンプにおいて生徒は積極的に参加し、「自分の意思を持って行動する生徒が増えた」。所沢社協でも、選択作業に管理職職員が熱心に参加した。MSCへの関係者への参加が、彼らのオーナーシップや意識向上に一定程度寄与していると推察できる。

「評価専門家の役割」に関しては、3事例全て、MSC実施は組織やプログラム内部の利害関係者中心で行われた。従来型評価のように、査定者としての評価専門家は存在しなかった。しかしながら、前述のように評価専門家はMSCを側面支援するトレーナー、コーチとしての役割を担った。

所沢社協のK氏は「外部専門家による研修や、設問の領域の設定などに関する助言が効果的であった。仮に当会のみでのMSCを実施した場合、進行で苦慮したと考えられる」とコメントしている。

③ 質的分析手法としてのMSC

震災復興では、被災者が主体的にイベントに参加したり、被災者間の「心の壁」が取り除かれたりしたという質的変化や「異なる地域間の人的交流が増えた」という想定外の変化が把握できた。キャンプでは、参加した生徒の学びの深化が表現された。所沢社協では「報告書を作成し活用したことで、市役所職員や市民に『CSWの見える化』を実現した」変化を把握することができた。関係者の参加で、質的変化や波及的な変化を生き生きと把握できるMSCの長所が発揮されたといえる。

④ 組織学習の手段である

震災復興では、「団体の活動の目的を再確認できた」「自分たちの活動が間違っていないかを考える『ものさし』になる」など、連携団体が組織の価値を改めて確認する機会をMSCが提供した。キャンプでは参加者が「自分たちが大切にしたい価値観について」確認することができた。所沢社協では最も重大な変化の選択作業は、組織の使命にある「市民が一緒に作る」という理念を関係者間で再確認した過程であった。

⑤ 3事例全てに共通すること

全事例で確認できるのは受益者や連携団体スタッフ、キャンプ参加生徒、社協職員などの能力向上、意識啓発など利害関係者のエンパワーメント効果である。そして参加型で質的変化を把握できる点である。震災復興、所沢社協においては、プログラムを改善させる教訓を学習することができた。これらは「エンパワメント評価」など他の参加型評価手法とも共通している。これに対して、MSCは評価プロセス全体に評価専門家が関与する他の参加型評価手法と比較して、導入時に専門家の指導を受ければ実施は内部人材だけで進行できる（比較的容易に内在化が可能である）という独自の特徴があるのではないか。

(3) MSCを効果的に実施するポイント

ここまでの検討の結果、MSCを効果的に行うには、以下の重要項目があると考える。

① **評価目的**：手法の特色に合わせ設定するべきである。

② **評価専門家の役割**：MSC導入時における専門家のトレーナー、コーチの役割が重要である。それによって評価の妥当性を高め、比較的容易に内在化できる。内在化の場合、関係者にファシリテーター機能があることが望ましい。

③ **データ収集**：設問の領域などはMSCの基本に添いながらも評価目的やプログラムに合わせて柔軟に設定する、また現場に精通した関係者から情報を収集することで関係者本意の評価ができる。

④ **データ分析**：選択者に外部者を加えたり、2段階で選択するなどの工夫で、結果の妥当性を高めたり、関係者の内省を深めたりすることが期待できる。情報提供者と選択者が同一の場合と異なる場合があり、それぞれ長所・

短所がある。

(4) 今後の展望と課題

今後、日本国内で MSCを有効に活用するための展望と課題を簡潔に記す。

① 今後のMSC国内活用の展望

震災復興、教育、福祉など国内での異なった分野における社会セクター事業の、学習・改善目的の評価や、利害関係者の学習・能力強化にMSCが効果的とわかった。特にロジックモデルでは把握が難しい波及的な変化の把握や、変化の背景を分析し教訓を学び、改善につなげていくことにMSCが適している。社会的インパクト評価の浸透により、日本の社会セクター事業に対して評価の必要性が叫ばれているが、多くはロジックモデルの評価である。それだけでは実現が難しい関係者のエンパワーメントを目指す手段として、MSCの併用が有益と考える。

② 今後の課題

課題を2つ述べる。まずMSCは日本では認知度の低い手法であり、適切に指導できる評価専門家が少ないという問題があげられる。しかし一度指導を受ければ、MSCは外部の専門家に頼らずとも、組織やプログラムの内部人材だけで内在化が可能である。他の課題として、フィードバックのあり方がある。筆者のバングラデシュでのMSC実施では、フィードバック後に受益者とスタッフの対話が促進される効果が観察された（田中 2014）。このことから筆者はフィードバックの効果に着目している。しかし、今回は3事例とも、「最も重大な変化」は受益者（参加者）に共有されたが、そのことによる直接的な効果を把握できない事例があった。どのような条件のもとで、フィードバックが効果的に行えるか、焦点を絞ったさらなる研究が必要と考える。

注記

1 日本では民間非営利組織の中で、主に国際協力に従事するものをNGO、国内で福祉や教育活動などを行うものをNPOと呼ぶ使い分けが一般的である。

2 国内のMSC活用は、2020年4月現在まで筆者が関与する範囲で本稿3事例を除き、トヨタ財団国内事業（2016）、環境省ESD事業（2016-8）、財務省支援日本劇団協議会（2017-8）などがある。

3 日本語訳である「モスト・シグニフィカント・チェンジ（MSC）手法・実施の手引き（日本語版2013）」は、http://blog.livedoor.jp/sankagatahyouka/MSCGuide-Japanese-2013.pdf　でダウンロードできる。

4 連携団体は、（特活）陸前たがだ八起プロジェクト、（特活）@リアスNPOサポートセンター、（特活）ウィメンズアイ、である。

5 プログラムの目的と手段の因果関係についての仮説を整理したもの。国際協力に使われるロジカル・フレームワークは、ロジックモデルを応用したプロジェクトの計画表である。

6 選択メンバーは、第3者的な視点を議論に含めて選択結果の妥当性・信頼性を高めるために、状況を知っている関係者（理事、他組織、ボランティアなど）を含めた。

7 CSW（コミュニティソーシャルワーカー）：個別の支援を行いながら、ニーズの共通性に着目し、地域の生活課題解決に向け、住民と協働して新しい仕組みづくりに取り組む専門職。

8 K職員によると、傾聴など対人援助技術を活用し仕事を進めている社協職員は、MSC評価に親和性があるという。

参考文献

大野浩志・向敦史（2016）「English Camp参加者に対するMSC評価実践報告」

古賀正義（2008）「質的調査の四半世紀－『質対量』論争を超えて、ポストモダンな時代へ」、『質的調査法を学ぶ人のために』、世界思想社

国際協力事業団（JICA）（2001）「参加型評価とは何か」、『国際協力と参加型評価』、国際協力事業団

白石弘幸（2009）「組織学習と学習する組織」、『金沢大学経済論集』、29(2)

田中博（2014）「参加型モニタリング・評価手法MSC（Most Significant Change）－バングラデシュ NGOでの実践から4つの特色を考察する－」、『日本評価研究』、14(2)

所沢社協（2018a）「所沢市社会福祉協議会－平成30（2018）年度事業のしおり」、社会福祉法人所沢市社

会福祉協議会
（2018b）「第3次発展・強化計画（平成30年度～平成32年度）」
長尾眞文（2009）「データ収集・分析」、『第三期評価士養成講座テキスト』、日本評価学会
日本NPOセンター（2017）『復興の先を見据えた地域づくり～住民の主体性を支えた復興支援のあり方～』、特定非営利活動法人日本NPOセンター
源由理子（2016）「参加型評価とは」、『参加型評価―改善と変革のための評価の実践』、晃洋書房
源由理子（2008）「参加型評価の理論と実践」、『評価論を学ぶ人のために』、世界思想社
三好皓一（2008）「評価とは何か」、『評価論を学ぶ人のために』、世界思想社
三好皓一・田中弥生（2001）「参加型評価の将来性～参加型評価の概念と実践についての一考察」、『日本評価研究』、1（1）
Davies, Rick & Dart, Jess（2005）. *Most Significant Change（MSC）Technique: A Guide to Its Use*
Estrella M.,and Gaventa,J.（1998）. *Who Counts Really?*, Participatory Monitoring and evaluation: literature review, IDS
Humentum（2019）. Two types of Accountability. https://www.humentum.org/free-resources/guide/accountability（2019/10/21閲覧）

（2020.6.13 受理）

MSC (Most Significant Change), participatory and qualitative evaluation method
- Consider three cases in Japan focusing on learning and improvement objective evaluation -

Hiroshi, Tanaka

Participatory Evaluation Center
tanaka.pecenter@gmail.com

Abstract

MSC (Most Significant Change) is a participatory and qualitative evaluation method used by Western NGOs and is suitable for evaluation of learning and improvement purposes, but there are few examples of its use in Japan. In this paper, Author compared and reviewed the three cases of MSC evaluation conducted in the disaster recovery, education, welfare in domestic social sectors, and present reference information for effective use of MSC in Japan. As a result of case studies, following points are revealed. The existence of evaluation experts as trainers and coaches, the setting of evaluation purposes according to the characteristics of the method, the collection of data from the viewpoint of the field, the need to intervene to improve the validity of data analysis, etc.. In addition, the difference from other participatory evaluation methods that internalization is relatively easy was recognized. It is expected that MSC will be utilized in domestic social sector projects in the future, but there are issues such as a lack of the experts to guide and appropriate feedback style.

Keywords

Participatory evaluation, Qualitative evaluation, Learning / Improvement, Social sector, NGO / NPO

日本評価学会第 20 回全国大会
「成果指標の課題～ Issues in Outcome Indicators ～」
開催の報告とお礼

2019年12月7日、8日開催（於：高知大学）の日本評価学会第20回全国大会には、104名の方々にご出席いただきました。誠にありがとうございました。会員各位の日頃の研究や実践活動の報告をもとに、評価研究者、実務者の間の経験、情報、知識を共有化する場として、皆様にとって有益な機会となりましたら幸甚に存じます。多くの方のご尽力により本大会を開催することができましたこと、心より感謝申し上げます。今後とも当学会の活動にご高配賜りますようどうぞ宜しくお願い申し上げます。

実行委員長　　新納　宏（高知大学）
プログラム委員長　牟田博光（東京工業大学）

2019年12月7日（土）

09：30 - 10：00	受 付（114）玄関			
午前の部 10：00 - 12：00	＜共通論題1＞ イノベーションの 形成的評価 （大島巌） （107）講義室5-3	＜共通論題2＞ 国際協力機構(JICA)の 事業評価における 最近の取り組み （正木朋也） （202）講義室5-4	＜自由論題1＞ 人材育成・教育評価 （齊藤貴浩） （204）講義室5-5	
お昼休み 12：00 - 13：15	お昼休み （12：05 - 13：00　理事会　（116）講義室）			
午後の部Ⅰ 13：15 - 15：15	＜共通論題3＞ 国の施策・事業の評価 ～行政評価局調査の 最近の事例～ （佐伯美穂） （107）講義室5-3	＜共通論題4＞ 我が国地域社会における SDGsへの取り組みの 評価：成果と課題 （廣野良吉） （202）講義室5-4	＜自由論題2＞ ODA評価 （林薫） （204）講義室5-5	＜自由論題3＞ 自治体評価１ （佐藤徹） （105）講義室5-2
午後の部Ⅱ 15：30 - 17：30	＜シンポジウム＞ 成果指標の課題 （小島卓弥） （101）階段講義室5-1			
17：40 - 18：30	総会　（101）階段講義室5-1			
18：40 - 20：00	懇親会　生協食堂			

2019年12月8日（日）

9：30 - 10：00	受　付（114）玄関			
午前の部 10：00 - 12：00	＜共通論題5＞ 評価・行政・ アカウンタビリティ （南島和久） （107）講義室5-3	＜共通論題6＞ 学校評価士は 学校運営における 評価的思考の利用を いかに支援できるか （小澤伊久美） （202）講義室5-4	＜共通論題7＞ 資金誘導に活用される 評価の動向とその功罪 ～バリュー・フォー・ マネーの議論を 出発点に～ （今田克司） （204）講義室5-5	＜ラウンドテーブル＞ 評価ディベート： 評価のお悩み相談会 2019 （小島卓弥） （105）講義室5-2
お昼休み 12：00 - 13：15	お昼休み （12：05 - 13：00　編集委員会　（116）講義室）			
午後の部Ⅰ 13：15 - 15：45 （※自由論題４は 15：15 終了）	＜自由論題4＞ 自治体評価２ （南島和久） （107）講義室5-3	＜自由論題5＞ 評価手法・評価デザイン （佐々木亮） （202）講義室5-4	＜自由論題6＞ 行政評価 （窪田好男） （204）講義室5-5	

（　　）は座長です。

日本評価学会第 20 回全国大会
「成果指標の課題～ Issues in Outcome Indicators ～」
プログラム詳細

○印は共同研究の代表者

2019 年 12 月 7 日（土） 受付　　9：30-10：00

2019 年 12 月 7 日（土） 午前の部 10：00-12：00

共通論題1　イノベーションの形成的評価　（107）講義室5-3

		座長	大島巌	日本社会事業大学
		討論者	今田克司	CSOネットワーク
K1-1	対人サービスのイノベーションをEBPプログラムに発展させる形成的評価の方法論～ CD-TEP法:実践家参画型エンパワメント評価の活用～	○	大島巌	日本社会事業大学
			新藤健太	群馬医療福祉大学
			源由理子	明治大学
K1-2	現場実践をEBP効果モデルに発展させる評価ガイドラインの開発－CD-TEP法を活用して－	○	清水潤子	日本ファンドレイジング協会
			新藤健太	群馬医療福祉大学
K1-3	対人サービスでのイノベーションをEBPに発展させるための英国での取組み～ What Works Centreを中心に		家子直幸	三菱UFJリサーチ&コンサルティング
K1-4	実践家参画型エンパワメント評価を進展させるために必要な中間支援組織の役割～ EBP技術支援センターの役割に注目して～	○	新藤健太	群馬医療福祉大学
			大島巌	日本社会事業大学
			鴨崎貴泰	日本ファンドレイジング協会

共通論題2　国際協力機構(JICA)の事業評価における最近の取り組み　（202）講義室5-4

		座長	正木朋也	国際協力機構
K2-1	JICA事業評価の現状と今後の方向性		阪倉章治	国際協力機構
K2-2	JICA事業評価の歴史と内部評価について		大貝隆之	国際協力機構
K2-3	タイ「人身取引被害者保護・自立支援促進プロジェクト」:「プロセスの分析」による学びの抽出		小泉高子	国際協力機構
K2-4	衛星データを用いた国際協力事業の効果検証：カンボジア水力発電所建設・改修事業の事例分析	○	石本樹里	国際協力機構
			倉田正充	上智大学
K2-5	質的比較分析(QCA)の応用：JICA無償資金協力事業のソフトコンポーネントと持続性の因果関係の推論	○	氏家慶介	国際協力機構
			今吉萌子	国際協力機構
			正木朋也	国際協力機構

自由論題1　人材育成・教育評価　（204）講義室5-5

		座長	齊藤貴浩	大阪大学
J1-1	留学生30万人計画の評価の試み		佐藤由利子	東京工業大学
J1-2	学校評価における評価指標の在り方についての実態分析		橋本昭彦	国立教育政策研究所
J1-3	防災・減災／復興を牽引する起業家育成プログラム開発と評価デザイン	○	加藤知愛	北海道大学
			武田浩太郎	東北大学
			祇園景子	神戸大学
J1-4	身体運動能力におけるけん玉の練習効果に関する試行的研究	○	關谷武司	関西学院大学
			坂口将太	聖和短期大学
			江嵜那留穂	関西学院大学
			Thet Mon Myat Myint Thu	関西学院大学
			神前駿太	関西学院大学
			西川智貴	関西学院大学
			木内隆裕	関西学院大学
			桑原志帆	関西学院大学
			越仲舞	関西学院大学
			大木康平	関西学院大学
			奥川菜々	関西学院大学
			安井志保美	関西学院大学
			鉄谷紗理	関西学院大学

2019 年 12 月 7 日（土） お昼休み　12：00-13：15（12：05-13：00 理事会　（116）講義室）

2019 年 12 月 7 日（土） 午後の部 I 13：15-15：15

共通論題3　国の施策・事業の評価～行政評価局調査の最近の事例～　（107）講義室5-3

		座長	佐伯美穂	総務省
		討論者	西出順郎	明治大学
K3-1	最近の行政評価局調査の事例 －空き家対策に関する実態調査－		吉田大志	総務省
K3-2	最近の行政評価局調査の事例 －女性活躍の推進に関する政策評価－		田中俊充	総務省
K3-3	最近の地域計画調査の事例 －南海トラフ巨大地震対策に関する実態調査－		荒木和久	総務省

共通論題4　我が国地域社会におけるSDGsへの取り組みの評価：成果と課題				（202）講義室5-4
		座長	廣野良吉	成蹊大学
K4-1	我が国の自治体のSDGsへの取り組みのオーバービュー		鈴木克徳	ESD活動支援センター
K4-2	静岡市におけるSDGs推進の取組		稲葉博隆	静岡市
K4-3	SDGsのローカライズを促進する参加型評価の可能性 －珠洲市（石川県）を事例に－		高木超	慶應義塾大学大学院 （明治大学大学院博士課程）
K4-4	持続可能な地域づくりにSDGsは必要か？愛媛県内子町の取り組みにおける成果と課題		堀江由美子	セーブ・ザ・チルドレン・ジャパン

自由論題2　ODA評価				（204）講義室5-5
		座長	林薫	文京大学
J2-1	SDGs評価：未来の世代に、素晴らしい世界を4：SDGs目標4.1への貢献：ミャンマー CREATEプロジェクトのインパクト評価（英語による発表）	○	佐々木亮	JICA-CREATE Team in Myanmar / International Development Center of Japan (IDCJ)
			渡辺真美	
			Aye Yie Mon	
			Htay Htay Han	
J2-2	SDGs評価：未来の世代に、素晴らしい世界を4：インドネシア政府によるSDGsグローバル指標への取組		石田洋子	広島大学
J2-3	地方開発プロジェクトにおけるエンパワメント評価の機能：タイ王国D-HOPEプロジェクトの評価活動からみる課題と可能性	○	岡部有美子	インスティチュート・フォー・コミュニティ・デザイン
			米原あき	東洋大学
J2-4	DAC評価項目のSDGsへの適合（報告）	○	村岡敬一	外務省
			長谷部（蟹川）わかな	外務省

自由論題3　自治体評価１				（105）講義室5-2
		座長	佐藤徹	高崎経済大学
J3-1	評価表のチェックリスト活用による自己評価のレベルアップ		池田葉月	京都府立大学
J3-2	新しいツールによる自治体評価の有効性向上		窪田好男	京都府立大学
J3-3	自治体障害者就労支援策のインパクト評価		後藤玲子	茨城大学
J3-4	実証研究に基づくメタ評価チェックリスト －図書館評価を事例として－		田辺智子	国立国会図書館／筑波大学

2019 年 12 月 7 日（土）　午後の部 II　15：30-17：30

シンポジウム　成果指標の課題				（101）階段講義室5-1
		座長	小島卓弥	NTTデータ経営研究所
		討論者	西出順郎	明治大学
			米原あき	東洋大学
	高知大学よりご挨拶		新納宏	高知大学
S-1	政府の評価実務から見た「成果指標」		砂山裕	総務省
S-2	自治体における成果指標（高知県の地方創生・産業振興計画について）		澤田博睦	高知県
S-3	教育における成果指標の設定		齊藤貴浩	大阪大学
S-4	ODAにおける成果指標		江口雅之	国際協力機構

2019 年 12 月 7 日（土）　総会　17：40-18：30（101）階段講義室 5-1

2019 年 12 月 7 日（土）　懇親会　18：40-20：00（生協食堂）

2019 年 12 月 8 日（日）　受付　　9：30-10：00（114）玄関

2019 年 12 月 8 日（日）　午前の部　10：00-12：00

共通論題5　評価・行政・アカウンタビリティ				（107）講義室5-3
		座長	南島和久	新潟大学
		討論者	西山慶司	山口大学
K5-1	自治体における行政と政策の統制		山谷清秀	浜松学院大学
K5-2	日本の府省横断的予算における政策の主流化		橋本圭多	神戸学院大学
K5-3	行政と評価のアカウンタビリティ「再考」		山谷清志	同志社大学
K5-4	評価学と政策過程論：再論		南島和久	新潟大学

共通論題6　学校評価士は学校運営における評価的思考の利用をいかに支援できるか				（202）講義室5-4
		座長	小澤伊久美	国際基督教大学
		討論者	鄭廣姫	韓国教育開発院
K6-1	学校評価士の現況－養成講座開設からこれまで		橋本昭彦	国立教育政策研究所
K6-2	学校評価士の実践と今後の展望－初期的考察		石田健一	
K6-3	大学生の学修成果の可視化 －プログラム運営の発展的・主体的議論を促す－		高林友美	サイバー大学
K6-4	評価的思考による大学初年次プログラム運営の実践例 －卒業時インタビュー－		小澤伊久美	国際基督教大学

共通論題7　資金誘導に活用される評価の動向とその功罪 ～バリュー・フォー・マネーの議論を出発点に～				（204）講義室5-5
		座長	今田克司	CSOネットワーク
K7-1	Evaluation and Value for Money: Thinking Beyond Measurement（英語による発表）		Julian King	Julian King & Associates Limited
K7-2	新自由主義下における社会介入の評価		津富宏	静岡県立大学

ラウンドテーブル	評価ディベート：評価のお悩み相談会2019 ～研究・実践活動において日々生じている評価の課題や疑問（あるいはお悩み）を共有し、解決の糸口を探る～			（105）講義室5-2
		座長	小島卓弥	NTTデータ経営研究所
			西出順郎	明治大学

2019年12月8日（日）お昼休み　12：00-13：15（12：05-13：00 編集委員会　（116）講義室）

2019年12月8日（日）午後の部Ⅰ 13：15-15：45（※自由論題4は15：15終了）

自由論題4　自治体評価２				（107）講義室5-3
		座長	南島和久	新潟大学
J4-1	県民の幸福に着目した総合計画の策定と評価 ～岩手県の取組事例～		和川央	東北公益文科大学大学院／岩手県
J4-2	地方自治体におけるメリハリある行政評価を目指した相対評価手法の研究 ～長野県岡谷市商業活性化計画を例として～		飯田洋市	公立諏訪東京理科大学
J4-3	岩手県における県民協働型評価の実態 －参加型評価の事例－	○	鎌田徳幸	政策21
			熊谷智義	政策21
J4-4	岩手県北上市における協働事業の検証 －第三者評価の結果より－	○	熊谷智義	政策21／合同会社地域計画
			鎌田徳幸	政策21／岩手県立大学
			岩渕公二	政策21

自由論題5　評価手法・評価デザイン				（202）講義室5-4
		座長	佐々木亮	国際開発センター
J5-1	日本における評価システムの特徴とガバナンス機能の有効性		三上真嗣	同志社大学大学院（博士課程）
J5-2	外部評価者の独立性－JICA外部事後評価における評価者の独立性確保－		河野摂	国際協力機構
J5-3	オープンデータ施策に関わる目標と評価の設定		本田正美	東京工業大学
J5-4	米国における根拠に基づく政策（EBPM）推進に果たす非営利組織の役割とソーシャルインパクト－米国連邦・地方政府へのデータ利活用の能力開発・支援活動を行う米国非営利組織Results for America－		白川展之	文部科学省科学技術・学術政策研究所
J5-5	Testing The Capital Asset Pricing Model: Empirical Evidences in the Vietnam stock Market（英語による発表）		Nguyen Thi Thanh Nga	Reitaku University（麗澤大学）

自由論題6　行政評価				（204）講義室5-5
		座長	窪田好男	京都府立大学
J6-1	日本の公的セクターにおける財政運営の持続可能性について		吉田素教	大阪府立大学
J6-2	行政事業レビューにおけるEBPMの展開		杉谷和哉	京都大学／国際高等研究所
J6-3	政策評価制度の日韓比較－比較政策学から		韓廷旼	同志社大学大学院（博士課程）
J6-4	独立財政機関（IFI）の有効性に関する文献レビュー		上野宏	国際開発センター／神戸大学
J6-5	防潮堤評価のための試論：気仙沼大島地区の事例から		上野真城子	アジア都市コミュニティー研究センター

日本評価学会第20回全国大会セッション報告

シンポジウム

シンポジウム「成果指標の課題」

座長　小島 卓弥（NTTデータ経営研究所）
討論者　西出 順郎（明治大学）
米原 あき（東洋大学）

評価を実施するにあたり、隘路に陥りやすいポイントの1つが「成果指標の設定」ではないだろうか。評価が導入される多くの分野において「売上」や「利益」等、万人に分かりやすく設定しやすい指標を得ることが難しく、多くの評価担当者が苦慮しているところである。一方で、不適当な評価が設定されることで、本来想定されない、誤った方向に事業や施策が誘導されてしまうこともまた少なくない。

以上を踏まえ、本シンポジウムでは高知大会における共通テーマでもある「成果指標の課題」について、政府、地方自治体、ODA、教育の各分野の評価の現場における現状と課題について報告してもらうとともに、それらの解消に向けて本学会が果たすべき役割について議論を深めていくことを目的に開催したものである。

まず、発表者による発表に先立ち、本テーマを提案した大会実行委員長の新納宏会員（高知大学）より趣旨や問題意識に関して発表した。それを踏まえ、各発表者から以下の発表があった。

第1発表者の総務省砂山裕氏（総務省行政評価局）からは「政府の評価実務から見た「成果指標」」と題し、政府におけるEBPMの取組について紹介するとともに、行政評価局と各府省で共同実施しているEBPM実証的共同研究の成果を題材としての分析指標例や、政策評価の総務省による点検結果から指標に関する指摘例等が発表された。

第2発表者にはご当地である高知県澤田博睦氏（高知県前産業振興推進部副部長（地方創生担当））より「自治体における成果指標（高知県の地方創生・産業振興計画について）」と題し、高知県の地方創生（産業振興計画の取組の総括）について発表があった。高知県における人口減少や経済状況、特に厳しい中山間地域の人口減少等の状況を整理した上で、高知県まち・ひと・しごと創生総合戦略における成果指標の設定の状況や、実際の成果や今後の取組の方向性について発表があった。

第3発表者の齊藤貴浩会員（大阪大学）からは「教育（大学）における成果指標の設定」と題し、「教育の成果」の難しさ、学校教育と評価の変化や現状、多くの利害関係者が交錯する大学評価における現状等を紹介するとともに、「教育で「万人」にわかりやすい（納得できる）成果指標の設定は可能か？」という課題提起をし、万人が納得できる成果指標（価値観）はないと考える一方で、できるだけ多くの人が納得できるような価値観を、多様な利害関係者と評価に基づくコミュニケーションをとりながら、ともに探し続けることの必要性について発表した。

第4発表者の江口雅之氏（国際協力機構（JICA）評価部審議役）からは「ODAにおける成果指標」と題し、JICAにおけるODA成果指標、JICA自体の業績指標、JICAプロジェクト評価においてどのような取組を実施しているのか整理するとともに、業績評価の近年の特徴として、SDGsとの関係性や「目標・評価の具体性・客観性・明確性等の確保」を意識した2015年の独立行政法人通則法の改正を受けた取組等についても発表があった。

これらを受け、コメンテーターの米原あき会員（東洋大学）からは、「評価の目的と方法」という2軸で

整理して各発表者に質問するとともに、指標によって価値を可視化し、議論／コミュニケーションできるようにすることの必要性について言及があった。

同じくコメンテーターの西出順郎会員（明治大学）からは、各発表者に評価における技術的な課題や外的要因と成果指標の関係性、評価指標がもたらす悪影響等について質問があった。

本シンポジウムでは、政府、地方自治体、教育、ODAの各分野における評価の実践者から、現場における評価の取組について「成果指標」にフォーカスして事例紹介が行われ、この分野における多くの新たな気付きがあり、また様々な課題が明らかとなった。

評価における成果指標の課題は本シンポジウムだけで語りつくせるものではなく、本学会において今後も継続的に取り上げられるべきテーマであると考えている。今後も同じようなテーマのセッションを定期的に開催され、課題や新たな取組の共有が行われることが期待される。

共通論題セッション

共通論題1 「イノベーションの形成的評価」

座長　　大島 巌（日本社会事業大学）
討論者　今田 克司（CSOネットワーク）

本セッションでは、近年、既存の制度では対応できない様々な新しい社会課題が拡大し、有効な解決方策が模索されている中で、そのような課題解決に向けて非営利組織等民間セクターの役割への関心も高まっている。これに対して、社会的インパクト評価やエビデンスに基づく政策立案（EBPM）に関心が集められ、課題解決のために新たな発想で生み出された「社会的イノベーション」を、より効果的なプログラムモデルに発展させるために有効な形成的評価の方法論が求められている。しかしながら現状では、世界的に見ても「効果モデル」を適切に形成・発展させるために有効な形成的評価の方法論は確立していない。そのために、社会課題の解決が十分に進展しない状況がある。このような中、本セッションでは、参加者約30名を得て、社会課題解決のために開発された社会的イノベーションを育て形成・改善し、EBP（Evidence-Based Practice）プログラムへと発展させるために求められる、有効な形成評価のアプローチのあり方について検討した。

第1報告では、大島巌会員（日本社会事業大学）らが「対人サービスのイノベーションをEBPプログラムに発展させる形成的評価の方法論～ CD-TEP法:実践家参画型エンパワメント評価の活用～」を報告した。この報告では、大島らの研究チームが開発した、ボトムアップ評価（Chen、2015）、実践家参画型評価の方法を用いて、様々な利害関係者（stakeholders）が参画し、社会が協働して「EBP効果モデル」を形成・発展させる形成的評価の方法論「CD-TEP法」について、社会課題解決のために開発されたイノベーションを形成・発展させる方法の1例としてまず検討した。

その上で、第2報告では、清水潤子会員（日本ファンドレイジング協会）らが、「現場実践をEBP効果モデルに発展させる評価ガイドラインの開発－CD-TEP法を活用して－」を、イノベーションの形成的評価の実例として、その適否と可能性を報告した。

第3報告では、家子直幸会員（三菱UFJリサーチ&コンサルティング）が、「対人サービスでのイノベーションをEBPに発展させるための英国での取組み～ What Works Centreを中心に」として、英国の対人サービス・イノベーションの形成的評価の取組みを、EBP仲介機関の取組みから紹介し、日本での適用可能性を検討した。

第4報告では、新藤健太会員（群馬医療福祉大学）らが、「実践家参画型エンパワメント評価を進展させるために必要な中間支援組織の役割～ EBP技術支援センターの役割に注目して～」を報告した。この報告では、実践家参画型エンパワメント評価が、実践家等が十分な評価キャパシティを形成して初めて適切な形成的評価が実施できることを踏まえて、どのようにこのような状況を作り出すことができるのか、中間支援組織、技術支援センターのあり方、可能性等についても発言した。これは英国のEBP仲介機関の取組みとも関連する議論になっている。

討論では、指定討論者・今田克司会員（CSOネットワーク）からの発言を皮切りに、多くの活発な議論が行われた。CD-TEP法はEBPプログラムを目指すために一定の科学的厳密さが追求される、それに対して地域ごとの実情をどのように訣えるのか、スケールアウト（モデルの横展開）をどう位置付けるか、技術支援センターの「技術支援」とは何か（支援する側、される側ではなく、協働ではないか）、誰の価値を優先するのか？（障害のある人のニーズ等をどう反映するのか）、一般市民の関心を高めるために何を行えば良いか？といったフロアからのコメントも含めて活発な意見交換がされた。今後重要性が増すイノベーションの形成的評価について貴重な議論の場になったと考える。

共通論題２「国際協力機構（JICA）の事業評価における最近の取り組み」

座長　正木 朋也（国際協力機構／北里大学）

国際協力機構（JICA）は2015年以後、共通セッションとしてこれまでに述べ7回の企画セッションを開催している。そこでは、単にJICAの事業評価現場の最新動向を公表するのみならず、参加者との直接的な質疑応答と議論を経て、事業評価システムと評価活動の改善のためのフィードバックをいただくことも重要な目的の1つとしている。JICA評価の最新の取組のレビューをもとに、今回は、発表者を含め30名ほどの参加者により活発な議論が行われた。

座長よりセッション企画の経緯と趣旨説明の後、それぞれ15 ～ 20分程度で発表を行った。

第1報告者の阪倉章治氏（評価部部長）は「JICA事業評価の現状と今後の方向性」と題して、今後のJICA事業評価を考えるために、まず評価の世界動向をレビューし、次いで現状のJICA事業評価の概要を共有して当セッションの基調を行った。

第2報告の大貝隆之氏（国際協力専門員）は「JICA事業評価の歴史と内部評価について」と題して、JICA事業評価の歴史と定着の経緯を振り返りつつ、とりわけJICAの自己評価である「内部評価」の実施状況を概観し、その内容と課題等を紹介した。

第3報告は、小泉高子氏（評価2課課長）より、「タイ「人身取引被害者保護・自立支援促進プロジェクト」:「プロセスの分析」による学びの抽出」と題して、プロセス分析手法の応用により通常のODA事業評価とは異なる観点・手法による学びの抽出の試みを報告し、今後のJICAによる協力の改善に向けての示唆を共有するものであったが、都合により口頭発表は行われなかった。

第4報告の石本樹里会員（評価2課）と倉田正充氏（上智大学）は、「衛星データを用いた国際協力事業の効果検証：カンボジア水力発電所建設・改修事業の事例分析」と題して、事業効果検証のために十分な指標が設定されていない場合や、治安等の問題によりデータを収集することが困難な場合に、衛星データや地理データを活用することにより、事業効果を定量的に可視化して評価につなげる試みの技術面の紹介

とともに今後の多方面への応用可能性を紹介した。

第5報告では、氏家慶介氏（評価1課）、今吉萌子氏（評価1課）、正木朋也会員（国際協力専門員）らが、「質的比較分析（QCA）の応用：JICA無償資金協力事業のソフトコンポーネントと持続性の因果関係の推論」と題して、質的比較分析（QCA）の手法を、ソフトコンポーネントの持続性評価に関わる原因探索の試みへの適用事例をもとに、QCAの今後の応用可能性と限界について紹介した。

全ての発表の後、フロアからQCAデータの作り方とベン図の描き方について、また、地理データの活用に関わる技術的及び実施コストに関わる質疑応答がなされた。次いで、省庁では内部評価が一般的であるところをJICAが外部評価を行うようになった経緯、及びそのメリット・デメリットに関わる質疑があり、順次、演者より回答ののち、過去議論も踏まえた座長からの補足を加えた。

以上、これまでのJICAの取組みにかかわる過去6回のセッションに加えて、今回、改めて歴史的視点も含めたレビューを行い、参加者からの多岐にわたる質疑とコメントが共有された。JICA評価における最新の科学的な取組みについても、その歴史的背景と経緯も踏まえて再考し、意義ある議論を行うことができた。新たな試みは試行錯誤のもとで行われる傾向にあるが、場当たり的であってはならない。それらアプローチが評価論におけるどの位置付けにあたるかも明確にしつつ説明責任を果たしてゆくことが望まれる。さらに、それらがJICA評価システムにおいて、どのような位置付けで行われているかを自ら意識しておくことも重要である。

これまでにも議論されたこのような課題については、今後も参加者らとの議論を重ねながら整理して、改善・解決のための努力を続けることを再確認した。改めて、JICA評価システムのさらなる改善と、現場ニーズにも同時に対応する努力を続けることをコミットして閉会した。

共通論題3「国の施策・事業の評価 ～行政評価局調査の最近の事例～」

座長／司会　佐伯 美穂（総務省）
討論者　　西出 順郎（明治大学）

本セッションでは、総務省が実施する「行政評価局調査」の最近の事例を報告することを通じ、学術研究に実務の実例を素材として供するとともに、実務に同調査の動向を紹介することを企図したものである。「行政評価局調査」は、担当府省とは異なる立場から、複数府省にまたがる政策や各府省の業務の実施状況を実地に調査し、その課題や問題点を実証的に把握・分析し、改善方策を提示するものであり、全国的な調査（全国計画調査）と地域の行政上の課題についての調査（地域計画調査）の双方の事例を取り上げた。

第1報告は、吉田大志氏（総務省行政評価局）による「最近の行政評価局調査の事例－空き家対策に関する実態調査－」である。同報告では、空き家対策について、行政手法等の蓄積が途上にあること、空き家や市町村の状況は区々であり「ベストプラクティス」を見出し難いこと、市町村には先進事例を調べるリソースが不足していること等の特徴を踏まえ、同調査では、空き家対策の実情を明らかにし、市町村の参考となる取組事例を取りまとめ、情報提供するというアプローチが採られたこと等について説明がなされた。

第2報告は、田中俊充氏（総務省行政評価局）による「最近の行政評価局調査の事例－女性活躍の推進に関する政策評価－」である。同報告では、女性活躍の推進について、国・地方・事業主の取組等を定める法律の施行後3年の検討に際し、同政策評価において、女性活躍推進の取組の効果等を可能な限り定量的に把握・分析するべく、差の差分法、回帰不連続デザイン等の統計分析手法を用い、EBPMの観点からリーディングケースの提示に向けた試みが行われたこと等について説明がなされた。

第3報告は、荒木和久氏（総務省高知行政監視行政相談センター）による「最近の地域計画調査の事例－南海トラフ巨大地震対策に関する実態調査－」である。同報告では、四国における重要かつ喫緊の課題である南海トラフ巨大地震対策について、想定される死者の8割は津波が原因であること、いつどこで災害に遭遇するか分からないこと等を踏まえ、同調査において、地域住民のみでなく、観光客・通行者等の視点から、災害情報伝達体制、緊急避難場所への誘導・案内等について、昼夜の2回、実際に現地を訪問する等して調査が行われたこと等について説明がなされた。

コメンテーターの西出順郎会員（明治大学）からは、第1報告について、「実態調査」において現状・原因の分析に注力することは評価の質の向上の契機となる旨、また、市町村の実態を全国的・機動的に把握することにより市町村の取組のベンチマークに資する旨の指摘があった。また、第2報告については、分析に関し、公共政策の分野において介入群と統制群の比較を行う場合の留意点、経済状況等の外部要因の取扱い、入手可能なデータによる制約等についてコメントがあった。第3報告については、いつ発生するか分からないという災害の性質上、対策に係るアクティビティ・アウトプットを定期的に調査することの意義や、同様の調査を他の地域へ展開する可能性について指摘がなされた。

フロアとの意見交換においては、担当府省による実態や効果の把握と総務省による実態調査等との関係、統計分析等に要するマンパワーや期間等について質疑応答が行われた。

共通論題4「我が国地域社会におけるSDGsへの取り組みの評価：成果と課題」

座長／司会　廣野 良吉（成蹊大学）

最初に企画・司会者による本セッションの目的、パネリストの紹介、セッション運営方式の説明があり、その後4人のパネリスト報告に引き続き約40分の質疑議応答が活発に展開された。

1）司会者による報告では、①SDGs2016-30の誕生に至る国連での背景とSDGsに先行したMDGsとの共通点と相違点と、②2019年7月国連本部で開催されたSDGs「HLPF＝ハイレベル政治フォーラム」と9月のSDGs首脳会議の議論、③国連経済社会局のSDGs報告2019と国連アジア太平洋経済社会委員会のSDGs報告2019で指摘されたSDGs実施成果と課題、④今後10年間の一層の努力目標に向けた国連事務総長の各国首脳へのアピール、特に本共通論題セッションの主テーマである地方自治体におけるSDGs実施体制の強化の重要性への言及が紹介された。

2）パネル討議では、4人の異なった職場経験をもつ専門家による報告発表があった。

鈴木克徳氏（ESD活動支援センター）は、我が国の地方自治体のSDGsへの取り組みの概観が提示され、①SDGsが2030アジェンダの一部であり、2030年の望ましい社会の実現に向けた方策のメニューを示して、チェッククリストとして169のターゲットから適切なものを選ぶことが重要であること、②企業や自治体のような異なるステークホルダーにとってSDGsが有する意義が異なること、③SDGsを考える上で重要な地域のビジョンづくりは伝統的に自治体が担ってきたことを明らかにした上で、選定されたSDGs未来都市については、2030年のビジョンづくりを重視していることを指摘し、モニタリング・評価の仕組みを明らかにすることの重要性と、SDGs推進に向けた人材育成が鍵であることを強調された。なお、SDGsウォッシュを回避するようSDGsの理解の促進に向けたあらゆるステークホルダーの対話の場づくりが重要という指摘があった。

稲葉博隆氏（静岡市）からは、静岡市のSDGsについて①市政への組込み、②普及啓発、③情報発信の3本柱で取り組んでいる現状が紹介された。①については、SDGs推進は目的ではなく手段であり、SDGsは各団体の設置目的を達成するための有効なツールであることが市民の間で共有された。そのため、各団体は17の目標全てに取り組む必要がなく、いくつかの目標との関連に留まるであろうという議論があったが、

SDGsは世界の課題集であり、チェックリストであると捉えた場合には、SDGsの17の目標、169のターゲットの視点から自らの団体運営、あるいは政策・施策を見つめ直してみるという活用も可能であることも認識された。総合行政を推進する責務を負う地方自治体においては、両方のアプローチで、SDGsを有効活用していくことが求められているといってよいであろう。そして、地方自治体には、ムーブメントを作っていくため、②普及啓発に短期的・集中的に取り組んでいくことが求められており、その過程や結果として、③情報発信を行なっていくことも肝要である。

高木超氏（慶応義塾大学大学院）は、参加型評価を用いて住民参加によるSDGsのローカライズに挑戦した石川県珠洲市の事例を報告した。珠洲市（2018年度SDGs未来都市に選定）は、SDGs推進の要となる産官学金（産官学＋金融機関）からなる能登SDGsラボを開設し、「2030年における理想の珠洲の姿」に関する検討を行った。その背景には、高齢化率が約47%に達し、人口減少が深刻化しているという珠洲市の主要課題がある。そこで、中学生から60代までの住民とともにワークショップを実施して、目的と手段のつながりを論理的に構造化して示すロジック・モデルを作成し、参加型評価によって持続可能な珠洲市を目指すための戦略を検討した。特に、「20代の若者が地域で新しいアイデアや「面白いこと」を始められるような環境を整える」という目標を設定し、この目標の達成度合いをどのような指標で判断するか検討した。SDGsの重要な要素であるマルチステークホルダーによる連携の重要性が参加者の中で認識・共有され、住民参加型手法の展開が、まちの持続可能性につながる手ごたえを得たと力説された。

堀江由美子氏（セーブ・ザ・チルドレン・ジャパン）は、「持続可能な地域づくりにSDGsは必要か？」という根本的な質問を提示し、愛媛県内子町の取り組みにおける成果と課題を分析した。愛媛県内子町における、行政と地域住民の協働による町並み、村並み保存運動やエコロジータウン推進を通した持続可能な地域づくりと、SDGsが目指す持続可能な社会との親和性を紹介した上で、SDGsが採択された直後よりフォーラムやセミナー等を実施してきた経緯から、内子町がSDGs先進地として早い時期から注目されてきた経緯を紹介した。こうした経緯にも関わらず、SDGsは内子町の政策に取り込まれるには至っておらず、第2期総合計画後期計画策定プロセスの紹介を通して、SDGsが重視するアプローチが多く採用されながらも、明示的に組み込まれなかった要因を分析した。さらに、こうした経験を通して見えてきた、地域でSDGsに取り組む上での課題や必要な要素について考察し、中でも「SDGsへの貢献と、目の前にある地域課題の解決策とがうまくつながらない」という地域の声にいかに応えるかを考察した結果、SDGsの17という個々の目標よりも、5原則（普遍性、包摂性、参画性、統合性、透明性）を重視したアプローチの重要性が指摘された。

フロアから、幾つかの質問とコメントがあったが、本セッション司会者として最も印象に残ったことは、①各市町村におけるSDGs地方版の導入では、「誰一人取り残さない」というSDGsビジョンと「5原則」の共有が重要である、②人口減に直面している地方都市において「若者が戻って来たくなる地域にするには、どうしたら良いか」という考え方への発想の転換が必要、③自治体ではSDGs達成に向けた宣言ではなく、積極的な行動が不可欠、④SDGsのグローバル指標や政府が作成した「ローカル指標」に従うのではなく、サブ・ローカル指標策定に向けた方法論の確立を進めて、自治体SDGsのモニタリングと評価体制を強化する必要性等が再認識されたことである。SDGsの達成に向けて総ての住民、組織を巻き込んでいくプロセスが肝要という点で、セッションの報告者のみならず、参加者一同が共有していた。我が国におけるSDGsが、サイロメンタリティ（縦割り）の発想や政策を横断的につなぐ「かすがい」の役割を果たすことを願って止まない。

共通論題5「評価・行政・アカウンタビリティ」

座長　　南島 和久（新潟大学）
司会／討論者　西山 慶司（山口大学）

本共通論題は、行政学の蓄積を土台とし、評価学に行政学の蓄積を接合できないか、できるとすればそれはどこなのかについて、議論するのであった。

第1報告は、山谷清秀会員（浜松学院大学（2020年４月より青森中央学院大学へ移籍））による「自治体における行政と政策の統制」であった。行政学では伝統的に、制度的統制－非制度的統制、外在的統制－内在的統制の軸に分かれる4つの象限に、行政に対するコントロールを分類整理してきた。本報告では、この構図に対する批判的な検討を行うとともにこの構図と政策評価との関係を議論するものであった。行政に対するコントロールは、行政機関が担う政策のアカウンタビリティと表裏の関係にある。また、その一部は政策評価ともリンクする。このような意味において、本報告は、この構図と政策評価との関係を議論するものであった。

第2報告は、橋本圭多会員（神戸学院大学）による「日本の府省横断的予算における政策の主流化」であった。これは行政学における総合調整論の派生論である。従来、一体的な行政のあり方を阻害するものとしてセクショナリズムが問題視されてきた。さらに、行政の専門性が高まるほどセクショナリズムが深刻化することが指摘されてきた。また、複数の府省をまたぐ政策については、この問題はいっそう深刻であった。政府部内では、これらの問題に答えるため、計画や予算のあり方が注目されてきた。本報告では、その延長線上において、「横断的予算」に注目し、政策評価でもよく取り上げられる「プログラム予算」について実証的な議論を行うものであった。

第3報告は、山谷清志会員（同志社大学）による「行政と評価のアカウンタビリティ「再考」」であった。本報告において重要なのは、アカウンタビリティ概念に階層性があるという指摘を行っている点であった。また、本報告はその階層の中において「プログラム・アカウンタビリティ」が位置づけられる点、及びそれが政策内容に関するコントロールを伴うものであるがゆえに難しいとされる点を指摘していた。本報告では、「日本評価学会に課された使命」として、評価のレスポンシビリティの啓蒙、評価を通じたアカウンタビリティ・レジームの整備充実を提唱するものであった。

第4報告は、南島和久会員（新潟大学）による「評価学と政策過程論：再論」であった。本報告において重要な点は、政策過程論と呼ばれる、「政策立案→政策決定→政策実施→政策評価→政策終了」のなかで政策評価が位置づけられること、及び「政策形成」「政策実施」と「政策評価」との関係がコインの表裏となっていることが、評価学会において十分に共有されていない点を指摘したことであった。その上で本報告では、EBPM（証拠に基づく政策立案）の登場と共に、「政策評価」と「政策立案」との関係がいっそう曖昧になっていること、また、「政策評価」の延長線上で語られる「政策終了」と「政策決定」との関係についてあらためて整理が必要であることが指摘されていた。

共通論題6「学校評価士は学校運営における評価的思考の利用をいかに支援できるか」

座長　　小澤 伊久美（国際基督教大学）
討論者　鄭 廣姫　　（韓国教育開発院）

本セッションでは、本学会における学校評価士養成講座発足時から現在までの学校評価士の現況を概観した上で、学校評価士らが学校現場に入る際の課題を報告し、その経験に基づく今後の展望を述べた。学

校評価士にとって特に困難な課題は、多忙な日常においても評価的思考による学校運営が実行可能だと学校側に認識してもらい、学校評価士を現場に受け入れようと決断してもらう点にあり、その突破には、良い事例蓄積と学校側への発信が考えられる。そこで、セッション後半では、「評価的思考」を用いた学校運営の事例を2つの大学における実践から報告した。最後に、指定討論者から、韓国における学校評価を踏まえたコメントをいただき、学校評価士の今後の可能性について、議論した。

第1報告は、橋本昭彦会員（国立教育政策研究所）による「学校評価士の現況－養成講座開設からこれまで」である。本報告では、学校評価士養成講座の企画者の立場から、学校評価士を中核メンバーとする学校評価分科会の活動の軌跡と共に学校と評価士の現状を踏まえ、学校に、学校評価の制度や形式ではなく、学校運営に評価の視点を活かす思考法を導入する支援をしようと考えた経緯が発表された。

第2報告は、石田健一会員による「学校評価士の実践と今後の展望－初期的考察」である。本報告では、まず、学校評価士数名が参加した学校評価を振り返り、個々の体験学習の評価（事業評価）では、観察・インタビュー・作文や資料の分析等により、学校から評価対象として依頼のあった体験学習の成果の可視化はされたが、学校側と評価者の合理性のずれがあったことが判明し、学校評価士が期待したほど評価の成果が学校運営に生かされなかったことが報告された。また、それを踏まえ、今後を展望する上で念頭に置くべき学校評価士の強みと弱み、機会と脅威とが示された。

第3報告は、高林友美会員（サイバー大学）が「大学生の学修成果の可視化－プログラム運営の発展的・主体的議論を促す－」と題して、大学等に質保証や説明責任のために要請され、他律的に導入される学修成果の可視化を、エビデンスに基づく実質的な改善活動に結びつける試みが示された。具体的には、評価的思考の活用と主体的な議論を持つために、評価的思考に基づく指標の捉え直しの実践、その際にプログラムのメンバーをいかに巻き込んだかが話された。

第4報告は、小澤伊久美会員（国際基督教大学）による「評価的思考による大学初年次プログラム運営の実践例－卒業時インタビュー－」である。本報告では、多忙な日常の中に評価的思考を埋め込み、プログラム運営に生かした例として、大学初年次プログラムにおいて、目標や現時点での課題をメンバーで確認し、定例会議等の議事の1つにデータ収集や議論を時機に応じて部分的に組み込み、また、卒業時の学生インタビューを実施し、活用した事例が報告された。

その後、本セッションの指定討論である鄭廣姫氏（韓国教育開発院）により、韓国における学校評価の現状、学校コンサルタントの養成と活動が紹介された。また、今日において再び学校評価を考える意義や、学校側の多様な評価需要・レベルに応える学校評価士の力量形成（持続的成長のシステムの構築）を含めた展望が示され、フロアを含む全体で議論がなされた。

共通論題7「資金誘導に活用される評価の動向とその功罪～バリュー・フォー・マネーの議論を出発点に～」

座長　今田 克司（CSOネットワーク）

本セッションでは、バリュー・フォー・マネー（VFM: Value for Money）の議論に焦点を当て、ニュージーランドから参加のメイン報告者（1）とオンラインでつなぎ、ビデオ会議方式で報告をしてもらった。必要な通訳は座長が行った。続けて会場からの論考をメイン報告者（2）が行った。約40人が参加し、熱のこもった議論の場となった。

社会課題解決における民間セクターの役割に注目が集まるなか、ESG投資や社会的インパクト投資における投資対象の選別（デュー・ディリジェンス）や成果評価の手法の精緻化の動きが進みつつある。日本においても、事業成果と資金調達や金銭的報酬を直接結びつけようという機運も高まっている。そんな中

で、ニュー・パブリック・マネジメント（NPM）の考え方と共に導入されたVFMアプローチは、事業によって生み出された価値を算出する方法で、広く公共事業等の分野で活用されてきている。この動向について、メイン報告者2名が議論を展開した。

メイン報告者（1）のジュリアン・キング氏は、本報告において、これまでのVFMのアプローチが不十分であったことを指摘し、特に価値判断基準やステークホルダーの参加を進化させたValue for Investment アプローチを提唱した。従来のVFMが、物事の経済分析に止まるのに対し、Value for Investment アプローチでは、経済学に評価学を掛け合わせ、ロジックとエビデンスの活用により、事業により生み出される価値の可視化と共有を試みる。

メイン報告者（2）の津富宏氏は、そもそも「社会」という公的なものに対する介入に、投資のロジックはなじむのだろうか、VFMは求められるべきなのだろうか、という問題意識を投げかけた。VFMアプローチは新自由主義の核心的事項であり、それにより、市民の客体化が推進されていると分析する。

2つの報告が問題提起となり、セッション参加者が教育、福祉、国際協力等の専門領域における評価のあり方を問うやりとりがなされた。質疑応答というよりも、提示された社会の流れやそこで評価が担わされている役割についての問題意識を共有する時間となった。整理されたことは、メイン報告者（1）がVFMアプローチに代表されるようなNPMのあり方に対する reformist（改革者）であろうとし、具体的な手法を編み出すことに専心してきているのに対し、メイン報告者（2）がこれに対する transformist（変革者）であろうとし、NPMが依って立つ土台を揺るがそうとしていることだった。

近年、日本においては、成果連動型委託契約の一種と解説されるソーシャル・インパクト・ボンド（SIB）が、ヘルスケア、就労支援、起業支援等の分野で試行的に始まっており、2016年度に法制化された休眠預金等活用法においては、社会的インパクト評価を活用し、評価結果によって適切な予算・人材等の資源配分に反映することが評価の目的の1つになっている。本セッションではVFMを端緒に会場とのディスカッションを試みたが、時間が限られていたことが残念だった。本セッションで糸口をつかんだような評価に対する時代的要請に、評価専門家はいかに答えていくべきか、さらなる議論が求められることが浮き彫りになった。

ラウンドテーブル「評価ディベート：評価のお悩み相談会2019」

座長　　小島 卓弥（NTTデータ経営研究所）
進行補助　西出 順郎（明治大学）

本セッションは、第19回全国大会に引き続いて開催され、評価の理論と実務の有機的結合を具現化する取り組みの1つとして実施したものであり、3回目の取組み（試み）となる。

学会員各位の日々の研究もしくは実践活動において日々生じている課題や疑問（あるいはお悩み）を広く共有し、また、当該議論を多様な活動領域や実務、理論といった複眼的なレンズを通じ、これらの解決の糸口を模索するものである。

前回同様、特定の評価領域を設けず、政策評価、自治体評価、NPO評価、施設評価、EBPM等幅広い分野における評価活動における課題、疑問、問題意識が寄せられ、それぞれに対し、参加者から自らの研究、あるいは実践経験から助言が供される等、評価の現場で発生している様々な課題に対して活発な議論が行われた。

評価の実践者は相談する相手がおらず、一人で苦悩するケースも少なくないようだ。このような場を通じてそれぞれが抱える課題に対し、研究者と実践者がそれぞれの知見を持ち寄り、課題解決を図ることは非常に重要であると考えている。

また、本学会において分野、研究者、実践者の壁を越えて議論が行われることは、我が国における評価の発展という側面からも非常に重要だと考えており、このような機会を継続することができればと考えている。

自由論題セッション

自由論題1「人材育成・教育評価」

座長　齊藤 貴浩（大阪大学）

本セッションでは4つの研究報告に対して、フロア（約30名）からの質疑を得て活発な議論が行われた。

まず、第1報告者である佐藤由利子会員（東京工業大学）の「留学生30万人計画の評価の試み」は、2008年に掲げられた「留学生30万人計画」について、主に留学生の出身国と受け入れ先の学校種別、就職率、満足度に関して定量的指標を中心に検証を行った報告であった。現在、30万人という数値目標はほぼ達成されたものの、それは途中から日本語教育機関で学ぶ留学生を含めたことによる増加が大きな影響を与えていること、そしてその教育の質に懸念があることが指摘された。また、日本国内での就職率の分析から、高度人材の獲得という視点では成功していないという問題提起がなされた。今後の日本社会の在り方に影響を与える施策でもあり、より詳細な分析と政策への反映が期待される。

第2報告者である橋本昭彦会員（国立教育政策研究所）の「学校評価における評価指標の在り方についての実態分析」では、自治体のガイドラインや規定を中心に評価指標に関する情報収集を行った報告がなされた。評価指標の捉え方の違い、プロセス（取組）を指標にしているケースが多いこと、数値目標に対しては慎重な姿勢、積極的活用等、ガイドラインでも自治体によって立場が異なること等が明らかにされた。昨今の教育委員会改革の影響についての質問には、首長の影響からか重点目標等の評価への反映があるとのことであり、今後の学校評価の行く末を見守る必要があると考えられる。

第3報告者である加藤知愛会員（北海道大学）他2名の「防災・減災／復興を牽引する起業家育成プログラム開発と評価デザイン」は、複数の大学が連携して提供する「レジリエント社会を牽引する起業家精神育成プログラム」に関する実践研究の報告であった。災害時や災害からの復興にも役立つ人材の育成を狙った教育プログラムの開発や、教育効果を測定するためのルーブリックの開発が紹介され、今後の展開が期待される。今後の課題として、災害のような逆境を意図的に作り出すことができないことから真の能力が獲得できているかどうかの評価や、またレジリエンスを目的としても通常は組織として存在しなければならず、そのような組織と教育プログラムの運用をどのように結びつけるかといった議論がなされた。

第4報告者である關谷武司会員（関西学院大学）他12名の「身体運動能力におけるけん玉の練習効果に関する試行的研究」は、けん玉を練習することが身体的運動能力を向上させる効果があるとの仮説のもと、けん玉の競技者と一般者の様々な能力を測定し、バランス感覚に有意差を認めたという報告であった。質疑応答の中で、被験者の属性の問題、またけん玉以外に複数の要因が関連しているのではないかとのコメントがあり、現在は試行であって今後検討をしていきたいとの回答があった。伝統的な遊びであるけん玉は、今は新しいスポーツとして発展しているとのことであり、今後の健康・運動能力面での評価手法の開発にも期待したい。

本セッションでは、「人材育成・教育評価」をテーマとして、人間の学習と教育、能力獲得に関する様々

な報告が行われた。教育の評価は評価研究の原点の1つでもある。新しい時代に対応するための人材育成・教育がどのようなものであるべきか、そしてそれに対応する評価がどうあるべきかを切り口として、新しい社会の形成と、更なる評価学の発展に寄与していくことが期待される。

自由論題2「ODA評価」

座長　林 薫（文教大学）

本セッションではODAに直接、間接に関係する4件の報告が行われた。

第1報告の佐々木亮会員による「SDGs評価：未来の世代に、素晴らしい世界を4：SDGs目標4.1への貢献：ミャンマー CREATEプロジェクトのインパクト評価」は、現在JICAの協力により進められているミャンマーの初等教育のカリキュラム改革の評価を試みたものである。ミャンマーでは伝統的に教師の教えた内容、教科書に書かれたことを丸暗記し、質問も許さないというような一方的な教育が行われてきた。JICAの協力によってミャンマー政府が取り組んでいる改革は、グローバル化する現代において、このような教育では人材の育成は難しいという認識のもとに、ディスカッションやアクティブ・ラーニング等の方法を導入して、生徒が自ら考える双方向型の教育にしていこうとするもので、現在初等教育の第4学年まで完成している。本研究では、カリキュラム改定前後で、対象群、比較群の間に有意なパフォーマンスの差が生じていることが報告された。分析手法等についてフロアから質問があり、今後対象学年が拡大される中で、研究の深化が期待される。

第2報告の石田洋子会員による「SDGs評価：未来の世代に、素晴らしい世界を4：インドネシア政府によるSDGsグローバル指標への取組」は現在、2020年から開始する新国家開発5か年計画をSDGs達成のロードマップにアラインさせるべく、グローバル指標の見直しやメタデータ作成等、次期SDGs行動計画策定の準備が進められているインドネシアの取り組みを分析したものである。これまで3年間の経験を、どう評価して、どのように次の国家開発計画や行動計画に反映し、SDGs達成につなげようとしているのかについて分析を行い、多くの課題があることが指摘された。SDGsは早くも15年の期間の序盤が終わろうとしており、インドネシアの事例からみても、その達成は容易ではないことを再認識させる発表であった。

第3報告の岡部有美子会員、米原あき会員による「地方開発プロジェクトにおけるエンパワメント評価の機能：タイ王国D-HOPEプロジェクトの評価活動からみる課題と可能性」は地方開発プロジェクトの評価を行う際、その評価が地域コミュニティ成員にとって能力向上の機会となっているか、その評価が多様なコミュニティの特性を適切に捉え、それらの特性を尊重した評価活動を行っているか、という観点で評価の評価を行ったものである。報告では構成主義的なアプローチによる質的データの検討と、構成主義を出発点に持ちつつ、データの量化を目指す実証主義的なアプローチを融合することにより、地方コミュニティというミクロな生活圏におけるひとや社会の活動を対象とする評価活動の可能性を展望することができた、とする一方で、エンパワメント評価が変革的に機能するためには、政策体系のローカライゼーションが行われなければならないこと、ワークショップの手法では能力の高いファシリテーターが不可欠となり、その養成をどうするかが課題であること、地域の実情にあった質問票を策定するためには、専門性を持った社会調査士が地方各地を巡って調査設計を行う必要があること等を中心に専門人材の育成と配分が重要であることが指摘された。

第4報告の村岡敬一会員による「DAC評価項目のSDGsへの適合」はDAC評価5項目の改訂の動きについて報告がなされた。SDGs時代の到来により、四半世紀以上開発協力評価の基準となっていたDAC評価5項目の有用性に関する議論がOECD-DACで検討され、2019年11月には下部組織である開発評価ネットワーク（EvalNet）において評価項目の改定について合意が形成される予定となっているという状況の下で、公表可能な範囲で議論のポイントや改訂の方向性について紹介がなされ、今後の課題等について議論が行われ

た。

以上のように4報告では、ODA評価に関わる効果、評価手法、課題、キャパシティー、国際的動向等に関して、多様な論点が提供され、活発な議論が行われることによって、セッション全体としてODA評価の研究の深化に貢献するものになったと考えられる。

自由論題3「自治体評価1」

座長 佐藤 徹（高崎経済大学）

本セッションでは5会員による研究報告に対し、フロアー（約25名）からの質疑を得て、活発な議論が行われた。

第1報告者である池田葉月会員（京都府立大学大学院）の「評価表のチェックリスト活用による自己評価のレベルアップ」は、自治体評価が業績測定を中心とする評価制度であるため、自己評価の中心的な作業である評価表の作成に着目したものである。実際に作成された評価表から問題点を整理して作成する際のチェックリストを作成し、それを施策・事業の担当課または事務局の職員が活用することで自己評価の質を向上させられるのではないかといった観点からの検討がなされた。今後は、具体的にはどのようなチェックリストが考えられるか、またどのようなチェックリストが有効であるかについての研究報告が期待されるところである。

第2報告者である窪田好男会員（京都府立大学）からは「新しいツールによる自治体評価の有効性向上」と題した報告があった。はじめに評価学の想定する評価体系の中で、自治体評価がどのように位置付けられるのかを再確認した上で、自治体評価に関わる行政職員の負担軽減や評価への参加の促進等が課題として提示された。さらに報告者が京都府内の市町村や宝塚市で行ってきた評価の実践を事例として挙げつつ、課題解決にファシリテーションやAI（人工知能）等のツールが有効ではないかといった提案がなされた。大変興味深い報告であり、今後さらなる実証研究の報告が待たれるところである。

第3報告者である後藤玲子会員（茨城大学）の報告は「自治体障害者就労支援策のインパクト評価」と題するものであった。行政区域内における就労年齢の障害者のうち8割を超える人々が就労することを成果目標に掲げて、その目標を達成した岡山県総社市の「障がい者千人雇用事業」に着目し、一般指標デザインに基づく差の差法によって障害者就労支援策のインパクト評価を行った結果が報告された。自治体の障害者就労支援策が障害者就労の質及び量に与えたインパクトを定量的に評価した研究であり、また自治体施策を対象としたインパクト評価の事例としてみても貴重な報告であった。今後は研究論文としてまとめられることが期待される。

第4報告は、田辺智子会員（国立国会図書館／筑波大学）より、「実証研究に基づくメタ評価チェックリスト－図書館評価を事例として－」と題した報告があった。具体的には、公共図書館の業績測定を対象に、評価担当者が評価の有効性向上に活用できるメタ評価の有用なツールとして作成されたチェックリストについて提案・報告がなされた。そのチェック項目は、公共図書館を対象としたインタビュー調査やアンケート調査をもとに評価の有効性向上に影響する要因を抽出し、その実施状況を確認したものであり、2館の公共図書館での試行と必要な修正を加えて完成させたチェックリストは図書館評価を振り返り、改善可能な点を把握する上で有用であることが確認されたとの報告があった。社会実装という点において、大変興味深い報告であった。

本セッションでは、「自治体評価」を切り口としつつ多角的な観点からの研究報告があった。本セッションの特徴を挙げるとすれば、自治体評価の現状に対して新たな提案を行うタイプの研究が多かったように見受けられる。その意味では今後は更に深化した研究報告が期待されるところである。

自由論題4「自治体評価2」

座長　南島 和久（新潟大学）

本セッションでは4名の会員による研究報告が行われた。

第1報告者は和川央会員であり、論題は「県民の幸福に着目した総合計画の策定と評価～岩手県の取組事例～」であった。岩手県では2019年度の総合計画の策定に際し、幸福度指標と政策体系を一致させ、幸福度指標の測定結果を評価に反映させたという。このような取り組みは、理念としては理解できるものの、国内に類例を見いだすことは極めて難しい。今後の展開について注目を要する取り組みといえるだろう。

第2報告者は飯田洋市会員であり、論題は「地方自治体におけるメリハリある行政評価を目指した相対評価手法の研究～長野県岡谷市商業活性化計画を例として～」であった。その内容は、多基準意思決定手法の階層分析法（AHP）を取り入れた相対評価を活用した事例として、岡谷市商業活性化計画を紹介するものであった。AHPは「問題」を構造化し、的確に捉えるための手法として知られている。本報告では、「結果に妥当性が認められる」としており、実践性の高い手法として注目することができるだろう。

第3報告者は鎌田徳幸会員及び熊谷智義会員であり、論題は「岩手県における県民協働型評価の実態－参加型評価の事例－」であった。本学会ではこれまでに何度か岩手県のNPO、政策21の取り組みが紹介されてきた。今回の報告はその後の発展を示すものであった。今回の報告において注目すべきは、岩手県知事といわてNPO中間支援ネットワークとの間で締結された「県の施策の評価及び立案に関する協働協定」（2008年）に基づき協働型評価の取り組みが展開してきたという点だろう。

第4報告者は熊谷智義会員、鎌田徳幸会員、岩渕公二会員であり、論題は「岩手県北上市における協働事業の検証－第三者評価の結果より－」であった。本報告の内容は、岩手県北上市における協働事業を対象とする、協働手法の質的向上を図るための第三者評価の実践報告であった。「協働のまちづくり」を掲げる自治体にとっては、協働手法のイノベーションは喫緊かつ重要な課題といえるものであるが、「協働」という事業の性質上、行政側も当事者の立ち位置をもつものであるため、第三者の観点からの評価が求められることとなる。この意味において、本報告の実践事例は、極めて重要な取り組みであるといえるだろう。

自由論題5「評価手法・評価デザイン」

座長　佐々木 亮（国際開発センター）

本セッションでは、評価手法及び評価デザインに関する多様な発表が行われた。

「日本における評価システムの特徴とガバナンス機能の有効性」と題して、三上真嗣会員（同志社大学大学院（博士課程））から発表が行われた。日本の評価システムと特徴、そして改善すべき点に関して良く整理された報告であった。

続いて、「外部評価者の独立性－JICA外部事後評価における評価者の独立性確保－」と題して、河野摂会員（国際協力機構）から発表が行われた。日本評価学会評価者倫理・スタンダード分科会のアンケート結果を踏まえた発表であった。外部評価者の第三者性の認識は近年改善してきているが、氏が所属する組織の実践も踏まえた最新の動向の報告であった。

続いて、「オープンデータ施策に関わる目標と評価の設定」と題して、本田正美会員（東京工業大学）から発表があった。行政機関によるオープンデータの開示の状況と今後の課題を整理した発表であった。

さらに「米国における根拠に基づく政策（EBPM）推進に果たす非営利組織の役割とソーシャルインパクト－米国連邦・地方政府へのデータ利活用の能力開発・支援活動を行う米国非営利組織」と題して、白

川展之会員（文部科学省科学技術・学術政策研究所）から発表があった。米国の研究機関が主導するEBPMの最新の状況と、米国の非営利組織のデータ利活用の状況が手際よくまとめられていた。EBPMは日本でも普及させていかねばならない動きであり、日本ではどのように取り組んでいくべきかの示唆もあった。

最後に、「Testing The Capital Asset Pricing Model: Empirical Evidences in the Vietnam stock Market」と題して、Nguyen Thi Thanh Nga会員（麗澤大学）から研究発表があった。統計分析に基づく実証研究であったが、想定外の出来事による影響をどのようにモデルに取り込んでいくかが今後の研究課題となろう。

以上、5本の発表のあとに、フロア参加者との質疑応答が行われた。さらに司会者から各発表に対する質問とコメントもなされた。たいへん活発な議論がなされた。手法はそれぞれで違うが、お互いに学ぶことが多いセッションであった。今後も同様のセッションが開催されることを望む。

自由論題6「行政評価」

座長　窪田 好男（京都府立大学）

本セッションでは、5名の会員による研究報告に対し、フロアー（約10名）からの質疑を得て活発な議論が行われた。

第1報告は、吉田素教会員（大阪府立大学）による「日本の公的セクターにおける財政運営の持続可能性について」であった。日本の中央政府財政と地方政府財政の持続可能性に関して、公的セクター内における財政移転の効果を考慮した検証を試み、持続可能性が乏しいと結論づけている。財政学のディシプリンとスタイルによる報告であったため、評価学との関連を問う質問があった。報告者からは、効果検証を試みている点で評価学であるという回答がなされた。

第2報告は、杉谷和哉会員（京都大学／国際高等研究所）による「行政事業レビューにおけるEBPMの展開」であり、本大会で多く見られたEBPMをテーマとする報告の1つであった。行政事業レビューにおいて、日本型EBPMと呼ぶべきものが観察できるが、それには改善の余地があるという趣旨の報告であったが、先行研究の整理、日本型EBPMの概念とその説明といった点に改善の余地があるという指摘がなされた。フロアーの関心を最も集めた報告であり、今後の進展に期待したい。

第3報告は、韓廷旼会員（同志社大学大学院博士後期課程）による「政策評価制度の日韓比較－比較政策学から－」であった。比較政策学という公共政策学のサブディシプリンを成立させて、日韓の政策評価制度を比較しようという意欲的なテーマであり、今後に期待したい。一方で、今回の報告では、公開情報をもとに日韓の政策評価制度の紹介に留まっているのではないか、報告者の過去の報告との違いが見えにくいという指摘もあった。

第4報告は、上野宏会員（国際開発センター／神戸大学）による「独立財政機関（IFI）の有効性に関する文献レビュー」であった。OECD加盟国で27機関あり、わが国への導入も検討すべきとされるIFIの有効性について、先行研究の整理を試みた報告である。IFIの有効性を高めるためには、IFIが自由に活動するのではなく、IFIがよって立つ財政ルールが必要であるというのが到達点であることを明らかにした。

第5報告は、上野真城子会員（アジア都市コミュニティー研究センター）による「防潮堤評価のための試論：気仙沼大島地区の事例から」であった。東日本大震災の津波被害を受けて、被災地各地で建設されつつある大規模な防潮堤について、工学的な評価だけではなく、多面的な評価が必要であることを、気仙沼市大島地区を事例として論じた報告である。

本セッションでは、行政評価を切り口として、ディシプリン間の交錯もあり、大胆な試論の提示もあった。本稿では、それらについて、やや辛口に紹介してきたが、いずれも評価学を発展させ、より豊かにする可能性を秘めていることは言うまでもなく、報告者の各会員には、今後の研究報告を期待したい。また

一方で、ディシプリン間の架橋や対話、試論の提示というような報告の場合、セッションのあり方について、約10分の質疑応答時間が有効かという思いも持った。質疑応答の時間を長くするとか、こうした性格のものはラウンドテーブルディスカッションでまずは扱うとか、研究大会の運営においても検討の余地、可能性があるのではないだろうか。

日本評価学会
第21回全国大会のご案内

日本評価学会会長
山谷清志（同志社大学）

例年、秋季に開催しております日本評価学会第 21 回全国大会については、昨今の新型コロナウイルス感染症（COVID-19）の影響等に鑑み、非対面方式による開催を予定しております。詳細は決定次第、学会メーリングリストおよびホームページにてご案内いたします。

記

１．　日程：2020 年 11 月 28 日（土）、29 日（日）

２．　方法：非対面方式（予定）

以上

日本評価学会誌刊行規定

2005.2.15改訂
2002.9.18改訂
2001.9.9改訂

（目的および名称）

1. 日本評価学会（以下、「学会」という）は、評価に関する研究および実践的活動の成果を国内外の学界をはじめ評価に関心をもつ個人および機関に広く公表し、評価慣行の向上と普及に資することを目的として、「日本評価研究（仮名）」（英文仮名：“The Japanese Journal of Evaluation Studies”、以下、「評価研究」という）を刊行する。

（編集委員会）

2. 「評価研究」の編集は、後で定める「編集方針」にもとづいて編集委員会が行う。
3. 編集委員会は、学会会員20名以内をもって構成し、委員は学会理事会が選任する。編集委員の任期は2年とし、再任を妨げないものとする。
4. 編集委員会は、互選により委員長1名、副委員長2名および常任編集委員若干名を選出する。
5. 編集委員会は、最低年1回編集委員会を開き、編集方針、編集委員会企画、その他について協議するものとする。
6. 編集委員会は、その活動等について、随時理事会へ報告し、承認を受けるとともに、毎年1回学会年次大会の場で、過去1年の活動成果と翌年の活動計画に関する報告を行う。
7. 委員長、副委員長および常任編集委員は、常任編集委員会を構成し、常時、編集実務に当たる。

（編集方針）

8. 「評価研究」は、原則として、年2回刊行する。
9. 「評価研究」の体裁は、B5版とし、和文又は英文とする。
10. 「評価研究」に掲載する原稿（以下「論文等」という）の分類は、以下の5カテゴリーからなるものとする。
    （1）総説
    （2）研究論文
    （3）研究ノート
    （4）実践・調査報告
    （5）その他
11. 「評価研究」への投稿有資格者は、学会会員および常任編集委員会が投稿を依頼した者とする。学会会員による連名での投稿および学会会員を主筆者とする非会員との連名での投稿は、これを認める。編集委員による投稿はこれを認める。
12. 投稿原稿を上記分類のどのカテゴリーとして扱うかは、投稿者の申請等をもとに常任編集委員会が、下記の「作業指針」に従って決定する。
    （1）「総説」は、評価の理論あるいは慣行について概観する論文とし、その掲載については編集委員会が企画・決定する。
    （2）「研究論文」は、評価の理論構築あるいは慣行の理解について重要な学問的貢献となると認められる論文とし、その採否については次項に定める査読プロセスを経て常任編集委員会が決定する。
    （3）「研究ノート」は、「研究論文」作成過程での理論的あるいは経験的な研究の中間的成果物に相当する論考で、その採否については次項に定める査読プロセスを経て常任編集委員会が決定する。
    （4）「実践・調査報告」は、評価事業の実践あるいは評価にかかわる調査の報告で、その採否については次項に定める査読プロセスを経て常任編集委員会が決定する。

(5)「その他」には、編集委員会が独自に企画する特集に掲載する依頼原稿や学会誌の刊行に関する編集委員会からの学会会員への連絡等が含まれる。

13. 論文等は2名の査読者により査読することとし、その人選は編集委員会が行う。「研究論文」については、査読結果と編集委員会が査読者とは別に指名する担当編集委員1名の参考意 見をもとに、編集委員会が掲載に関する決定を行う。「総説」、「研究ノート」、「実践・調査報告」および「その他」の論文については、査読結果にもとづき編集委員会が掲載に関する決定を行う。

14. 編集委員が「評価研究」に投稿した場合には、当該委員はその投稿に係わる常任編集委員会あるいは編集委員会の議事に一切参加しないものとする。

15. 上記いずれのカテゴリーの投稿についても、常任編集委員会による掲載の判断は可・不可の二者択一で行うこととする。但し、場合によっては編集委員会の判断で、小規模の修正による掲載も認める。「研究論文」としての掲載が適当でないと判断された場合でも、投稿者が希望すれば、常任編集委員会は「研究ノート」あるいは「実践・調査報告」としての掲載を決定できる。

(投稿要領の作成公表)

16. 編集委員会は、上記の編集方針にもとづき投稿要領を作成し、理事会の承認を得て、広く公表する。

(配布先)

17. 「評価研究」は、学会会員に無償で配布するほか、非会員に有償で提供する。

(抜刷の配付)

18. 「評価研究」掲載論文等の抜刷り30部を、投稿者（原著者）に無料で配布する。それ以上の部数を希望する場合は投稿者（原著者）の自己負担とする。

(インターネット上の公開)

19. 「評価研究」掲載論文等は、投稿者（原著者）の了承を得て全文をインターネット上で公開する。

(著作権)

20. 「評価研究」に掲載された論文等の著作権は各投稿者（原著者）に帰属するものとし、編集権は本学会に帰属するものとする。

(事務局)

21. 「評価研究」編集及び配布の事務は、それに関連する会計も含めて学会事務局が担当する。

（以上）

『日本評価研究』投稿規定

2019.1.21改訂
2017.7.20改訂
2008.9.29改訂
2003.4.18改訂
2002.3.25改訂
2001.9.9改訂

1. 『日本評価研究』（The Japanese Journal of Evaluation Studies）は、評価に関する論文、論考、調査報告等を掲載する。

2. 『日本評価研究』は、会員間の研究成果交流の場を提供し、内外における評価研究の一層の発展に資することを主目的として発行されており、原則として会員による寄稿を掲載する。なお、依頼原稿を除き、ファーストオーサーは学会員でなければならない。また、投稿は、一時に一原稿に限るとともに、他学会誌などへ二重に投稿などのない未発表のものとする。

3. 投稿された原稿は、編集委員会の責任において審査を行ない、採否を決定する。審査にあたっては、1原稿毎に2名の査読者を選定し、査読結果を参考にする。（査読者には、投稿者名を伏せて査読を依頼する。）

4. 原稿料は支払わない。

5. 『日本評価研究』に掲載された論文等は、その全文をインターネット上の本学会のホームページに掲載する。

6. 投稿にあたっては、原稿の種類が、①研究論文、②総説、③研究ノート、④実践・調査報告、⑤その他のうち、どのカテゴリーに入るかを明記する。ただし、カテゴリーについての最終判断は、編集委員会で行なう。「研究論文」は評価の理論構築あるいは慣行の理解について重要な学問的貢献となると認められる論文、「総説」は、評価の理論あるいは慣行について概観する論文、「研究ノート」は「研究論文」作成過程での理論的あるいは経験的な研究の中間的成果物に相当する論考、「実践・調査報告」は評価事業の実践あるいは評価にかかわる調査の報告、「その他」は編集委員会が独自に企画する特集に掲載する依頼原稿等である。

7. 投稿方法
   (1) 使用言語は日本語または英語とする。
   (2) 著者校正は原則として第一校までとする。
   (3) 英文原稿については、ネイティブスピーカーによる英文チェックを済ませ、完全な英文にして投稿すること。
   (4) 下記送付先に投稿原稿の電子ファイル（図表を入れた印刷モードのPDFとMS word）をメールで提出する。その際、連絡先（住所、Tel、Fax、Email）と原稿の種類を明記すること。掲載可と判断された原稿については、必要なリライトを経た後に、最終原稿の電子ファイル（MS word）を提出する。その際、オリジナル図表（Excel等）を添付すること。なお、編集に関する最終的

なデザインは編集委員会が行うものとする。

(5) 日本語原稿の最大文字数は以下のとおり。①研究論文20,000字、②総説15,000字、③研究ノート15,000字、④実践・調査報告20,000字、⑤その他適宜。それぞれ和文要旨を400字程度、英文要旨を150words程度、及び和文・英文でキーワード（5つ以内）を別に添付する。最大文字数20,000字の「研究論文」と「実践調査報告」の場合は、図表を含め刷り上がりが14ページ以内とする。ただし、15 ～ 16ページの2ページは著者負担で増頁できる。最大文字数15,000字の「総説」と「研究ノート」の場合は、図表を含め刷り上がりが12ページ以内とする。ただし、13 ～ 14ページの2ページは著者負担で増頁できる。左記最大ページ数（16ページと14ページ）を超える投稿は認めない。なお、刷り上がりとは、第1ページの和文要約、第2ページ以降の本文等（図表も含め、20字×43行×2段で設定）、最終ページの英文要約を入れた印刷モードのことである。

(6) 英文ではA4版用紙に左右マージン30mmをとり、10ポイントフォントを使用し、1ページ43行のレイアウトとする（1ページ約500words）。論文冒頭に150words程度のAbstractをつける。14ページでは、7,000words相当になるが、タイトルヘッド等を考慮して、最大語数を約6,000words（図表、注、文献込み）とする。図表の量によっては、ページ数が予想以上に増える場合もあり得るので、注意すること。ページ数の上限については、日本語原稿の場合と同様である。

8. 送付先

〒108-0075
東京都港区港南1-6-41
品川クリスタルスクエア12階　（一財）国際開発センター内
日本評価学会事務局「日本評価研究」編集担当
E-mail: jes.info@evaluationjp.org

『日本評価研究』執筆要領

2019.1.21改訂
2002.9.18改訂
2002.3.25改訂

1.　本文、図表、注記、参考文献等

（1）論文等の記載は次の順序とする。

日本語原稿の場合

第1ページ：表題、著者名、所属先、E-mail、和文要約（400字程度）、和文キーワード（5つ以内）
第2ページ以下（1ページ20字×43行×2段で設定する）：本文（図表含む）、謝辞あるいは付記、注記、参考文献
最終ページ：英文表題、英文著者名、英文所属先、E-mail、英文要約（150words程度）、英文キーワード（5つ以内）

英文原稿の場合

第1ページ：Title; the author's name; Affiliation; E-mail address; Abstract（150 words）; Keywords（5 words)
第2ページ以下：The main text; acknowledgement; notes; references

（2）本文の区分は以下のようにする。

例１（日本語）

1.
（1）
①
（2）
（3）

例２（英文）

1.
1.1
1.1.1
1.1.2

（3）図表については、出所を明確にする。図表は原則として、筆者提出のものをそのまま写真製版するので、原図を明確に作成すること。写真は図として扱う。

例１：日本語原稿の場合

図１ ○○州における生徒数の推移

（注）
（出所）

表１ ○○州における事故件数

（注）
（出所）

例２：英文原稿の場合

Figure １ Number of Students in the State of ○○

Note:
Source:

Table １ Number of Accidents in the State of ○○

Note:
Source:

(4) 本文における文献引用は、「‥‥である（阿部1995、p.36）。」あるいは「‥‥である（阿部1995）。」のようにする。英文では、(Abe 1995, p.36) あるいは(Abe 1995)とする。

（5）本文における注記の付け方は、（…である[1]。）とする。英文の場合は、（….[1]）とする。
（6）注記、参考文献は論文末に一括掲載する。
注記
1 ……。
2 ……。
（7）参考文献は、日本語文献は著者の五十音順、外国語文献は著者のアルファベット順に記し、年代順に記載。参考文献の書き方については以下のようにする。

日本語単行本：著者（発行年）『書名』、発行所
（例）日本太郎（1999）『これからの評価手法』、日本出版社

日本語雑誌論文：著者（発行年）「題名」、『雑誌名』、巻（号）：頁－頁
（例）日本太郎（1999）「評価手法の改善に向けて」、『日本評価研究』、1（2）：3-4

日本語単行本中の論文：著者（発行年）「題名」、編者『書名』、発行所、頁－頁
（例）日本太郎（2002）「行政評価」、日本花子『評価入門』、日本出版社、16-28

複数の著者による日本語文献：著者・著者（発行年）『書名』、発行所
（例）日本太郎・日本花子（2002）『政策評価』、日本出版社

英文単行本：著者（発行年）. *書名*. 発行地：発行所.
（例）Rossi, P. H. (1999). *Evaluation: A Systematic Approach 6th* edition. Beverly Hills, Calif: Sage Publications.

英語雑誌論文：著者（発行年）. 題名. *雑誌名*, 巻（号）, 頁－頁.
（例）Rossi, P. H. (1999). Measuring social judgements. *American Journal of Evaluation,* 15(2), 35-57.

英語単行本中の論文：著者（発行年）. 題名. In 編者(Eds.), *書名*. 発行地：発行所, 頁－頁.
（例）DeMaio, T. J., and Rothgeb, J. M. (1996). Cognitive interviewing techniques: In the lab and in the field. In N. Schwarz & S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research.* San Francisco, Calif: Jossey-Bass, 177-196.

2名の著者による英語文献：姓, 名, and 姓, 名（発行年）. *書名*. 発行地：発行所.
（例）Peters, T., and Waterman, R.(1982). *In Search of Excellence: Lessons from America's Best Run Companies.* New York: Harper & Row.

3名以上の著者による英語文献：姓, 名, 姓, 名, and 姓, 名（発行年）. *書名*. 発行地：発行所.
（例）Morley, E., Bryant, S. P., and Hatry, H. P. (2000). *Comparative Performance Measurement.* Washington: Urban Institute.

（注１）同一著者名、同一発行年が複数ある場合は、(1999a)、(1999b) のようにa,b,cを付加して区別する。
（注２）2行にわたる場合は2行目移以降を全角1文字（英数3文字）おとしで記述する。

『日本評価研究』査読要領

日本評価学会　『日本評価研究』編集委員会
2005年9月10日決定

1.　本査読要領の趣旨

本査読要領は、『日本評価研究』における掲載論文等の審査の要である査読手続きについて、投稿する会員及び査読を依頼される会員に対して解説を行い、審査手続きを効率的かつ効果的に行うことを目的として、定めるものです。

2.　査読の目的と投稿者の責任

査読は、投稿原稿が『日本評価研究』に掲載される論文等としてふさわしいものであるか否かについての判定を当編集委員会が行う上で必要とされるものです。

査読に伴って見いだされた疑問や不明な事項について、必要な場合は修正意見をつけて、修正を求めることがあります。査読は、その意味で、投稿原稿の改善に資するものでもあります。ただし、修正が求められた場合においても、論文等の内容に関する責任は著者が負うべきものであり、査読者の責に帰するものではありません。

査読者は2名で、編集委員会において学会会員の中から当該分野の専門家を選び依頼されますが、学会会員以外に依頼することもあります。

3.　査読の視点

査読は、以下の5つの視点によりますが、投稿原稿の種類によって、重点が違います。

（1）テーマの重要性・有用度
（2）研究の独自性
（3）論理の構成
（4）実証法・方法論の妥当性
（5）評価理論・実践への貢献

・研究論文の査読については、上記の5項目全てに配慮する。
・研究ノートの査読については、上記5項目のうち、特に（1），（2），（3），（4）の諸項目に配慮する。
・実践・調査報告の査読については、上記5項目のうち、特に（1），（3），（5）の諸項目に配慮する。
・総説の査読については、上記5項目のうち、特に（3）と（5）の諸項目に配慮する。

4.　投稿に当たっての留意点

2. に掲げた査読の視点以外に、基本的な論文の完成度の問題があります。例えば、

●論文等として体裁が整っているか、
●執筆要領にしたがっているか、
●簡潔明瞭に記述されているか、
●実証的なデータは適切に位置づけられているか、
●注や参考文献は本文と対応しているか、
●専門用語の使用は適切か、
●語句や文法的な誤りがないか、
●誤字脱字はないか、
●句読点に誤りはないか、
●英文要約などの英文表現は適切か、（必ずしも和文要約の直訳である必要はなく、英文としてまとまっていること）
●字数は規程に従っているか、

など、内容及び形式に関する留意点があります。

大学院生及び実務家の投稿において、論文としての体裁が整わないまま送付されている例があり、査読そのものに至らないものもあるので、しかるべき指導を受けた後に投稿されるよう強く勧めます。

5. 査読にあたっての判断事例

(1) 完成度において不十分であるが掲載を考慮できる場合

萌芽的な研究、発展が期待できる論文等は評価論の発展のためにできるだけ評価してください。

- 検証は十分とはいえないが、理論や定式化が学問の発展に有用である。
- 考察は十分とはいえないが、新たな理論の形成・促進に有用である。
- 文献調査は十分とはいえないが、研究の位置づけは明確である。
- 比較研究は十分とはいえないが、適用例としては意義がある。
- 考察は十分とはいえないが、社会的、または、歴史的に重要な事例の評価として意義がある。
- 考察は十分とはいえないが、特定の社会活動の評価として意義がある。
- 論文の構成や表現は適切とはいえないが、内容は評価できるものがある。
- 論理性は十分とはいえないが、実務上の有用性がある。
- 有意義な実践・調査報告である。

(2) 掲載を考慮するのが困難と判断される事例

- 問題意識や問題の設定が不明確。
- 基本的な用語の概念の理解や分析枠組が不明確または不適切。
- 論拠とするデータ等の信頼性が乏しい。
- 論旨の明確さや論証の適切さがない。
- 論文の構成、表現（用語、引用、図表等）が適切でない（または整合性がとれていない）。

6. 判定

掲載についての判定は以下の4つの類型に分かれ、最終的に常任編集委員会において決定します。ただし、これらの判定は、評価できる項目や問題のある項目の多少によるものではありません。(3) 及び (4) にあるように、投稿論文の種類以外であれば、掲載を考慮できるとする場合があります。別の種類となる場合、字数の関係で、大幅に修正を要することがあります。

(1) 掲載可とする。

(2) 小規模の修正による掲載可とする。

(3) 大幅な修正による掲載可とする。

但し、（総説／研究論文／研究ノート／実践・調査報告）として掲載を考慮できる。

(4) 掲載不可とする。

但し、（総説／研究論文／研究ノート／実践・調査報告）として掲載を考慮できる。

Publication Policy of the Japanese Journal of Evaluation Studies

Last revised on 15th February 2005

The Purpose and the Name

1. The Japan Evaluation Society (hereinafter referred to as "evaluation society") publishes "The Japan Journal of Evaluation Studies (hereinafter referred to as "evaluation study") in order to widely releaseevaluation studies and outputs of practical activities to domestic and international academic societies,interested individual and institutions, and contribute to the advancement and prevalence of evaluationpractice.

Editorial Board

2. The editorial board administrates editing of evaluation study based on the editorial policy stated below.
3. The editorial board is formed with less than 20 members of the evaluation society who are assigned by the board of directors. Terms of editors are two years but can be extended.
4. The editorial board assigns one editor-in-chief, two vice-editors-in-chief, and a certain number of standing editors among the members.
5. The editorial board may hold at least one meeting to discuss the editing policy, plans of editorial board, and others.
6. The editorial board reports activities to the board of directors as needed and receives approval. Also it is required to report the progress of the past year and an activity plan for the following year at the annual conference.
7. The editor-in-chief, the vice-editors-in-chief and the standing editors organize the standing committee and administrate editing on a regular basis.

Editorial Policy

8. The evaluation study, as a principle, is published twice a year.
9. The evaluation study is printed on B5 paper, and either in Japanese or English.
10. Papers published in the evaluation study are categorized as five types;

10.1. Review
10.2. Article
10.3. Research note
10.4. Report
10.5. Others

11. The qualified contributors are members of the evaluation society (hereinafter referred to as "members") and persons whose contribution is requested by the standing editors. Joint submission of members and joint submission of non-members with a member as the first author are accepted. Submission by the editors is accepted.
12. Submitted manuscripts are treated as the above categories, however, the standing editors will decide based on the application of the contributors and the following guidelines;

12.1. "Review" is a paper, which provides an overview of evaluation theory or practice. The editorial board will make the decision regarding publication.

12.2. "Article" is considered as a significant academic contribution to the theoretical development of evaluation or understanding of evaluation practice. The standing editors committee makes adoption judgments following the referee-reading process described in the next section.

12.3. "Research note" is a discussion equivalent to the intermediate outputs of a theoretical or empirical enquiry. The standing editors committee makes adoption judgments following the referee-reading process described in the next section.

12.4. "Report" is the study report related to a practical evaluation project or evaluation. The standing editors committee makes adoption judgments following the referee-reading process described in the next section.

12.5. "Others" includes requested papers for special editions organized by the editorial board and announcements from editorial board to members regarding publication.

13. The editorial board selects two referee readers. For the "article" , the editorial board makes adoption judgments referring to the results from referee readings and comments provided by one editor assigned by the editorial board. For "review", "research note", "report" and "others", the editorial board makes adoption judgments referring to the results from referee readings.

14. When editors submit a manuscript, the editors are not allowed to attend any of the standing editors committee meetings or editorial board meetings regarding the manuscript.

15. The standing editors have alternative of approval or not-approval for adoption judgment of manuscripts submitted to any categories. However exception is permitted if the editorial board approves the publication after minor rewrite. Even if the manuscripts are considered insufficient as an "article", standing editors can decide whether the manuscripts are published as a "research note" or "report" if the authors wish to publish.

Formulation and Release of Submission Procedure

16. The editorial board formulates the submission procedure based on the editorial policy described above and release after approval from the board of directors.

Distribution

17. The evaluation study is distributed to all members for free and distributed to non-members for a charge.

Distribution of the Printed Manuscript

18. 30 copies of the respective paper are reprinted and distributed to the authors. The authors must cover any costs incurred by author's requests for printing more than 30 copies.

Release on the Internet

19. The papers published in the evaluation study are released on the internet with approval from the authors.

Copyright

20. Copyright of papers which appear in the evaluation study is attributed to the respective authors. Editorial right is attributed to the evaluation society.

Office

21. The office is in charge of administrative works for editing, distribution, and accounting.

Information for Contributors
(For English Papers)

Last revised on 20th July 2017

1. "The Japanese Journal of Evaluation Studies" is the publication for reviews, articles, research notes, and reports relating to evaluation.

2. "The Japanese Journal of Evaluation Studies" is primary published to provide opportunities for members of the Japan Evaluation Society (hereinafter referred to as "members") to exchange findings, and to contribute to further development of the study of evaluation both domestically and internationally. As a principle, this journal publishes the contributions submitted by the members. With the exception of requested papers, the first author must be a member. A submission (as the first author) is limited to one manuscript that has not been published or submitted in any form for another journal of academic association etcetera.

3. Adoption judgments of the manuscript are made at the discretion of the editorial board. Comments from two referee readers who are appointed for every manuscript are referred to in the screening process (the editorial board requests referee readers without notifying the author of manuscript).

4. Payment for the manuscript is not provided.

5. Papers published in "The Japanese Journal of Evaluation Studies" are released on the Internet at homepage of this academic society.

6. Regarding submission, manuscripts must be identified as one of the following categories: 1) article, 2) review, 3) research note, 4) report, and 5) others. However, the final decision of the category is made by the editorial board.
   "Article" is considered as a significant academic contribution to the theoretical development of evaluation or understanding of evaluation practice.
   "Review" is a paper which provides an overview of evaluation theory or practice.
   "Research note" is a discussion equivalent to the intermediate outputs of a theoretical or empirical study in the process of producing an "article".
   "Report" is the study report related to a practical evaluation project or evaluation.
   "Others" are manuscripts for special editions requested by the editing committee.

7. Manuscript Submission
   (1) Manuscripts may be written in either Japanese or English.
   (2) Correction by the author is only for the first correction.
   (3) English manuscripts should be submitted only after the English has been checked by a native speaker.
   (4) Submit manuscripts via email. Contact information including mailing address, telephone number, fax number, e-mail address, and the category of the manuscript should be clearly stated.
   For approved manuscripts, after necessary rewriting, the author needs to submit the final paper via email. Original figures, charts, and maps should be provided.
   (5) Total printed pages should not exceed 14 pages. Any cost incurred by printing more than 14 pages must be covered by the author.
   (6) The layout for English papers should be 30 mm of margin at left and right side, 10pt for font size, 43 lines on A4 paper (about 500 words per page). An abstract of 150 words should be attached to the

front. 14 pages are equivalent to 7,000 words but the body should not exceed 6,000 words to allow for the title, header, figure, chart, footnotes, and references. Please note that the number of pages may be more than expected depending on the number of figures included.

8. Mailing address

Office of Japan Evaluation Society at International Development Center of Japan
Shinagawa Crystal Square 12th Floor, 1-6-41 Konan, Minato-ku, Tokyo,
108-0075, Japan
E-mail: jes.info@evaluationjp.org

Writing Manual of the Japanese Journal of Evaluation Studies (For English Papers)

Revised on 18th September 2002

1. Text, Charts, Figures, Graphs, Diagrams, Notes, and References

(1) The paper should be written in the follow order:

First page: Title; the author,s name; Affiliation; E-mail address; Abstract (150 words); Keywords (5 words)

Second page: The main text; acknowledgement; notes; references

(2) Section of the text should be as follow:
1.
1.1
1.1.1
1.1.2

(3) Source of the charts, figures, graphs, and diagrams should be clarified. Submitted charts and others will be pzhotoengraved, therefore it is important that the original chart is clear. Pictures shall be treated as figures.

Figure 1 Number of Students in the State of ○○

Note:
Source:

Table 1 Number of Accidents in the State of ○○

Note:
Source:

(4) Citation of literature in the text should be, (Abe 1995, p.36) or (Abe 1995).
(5) Note in the text should be, (------.[1])

(6) Note and references should be written all together in the end.
Note
1 --------.
2 --------.

(7) Reference should list the literature in alphabet order, and arranged in chronological order. Follow the examples:

Book: author (year of publication). *Title of the book.* Published location: publishing house.

(e.g.) Rossi, P. H. (1999). *Evaluation: A Systematic Approach 6th edition.* Beverly Hills, Calif: Sage Publication.

Article from magazine: author (year of publication). Title. *Title of the magazine,* volume (number), page-page.

(e.g.) Rossi, P. H. (1999). Measuring social judgments. *American Journal of Evaluation,* 15(2), 35-37.

Article in Book: author (year of publication). Title. In editor (Eds.), *Title of the book.* Published location: publishing house, page-page.

(e.g.) DeMaio, T. J., and Rothgeb, J. M. (1996). Cognitive interviewing techniques: In the lab and in the field. In N. Schwarz & S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research.* San Fransisco, Calif: Jossey-Bass, 177-196.

Book by two authors: surname, first name, and surname, first name. (year of publication). *Title of the book.* Published location: publishing house.

(e.g.) Peters, T., and Waterman, R. (1982). *In Search of Excellence: Lessons from America,s Best Run Companies.* New York: Harper & Row.

Book by more than three authors: surname, first name, surname, first name, and surname, first name. (year of publication). *Title of the book.* Published location: publishing house.

(e.g.) Morley, E., Bryant, S. P., and Hatry, H. P. (2000). *Comparative Performance Measurement.* Washignton: Urban Institute.

(note 1) If some references are from the same author with the same publication year, differentiate by adding a,b,c as (1999a), (1999b).

(note 2) If the reference is more than a single line, each line from the second should be indented by three spaces.

(e.g.) DeMaio, T. J., and Rothgeb, J. M. (1996). Cognitive interviewing techniques: In the lab and in the field. In N. Schwarz & S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research.* San Fransisco, Calif: Jossey-Bass, 177-196.

Referee-Reading Guideline

The Japanese Journal of Evaluation Studies Editorial Board,
The Japan Evaluation Society
Approved on 10th September 2005

1. Content of the Referee-Reading Guideline

This Referee-Reading Guideline is to provide explanation of the main publication judgment, procedure of the referee-reading, to the members who submit the manuscript and for the members who are requested to conduct referee-reading in order to carry out the procedure efficiently and effectively.

2. Purpose of Referee-Reading and the Responsibility of the Author

Referee-reading is necessary for the editorial board to make decisions of whether submitted manuscripts are appropriate to publish in the Japanese Journal of Evaluation Studies or not.

If there is doubt or obscurity identified in manuscripts during the referee-reading corrections may be required. Therefore, referee-reading also contributes to the improvement of the submitted manuscripts. However, although the manuscripts are requested corrections, the author is still solely responsible in regards to the contents and it is not attributed to the referee-readers.

Referee-readers are two persons who are requested by the editorial board depending on the specialty or the field of the submitted manuscript. People who are not members of this academic society also may be requested.

3. Items of Consideration in Referee-Reading

Five points are considered in referee-reading, however, the importance of each may be different depending on the type of manuscript.

(1) Importance and utility of the theme
(2) Originality of the study
(3) Structure of the logic
(4) Validity of verification and methodology
(5) Contribution to evaluation theory and practice

- For the article, all of above five are considered.
- For the research note, especially (1), (2), (3), and (4) are considered.
- For the report, especially (1), (3), and (5) are considered.
- For the review, especially (3) and (5) are considered.

4. Attentions in submission of manuscript

Besides above five viewpoints, basic completeness as a paper is also considered, for example;

- appearance of the paper is organized
- written according to the writing manual
- described simply and distinctive
- verification data is appropriately used
- notes and references are corresponding with the text
- terminology is appropriately used
- no wording and grammatical mistakes
- no errors and omission
- no punctuation mistakes
- expression in English abstract is appropriate
- word count is according to the manual

The above mentioned forms and contents are also considered. There have been cases in which graduate students and practitioners posted without organizing the manuscripts as a paper. On those occasions, referee-reading was not conducted. Necessary consultation is strongly recommended prior to submission.

5. Judgment Cases in Referee-Reading

(1) In the case of the manuscript which is considered acceptable for the publication but is not yet complete:
The referee reader should evaluate carefully whether the paper can contribute to the development of evaluation theory or evaluation studies.

- Verification is lacking but the theory and formulation are useful for academic development.
- Analysis lacking but useful for formation and promotion of new theory.
- The literature review is not of a high standard but, the overall study is meaningful.
- Comparative study is not up to standard but is meaningful as an example of application.
- Analysis is lacking but it is meaningful as an evaluation of socially and historically important cases.
- Analysis is lacking but it is meaningful as an evaluation of particular social activities.
- Organization and expression are not up to standard as a paper but the contents are worthy to evaluate.
- Logic is not strong enough but useful in practice.
- The paper has significance as a report.

(2) In case of the manuscript which is considered as difficult for publication:

- Awareness of the issue or setting of the problem is indecisive.
- Understanding or analytical framework of notion of basic terminology is indecisive or inappropriate.
- There is a lack in credibility of data for the grounds of an argument.
- There is no clear point of an argument or appropriateness of proof.
- Organization of the paper and presentation (terminology, citation, chart, etc) are inappropriate (or not consistent).

6. Judgment

The final decision will be made on publication at the standing editors committee following one of four patterns (listed below). However, these judgments are not based on the number of errors but on the strength of the overall report. In the case of (3) and (4), there is a possibility to be published as a different type of paper. If it is published as a different type of paper, major rewrite concerning the number of words may be required.

(1) The paper will be published.
(2) The paper will be published with minor rewrite.
(3) The paper will be published with major rewrite, however as a different type of paper (review, article, research note, or report).
(4) The paper will not be published; however there is the possibility that it will be published as a different type of paper (review, article, research note, or report).

～投稿案内～

日本評価学会では、「日本評価研究」掲載のための投稿原稿を募集しております。投稿の締め切りは9月末日（翌年3月刊行）及び3月末日（9月刊行）です。

ご興味をお持ちの方は投稿規定・執筆要領・査読要領、並びに原稿見本をご参照のうえ、奮ってご投稿ください。http://evaluationjp.org/activity/journal.html

投稿先：特定非営利活動法人日本評価学会編集委員会事務局
E-mail： journal@evaluationjp.org

『日本評価研究』第20巻第2号
2020年７月31日

編集・発行　特定非営利活動法人日本評価学会編集委員会事務局
〒108-0075　東京都港区港南1-6-41
芝浦クリスタル品川12階
一般財団法人国際開発センター内

印　　刷　株式会社　研恒社

© 日本評価学会

本誌に掲載されたすべての内容は、日本評価学会の許可なく転載・複写はできません。

ISSN 1346-6151

Japanese Journal of Evaluation Studies

Vol. 20, No. 2, July 2020

CONTENTS

Special Issue: The Current Status and Issues on Evidence-Based Policy Making（EBPM）

Editor's Note: The Current Status and Issues on Evidence-Based Policy Making（EBPM） — Ryo Sasaki, Tomoya Masaki

Reporting Guidelines for Evidence-Based Policymaking（EBPM）and Improving Research Paper Quality — Tomoya Masaki, Kiichiro Tsutani

Furthering Evidence-Based PolicyMaking in Japan: Implications from experiences in the healthcare sector — Satoko Tanabe

Current Status and Issues of Evidence-Based Policy Making in Japan — Yohei Kobayashi

Study on Evidence-Based Policy Making（EBPM）in practice in the Government of Japan — Yutaka Sunayama

Current status and issues for Institutionalizing Evidence-Based Policy Making（EBPM）in Japanese Local Governments — Koji Nishihata

The characteristics of evidence institutions in education — Toshiro Mori, Yoshihiro Okazaki

The current status of Evidence-Based Policy Making（EBPM）at international development field — Ryo Sasaki

Article

Evaluation of Municipality Mergers and Cooperation as Policy Tools Based on the Survey Results Regarding the Local Revitalisation Policy in Japan — Yuichi Murakami

Report

MSC（Most Significant Change）, participatory and qualitative evaluation method - Consider three cases in Japan focusing on learning and improvement objective evaluation - — Hiroshi, Tanaka

Report of the 20th Annual Conference

Announcement for the 21st Annual Conference

Japan Evaluation Society