

博士論文

自治体経営と政策評価

—現場発のボトムアップ型改革の可能性—

後藤 玲子

(戸籍名：磯崎玲子)

目次

序章 より良い自治体経営とは何か—本論文の目的と射程	1
1. より良い自治体経営とは何か	1
1-1. 目的要件—人々の潜在能力（生き方の幅）の拡大	2
1-2. ガバナンス要件—公平性と開放性	4
1-3. パフォーマンス要件—有効性と効率性	6
2. 本稿の目的と射程	7
3. 日本の政策評価をめぐる先行研究と比べた本稿の特色	10
4. 本稿の構成	13
第1章 自治体ニーズ評価の理想と実態	15
1. ニーズ評価とは何か	15
1-1. 政策評価とは	15
1-2. 政策のロジック・モデルと5つの評価活動	17
1-3. 政策プロセスと政策評価	22
2. 理想的なニーズ評価とは何か	23
3. 自治体ニーズ評価の理想と実態のギャップの検証方法	26
3-1. 第一要件の検証方法	26
3-2. 第二要件の検証方法	29
4. 自治体ニーズ評価の実態と課題	34
4-1. 第一要件の検証結果と考察	34
4-2. 第二要件の検証結果と考察	38
4-3. 検証結果のまとめ	42
5. 自治体ニーズ評価の理想と実態のギャップを埋めるために	43
5-1. ニーズ評価の理想と実態のギャップが生じている原因	43
5-2. ニーズ評価の意義（1）—貧困の発見と再発見	44
5-3. ニーズ評価の意義（2）—柔道事故の可視化のインパクト	48
6. 事例分析から得られる示唆	49
第2章 達成目標が充たすべき要件	51
1. 達成目標が充たすべき要件は何か	51
1-1. 達成目標が充たすべき要件（1）—政策科学の文献に基づいて	51

1-2. 達成目標が充たすべき要件（2）—社会科学のエビデンスに基づいて	54
2. 岡山県総社市「障がい者千人雇用事業」の事例分析	57
2-1. 千人雇用事業を取り上げる理由	57
2-2. 千人雇用事業の概要と主要な成果	62
2-3. 千人雇用事業のインパクト評価方法	70
2-4. 千人雇用事業のインパクト評価結果	76
2-5. 千人雇用事業のインパクト評価結果の頑健性評価	77
2-6. 千人雇用事業の目標と達成目標が充たすべき7要件	86
3. 小括	88
第3章 バイアスとエビデンス	90
1. エビデンスをつくり、つたえ、つかう—エビデンスに基づく医療の取組み	90
1-1. エビデンスに基づく医療とは	91
1-2. 最良のエビデンスをつくり、つたえる—コクランの5つの特徴	93
1-3. エビデンスの確実性の評価と推奨度の評価の峻別	96
2. エビデンスに基づくインパクト評価の実践方法	99
2-1. 効果が確かな政策を選ぶ—エビデンスを「つかう」	100
2-2. 介入プロセスと介入結果を評価する—エビデンスを「つくる」	103
2-3. データやエビデンスをつかってもらう—エビデンスを「つたえる」	105
3. 個人と組織の認知バイアスに対処する	106
3-1. 政策に基づく証拠形成と集団浅慮	106
3-2. 確証バイアスによる判断ミスを組織的に防ぐ方法	107
3-3. 予想と現実のギャップを可視化して、バイアスの自覚を促す	109
3-4. 認知バイアスに組織的に対処する	115
4. 政策プロセスの透明性と開放性の大切さ	119
5. 小括	121
第4章 目標を達成するまでやり抜く組織の構築方法	124
1. 成功する人の特徴	125
2. 組織が失敗から系統的に学習するためにはどうすればよいか	129
3. 失敗から学習する能力を組織的に高める方法	132
3-1. 体験型の研修と採用基準の変更	132

3-2. マインドセットのリフレーミング	133
3-3. 失敗からの学習を促す業務システムへの変更	134
4. 失敗から進んで学習する意欲を組織的に高める方法	140
4-1. 理性ではなく感情に働きかける	141
4-2. 自律性の感覚を重んじる	142
4-3. 成果を素早く実感できるようにする	145
4-4. 外発的動機づけを用いる際に注意すべきこと	145
5. 小括	146
終章 結論と今後の研究課題	148
1. 結論	148
2. 今後の研究課題	150
3. 謝辞	151
参考文献	154

序章 より良い自治体経営とは何か―本論文の目的と射程

本稿の目的は、より良い自治体経営を実現するために地方自治体は何をすべきなのかを明らかにすることにある。その際、本稿のアプローチの特徴は、自治体経営の状態を測定評価するための地方自治体の意欲と能力に焦点を当てる点にあると言えるだろう。すなわち本稿は、より良い自治体経営を促す政策評価システムをどのように構想し実現すべきかを明らかにしようとするものである。そのためにはまず、より良い自治体経営というときの「より良い」とは何を意味するのかを明確にしておかなければならないだろう。より良い自治体経営とは、住民の平均的幸福度をできるだけ高めるような自治体経営のことだろうか。それとも、地球環境により優しい共生社会を実現するような自治体経営のことだろうか。あるいは、最も貧しい人々のニーズを最優先する自治体経営を指すのだろうか。本稿の目的は自治体経営のあるべき姿を道徳原理から規範的に論ずることではなく、より良い自治体経営を実現する方法について社会科学の方法論に基づいて論理的・実証的に論ずることであるが、本稿が依拠する福祉観を語らずに理想を達成する方法について論じては、目的と手段の倒錯を引き起こしかねない。よってまず、本稿の考えるより良い自治体経営とは何かについて簡潔な説明を与える。次に、本稿が政策評価に焦点を当ててより良い自治体経営の実現方法について検討する理由を述べる。続いて、日本の政策評価をめぐる先行研究と比べた本稿の特色を整理し、最後に本稿の構成を述べる。

1. より良い自治体経営とは何か

本稿でいうより良い自治体経営とは、(a)人々の潜在能力をより拡大することを目指し、その目標を、(b)より公平でより開かれた方法で、(c)より効果的・効率的に達成する自治体経営を指す。以下では、第一の要件(a)を「目的要件」、第二の要件(b)を「ガバナンス要件」、第三の要件(c)を「パフォーマンス要件」と呼ぶ。3つの要件は目的―手段関係にあるのではなく、すべてより良い自治体経営を実現するための基本目標であり、内在的価値をもつ。以下では対立する福祉観に言及しながら、なぜ本稿は(a)から(c)の要件を備えた自治体経営を望ましいと考えるのかを説明していこう。

1 - 1. 目的要件一人々の潜在能力（生き方の幅）の拡大

第一に、本稿でいうより良い自治体経営は、結果としての富や幸福よりも、人々の本質的な自由を拡大する潜在能力に着目する。17 世紀に確率論を創始した早熟の天才 Blaise Pascal は、例外なくすべての人間の行為の動機は幸せの追求であると考えた¹。人々の人生の目的が幸せに生きることにあるなら、市民をより幸せにすることを自治体経営の目標にすべきなのかもしれない。実際に、政府は人々の幸福の追求を目指すべきだとする考えは古くから存在していた。例えばフランス革命の最中に国民公会で採択された 1793 年憲法 (Constitution du 24 juin 1793) は、「社会の目的は一般の幸福である。(Le but de la société est le bonheur commun.)」と宣言し²、功利主義の創始者と言われる Jeremy Bentham は、最大多数の最大幸福の保証が政府の最重要課題だと主張した³。

社会全体の効用を最大化することこそが【善】だという功利主義者の考えは、個人の主観的効用を個人間で比較して集計できるという前提に基づいていた⁴。しかしそれは実証可能な命題ではないという Lionel Robbins らの批判を受けて⁵、主流派経済学は 1930 年代以降、パレート効率性という新しい厚生基準を経済システムの【善】の基準として採用した。その結果、個人の主観的効用の基数的性質に基づいて需要関数を導出するという考え方は辛うじて費用便益計算の分野で生き残ったものの、それ以外の経済学の分野ではすっかり影を潜めていた⁶。

¹ Gilbert (2006) pp. 34-35 (訳 p. 70) を参照。

² Bok (2010) p. 5 を参照。

³ Bentham (1891) pp. 93-125 を参照。

⁴ 「善」と「正」という道德概念を【 】で囲む表記法は鈴木 (2004) に倣った。

⁵ Robbins (1935) Chapter 3 を参照。

⁶ パレート効率性とは、集団内の誰かの効用を犠牲にしなければ、他の誰かの効用を高めることができない状態のことである。パレート効率性の規範的妥当性については、奥野編 (2008) 第 3 章を参照。Vilfredo Pareto が経済生活を諸要素の相互依存的連関態として記述することによって価値概念を回避し、意識的な経済活動を行う当事者としての人間を排除したことという指摘については、須藤 (1988) 第 2 章を参照。なお Sen (1985) は、パレート効率性を厚生基準とする顕示選好学派は、個人間の効用比較不可能性ゆえではなく、第一に、福祉の個人間比較を容易に許さないアプローチでは福祉概念を実体化することが難しいという理由から、第二に、後述するように効用は人々が逆境や不遇に適応した結果しか測定できないという理由から、福祉の経済学としての適格性を欠くと指摘している (Sen [1985] Chapter 2, 3, 4 and 7 を参照)。

そのような状況が大きく変化し、経済学で幸せの決定要因に関する研究が盛んに行われるようになったのは、1990年代後半のことであった⁷。幸福の経済学（happiness and economics）は、個人の主観的幸福度を福祉の尺度として用いるという古典的な考え方を復活させた。すなわち幸福の経済学は基本的に、「われわれが善と呼ぶものは、ほとんどがわれわれを幸せな気持ちにするいいもの」⁸であるという前提に立って、人々が自己申告する主観的幸福度を【善】の指標として用いる。幸せは多面的で掴みどころがないことを認めつつも、測定手法を工夫しながら何度も繰り返し大量に測定することによって人々の幸福度を科学的に測定評価できるという信念に基づいて、幸せの決定要因を明らかにしたり、実質所得に代わる新しい豊かさ指標を開発するという作業に挑んでいる⁹。

しかし、一時的な刺激や興奮に簡単に左右される主観的な幸福度を自治体経営の目標に据えるという考えは、本当に適切なのだろうか。ここで問題としているのは、測定バイアスの問題ではない。当てにならない主観的な感情を社会の目標にすべきかという問題である。長く停滞していた厚生経済学（welfare economics）に新たな地平を切り拓いた Amartya Sen は、幸福度を福祉の目標とすることに度々異論を唱えてきた。彼によれば、J. Bentham に代表される伝統的な功利主義者の効用—それは幸福ないし快楽または欲望充足を意味する概念である—は、福祉にとって手段的価値しかもたない富裕—財貨に対する人の支配権—を福祉の目標に据えるよりはましである。しかし第一に、人が快楽や苦痛を得る物理的条件を無視しているために、厳しい状況におかれた人々の欠乏や抑圧をむしろ覆い隠してしまうという点において、第二に、人がどのような生き方や在り方を幸福と考えるかという内省に基づく主観的評価に関わる概念ではなく、単なる感情を表すに過ぎないという点において、二重の欠陥がある。そこで彼は、効用や富裕に代わって福祉一般を評価する新しい尺度として、「潜在能力」（capability）という概念を提唱した。潜在能力とは、人間が潜在的に達成可能な生き方や在り方を反映する機能の組み合わせの集合—「生き方の幅」—のことである¹⁰。効用や富裕では選択の結果から【善】を判断する他ないが、潜在能力アプローチでは、人が自ら価値を認める生き方や在り方を実現する機会やプロセスをも

⁷ 幸福研究の歴史については Frey (2008) Chapter 1 を参照。

⁸ Gilbert (2006) p. 71 (訳 p. 124) より引用。

⁹ 幸福の経済学については主に、Frey and Stutzer (2002); Gilbert (2006); Frey (2008); Bok (2010)を参照。

¹⁰ Sen (1985)。capability に対する「潜在能力」という訳語については鈴木(2004)を、「生き方の幅」という訳語については Roemer (1996) 訳 p. 192 を参照。

【善】とみなすことができる。測定評価の難しさがあっても、自治体経営が福祉の究極的な目的に掲げるべきなのは、住民の平均的幸福度や平均実質所得の向上ではなく、住民と市民の人生の生き方や在り方を反映する機能の組み合わせを多面的に把握して、彼女／彼らの潜在能力を継続的に高めることだと本稿は考える。なお住民と市民としたのは、自治体経営では、その自治体の行政区域に居住する住民の利益のためだけではなく、これから生を享ける人々や他の地域に住む人々の利益を考慮に入れることが必要だと考えるためである¹¹。

1－2．ガバナンス要件－公平性と開放性

第二に本稿は、人々の本質的な自由を拡大するという目標と、それをより公平かつ開放的な方法で実現するという目標を同じくらい重要な目標だと考える。結果さえ良ければガバナンスがどうであれ構わないという立場をとらないのは、限られた集団に権力と富が集中している収奪的な政治経済制度の下で善い目的に向けた前進がみられたとしても、その結果のみをもってより良い自治体経営が行われたとみなすのは適当ではないと考えるからである¹²。基本的人権の保障を高らかに謳う社会において人種差別やジェンダー差別が存在し続けてきた歴史を振り返れば、ガバナンス要件を目的要件とは別個の独立した目標にすることの意義は明らかであろう。

他方で本稿は、ガバナンスの正しさが目的の善さに優先されとも考えない。これは例えば、1990年代以降の中国における絶対的貧困者の減少をどのように評価するかという問

¹¹ 佐々木（2010）は、専門的な政策評価においては住民と市民を峻別した上で、その双方の価値観を含むことが重要であるとする。例えば「女性は教育を受けるべきではない」と考える人々の多く住む地域で、女性が排除された教育制度を住民目線で評価することが妥当かという問題を考えれば、特定地域の住民だけの価値観ではなく、より広い範囲の市民の価値観を常に併せて検討する必要があることが分かるだろう（同書, pp. 28-33）。本稿も同じ立場に立つ。なお佐々木（2010）における専門的な政策評価とは、個人の価値観に基づく評価ではなく、評価に独特の論理を身につけた者による評価という意味である（同書, p. 26）。

¹² 「収奪的な政治経済制度」という概念については、Acemoglu and Robinson（2012）を参照。「収奪的な政治制度」とは限られた人々が権力を掌握し、その権力の行使にほとんど制約がない政治制度のことであり、「収奪的な経済制度」とは社会の中のある集団から収奪し、別の集団に利益をもたらすために設計されている経済制度のことである。対概念は「包括的な政治経済制度」である。

題に関わる。中国では、市場開放が進んだ 1993 年から 2012 年の 20 年間で人口に占める絶対的貧困者の割合が 57%から 6.5%に激減した¹³。表現の自由や政治的自由のような基本的な自由が承認されていない状況で成し遂げられた成果を手放しで称賛することはできないが、深刻な社会的剥奪状態にあった人々の基礎的ニーズが充たされ、選択の自由が拡大するような変化を前進ではないと評価することも適切ではないだろう。

ガバナンス要件としては、人々の潜在能力を広げるためのルールと機会が実質的に公平で、その決定プロセスが多様な人々に開かれていることが重要だと考える。民主的であるということの意味は多義的で、安易な民主化はむしろ弊害を招くことがありうるため、より民主的であることを要件にはしない¹⁴。多様性ではなく開放性を重視する理由は、結果としての多様性が実現しているかどうかよりも、多様性を積極的に許容する状態であることが重要だと考えるためである。

ネットワークの開放性は社会の創造力の源であり、人々が多様な個性を豊かに発展させる源でもありうる。須藤（1995）によれば、個性とは複数の活動領域の特性とその特性を関係づける様式によって規定され、創造性とは複数の活動領域の情報を能動的に関係づけ、さらには関係づける様式を変化させることによって発揮される。それゆえ、人々が豊かな個性と創造性を発揮するためには、ネットワークを複合化して諸主体間・諸ネットワーク間で相互作用・相互浸透が起きるような仕組みを構築する必要がある¹⁵。そこで大切なのは、ネットワークを何層にも複合化して参加主体が多様なネットワークに重複参加することではない。各ネットワークがノイズを受け入れる開放性を持ち、各参加主体が各ネットワークの開口部からもたらされる新しい情報を主体的に受け入れ、新しい関係性を自律的に構築することが相互承認されていることである。

参加主体の自律性を重んじる開放的なネットワークには、地域社会の自殺を予防する効果もある。日本有数の自殺希少地域である徳島県の旧・海部町（現・海陽町）は、大都市

¹³ ここでいう絶対的貧困者とは、2018 年現在の世界銀行の定める絶対的貧困ラインである一日 1.9 米ドル未満で暮らす人々のことである。数値は世界銀行の貧困データベースに基づく（Retrieved September 13, 2018, from <http://povertydata.worldbank.org/>）。中国における絶対的貧困者の減少については、Ravallion and Chen（2007）も参照。

¹⁴ 1990 年代以降の最底辺の 10 億人の国における選挙の普及という形態での民主主義が政治的暴力の危険性を高めたという関係性については、Collier（2009）を参照。民主的な集約ルールが多様性と多数決に基づく議会制民主主義の弊害については、フィッシュキン（2011）；坂井（2013）及び同（2015）に詳しい。

¹⁵ 須藤（1995） pp. 1-23 及び pp. 190-219 を参照。

と比べれば参加主体の多様性に欠けるかもしれないが、地域コミュニティの排他的傾向は小さく、そこで築かれている人的ネットワークは開放的かつ複合的で、緩やかな紐帯と呼べるような心地よい密度が保持されているという。そのため旧・海部町の地域コミュニティには、多様な個性を即時にありのままに受け容れる柔軟性と、問題を内に秘めずオープンに語る文化があり、人々の自己効力感が高く、地域の自殺率は低いのだという¹⁶。本稿では、このような開かれた社会の在り方を指して開放性という用語を用いる。

1-3. パフォーマンス要件—有効性と効率性

最後に本稿では、自治体経営が効果的・効率的であるというパフォーマンス要件を、目的要件とガバナンス要件と並ぶ3つ目の自治体経営の基本目標とみなす。パフォーマンス要件は「人々の潜在能力をより拡大する」という目的要件に既に含まれていると解釈することもできるが、目的が妥当かという問題と目的に沿った成果が出ているかという問題は切り離すことが可能であるし、目的要件は政策目的の妥当性に関係し、ガバナンス要件は政策プロセスの妥当性に関係し、パフォーマンス要件は政策の有効性と効率性に関係すると位置付けると、第1章で述べる政策のロジック・モデルに基づく政策評価と整合的に考えることが容易になる。

効率性とは単に無駄のないことを表すわけではなく、コスト見合いの効果が大きいという相対的な有効性を表す概念であるが、自治体経営の効率化では効果を測定しにくいことも影響して費用削減ばかりが強調される傾向にある。そのため、効率性を高めるためにはコストを節約することと効果を高めることの両方が大事だということが想起されやすくなるように、有効性と効率性の両方をパフォーマンス要件とする。

自治体経営の望ましさについて実証的に分析評価する場面では、所与の目的と所与のガバナンスにおけるパフォーマンスを比較評価することが多くなるかもしれない。しかし基本的には、価値観の議論と切り離してパフォーマンスの良さだけを論じるべきではない。それは例えば、法定受託事務の評価でも同様である¹⁷。法定受託事務では所与の目的を所

¹⁶ 岡（2013）を参照。同書で描かれている旧・海部町は、本稿の考えるより良い自治体経営が目指すべき地域社会を考える上で示唆に富む参考例の一つである。

¹⁷ 「法定受託事務」とは、地方自治体が処理する事務のうち、中央政府が本来果たすべき役割に関わるものであって、中央政府においてその適正な処理を特に確保する必要があるものとして法律またはこれに基づく政令に特に定めるもののことである。都道府県から市

定の方法で首尾よく達成することが求められるが、その事務処理において業務のパフォーマンスだけを追求すると、手続きの正確さや濫給の防止といった行政コストの削減に繋がる課題ばかりが認識され、福祉の受給資格があり手を差し伸べるべき人への漏給の抑止という社会の【善】や【正】に繋がる課題が見えにくくなってしまう。それは、潜在能力の拡大という観点からも、より公平で開放的な福祉ガバナンスという観点からも、望ましいことではないだろう。法定受託事務のような業務についても自治体経営の観点からより上位の目標と関係づけて運用すべきであり、目的やガバナンスを問わずにパフォーマンスを高めるだけでより良い自治体経営が行われているとは判断できないと本稿では考える¹⁸。

2. 本稿の目的と射程

本稿では、(a) 人々の潜在能力（生き方の幅）をより拡大するという目的を、(b) より公平で開放的な方法で、(c) より効果的・効率的に達成する自治体経営を、より良い自治体経営だと考える。では日本の自治体経営は、より良い方向へと進んでいるのだろうか。

この問いに答えるための研究アプローチは、大別して2種類ありうる。1つ目は、目的要件、ガバナンス要件及びパフォーマンス要件を操作化し、自治体経営の状態を測定して、経時的・横断的に比較評価するという方法である。例えば目的要件である潜在能力の測定指標については、国連の人間開発指数（Human Development Index：HDI）を嚆矢として様々な社会指標が提案されている¹⁹。どの指標を用いればより適切に社会進歩を測定できるかは、理論的にも実務上にも大きな問題になりうる。しかし、完璧な社会指標も完璧な評価もありえない。時間その他の資源制約を鑑みながら改善努力をし続けることを前提として測定指標を選び、測定評価を試みることによって初めて、自治体経営の状態は明らかになってくるものであろう。例えば、国内総生産（Gross Domestic Product：GDP）に

町村への法定受託事務も存在する。

¹⁸ 政策目標を合規性や行政活動の効率性ではなく、標的集団にとってのアウトカムに変えることで、福祉業務の在り方は刷新される可能性がある。かつて、米国オレゴン州の低所得家庭支援政策では、窓口に申請に来た人に対して遅滞なく正確な事務を行うことが重要とされていたが、政策目標を当事者目線で見直すことにより、福祉の潜在的受給者にアウトリーチし、人々の自立生活力を高めることが目指されるようになったという（米国行政学会・行政経営センター〔2001〕第4章を参照）。

¹⁹ 社会指標については、例えば Atkinson, et al. (2002) ; Stiglitz et al. (2009) を参照。

は、市場で取引されない価値が的確には反映されないという問題や、本質的に経済効率を測る指標であって所得分配の公正性を測る指標ではないといった限界のあることがよく知られている。潜在能力等を測るための様々な社会指標が開発されていれば、地方自治体が域内総生産を自治体活動のアウトカム指標に用いる際に、域内総生産では捉えきれない価値に目を向けることが容易になる。したがって社会経済の発展度合いを多面的に測定評価するための社会指標を開発するという研究アプローチには、大きな意義があると考えられる。

しかし一方で、指標選定の恣意性をどのように排除すればよいかという問題や、社会指標で測られるアウトカムと自治体経営との関係性をどのように特定すればよいかという問題がある。指標選定の恣意性について考えてみよう。潜在能力の測定指標については、近年では Joseph Stiglitz や A. Sen のような過去にノーベル経済学賞を受賞した学者らが 2009 年に公表した『経済パフォーマンスと社会進歩の測定に関する委員会報告書 (*Report by the Commission on the Measurement of Economic Performance and Social Progress*)』にいくつかの提案がなされている²⁰。権威のある学者が推奨している社会指標こそが潜在能力を測定評価するのに最も適した社会指標であるわけではないが、潜在能力を拡大する要因の中には人々が自覚できない経路がありうるから、何をもって生き方の幅が拡大したと言えるかを市民に選んでもらうだけでは心許ない。したがって指標選定に当たっては、先行研究を参考にしながら住民と市民の生き方の幅を規定する要因を実証的に調べ上げ、それを最もよく表す指標をいくつか選び、現実には人々の生き方の幅が拡大したことを適切に測れているかどうかを検証するという作業を繰り返すことになるだろう。

ガバナンス要件についてはどうだろうか。政府のガバナンスの測定指標としては、例えばトランスペアレンシー・インターナショナル (Transparency International) という非政府組織が 1995 年から年に一回調査・公表している「腐敗認識指数 (Corruption Perceptions Index : CPI)」という社会指標がある。腐敗認識指数とは、公共部門の汚職度合いに関する、ビジネスパーソンやその国の専門家による認識を調査した複数の情報源からのデータを集約して作成された総合指標で、政府の腐敗度を測る指標として世界で広く用いられている。例えば 2018 年版では、過去 2 年間に公表された 12 の独立した機関による 13 の情報源からのデータを集約して腐敗認識指数が作成されている²¹。事実を表すデータではな

²⁰ Stiglitz et al. (2009) Chapter 2 を参照。

²¹ Transparency International(2018) *Corruption Perceptions Index 2018*, Retrieved July 1,

く認識を表すデータが用いられているのは、汚職は通常意図的に隠されており明るみに出ることは稀であるため、客観的な汚職度合いを直接的に測定することは難しいからである。この例から分かるように、ガバナンス指標の場合には認識ではなく事実を表すデータで測ることが容易ではないため、調査対象や測定方法の恣意性はより深刻な問題になりうる。日本の地方自治体の公平性や開放性を測る指標については、管見によれば一般的によく用いられている社会指標はないから、調査を何度か繰り返して安定的な結果を得られる社会指標を作るという作業から始めなければならないだろう。

パフォーマンス要件についてはどうだろうか。前節で述べた通りパフォーマンス要件は潜在能力要件に含まれるとも解釈できるので、アウトカム指標には潜在能力の測定指標と同じ社会指標を用いることが可能である。パフォーマンス要件で問われるのは、そのアウトカムが効果的・効率的にもたらされたかという点である。つまり、自治体経営が効果的かつ効率的であったからこそアウトカムの指標値が改善されたという関係性の特定が必要になるが、管見によれば自治体経営の状態を測るための広く合意された社会指標は存在しない。のみならず、「援助は役に立つか」という問いは大きすぎて反証可能な方法で科学的に答えられないのと同じように²²、「自治体経営の効果効率は高まっているか」という問いは大きすぎるため、観察されない交絡因子が多すぎて仮説を共同主観的にテストすることは非常に難しいのではないかと思われる²³。

2019, from <https://www.transparency.org/cpi2018#summary>。2018年版の腐敗認識指数の作成には、アフリカ開発銀行の調査（African Development Bank Country Policy and Institutional Assessment 2016）、世界銀行の調査（World Bank Country Policy and Institutional Assessment 2017）、世界経済フォーラムの調査（World Economic Forum Executive Opinion Survey 2018）等が用いられている。

²² 「援助は役に立つか」という問いは大きすぎて反証可能な方法で科学的に答えられないという主張については、Banerjee and Duflo（2011）を参照。米マサチューセッツ工科大学「貧困アクションラボ（The Abdul Latif Jameel Poverty Action Lab: J-PAL）」の創始者である Abhijit V. Banerjee と Esther Duflo は、「援助は役に立つか」という大きな問いを巡って論争を繰り広げるのではなく、具体的な問題とその個別の答えを実験で得られたデータに基づいて辛抱強く一步一步考えることこそが、世界最大の問題である貧困問題の撲滅に繋がると主張したのだった。

²³ 交絡因子とは、原因と結果の両方に影響を与える要因のことである。カール・ポパーの反証可能性（falsifiability）原理については、須藤（1988）第1章を参照。反証可能性原理は、科学理論といえども価値負荷的であり反証可能な憶測としての仮説的性格を失うことはないが、理論を科学者共同体が学界という公共の場で批判し合って経験的にテストできるという事実によって客観的な理論として承認されると考える。

これまで検討してきたように、目的要件、ガバナンス要件及びパフォーマンス要件を測定するための社会指標を開発して自治体経営の状態を測定評価するという研究アプローチは有益でありうるが、原理的な困難さを抱えている。よって本稿は、良い自治体経営が充たすべき要件の特定化方法に焦点を当てるのではなく、自治体経営の状態を測定評価するための地方自治体の意欲と能力に焦点を当てるという、もう一つの研究アプローチを採用する。すなわち、より良い自治体経営を実現するために政策評価の在り方をどのように構想し実現すればよいかという問いを本稿の主題とする。とりわけ、地方自治体の職員がボトムアップで変えることのできるソフトな制度に注目して、政策プロセスの各段階で自治体評価をどのように行えばよいかについて考えていこうと思う。

より良い自治体経営が行われているかどうかを地方自治体自らが評価する必要は必ずしもないし、むしろ内部評価は必然的に深刻な利益相反をもたらすから、ボトムアップ型の改革だけでより良い自治体経営に向けた取組みが進展すると期待することはできない。がん患者の状態をより正確に診断するためには専門的な知識とスキルを持つ臨床医の力が不可欠であるのと同じように、自治体経営の状態をより正確に診断するためには専門的な知識とスキルを持つ専門家の力が不可欠であろう。しかし他方で、いかなる治療やケアが望ましいかを患者抜きで決めることは不適切であるし、患者自身が自らの健康状態を自分である程度把握して管理できるようになることは患者の健康維持に役立つ。それと同じように、地方自治体の経営状態の望ましさを地方自治体が関与せずに決めることは不適切であろうし、地方自治体が自らの経営状態をある程度把握して管理できるようになることは、より良い自治体経営を実現するために不可欠だと考えられる。したがって本稿では、特に地方自治体がボトムアップ型で変えることのできるソフトな制度に焦点を当てて、自治体評価のあるべき姿について考察していこうと思う。

3. 日本の政策評価をめぐる先行研究と比べた本稿の特色

村松（1999）によれば、政策評価は日本においては学界でも実務でも 1990 年代に入ってから盛んになった比較的新しいテーマである。その大きな理由は、明治以後の日本政府にとって最大の関心事は西洋へのキャッチアップだったことにある。政府は評価制度を整備して効果を確かめることよりも西洋の優れた生産技術や法制度を輸入することを急ぎ、そして実際に西洋への急速なキャッチアップによって産業構造の高度化が実現し、第二次

世界大戦後も長きにわたって右肩上がり成長が続いたために、学界においても政策効果を科学的な方法で確かめるべきだという主張が大勢を占めることはなかった²⁴。

村松（1999）が言及しているのは、行政学界の動向である。政策評価を研究テーマとする学問分野は他にも数多あるが、その中で行政学と同じように政策評価の対象と手法の両方に強い関心を持つ代表的な学問分野をもう一つ挙げるとすれば、経済学であろう。では日本の経済学界において、政策評価というテーマはどのように扱われてきたのだろうか。

経済学にとって、市場を補完する経済主体としての政府が効率的で公正であるかという問いは、その学問の生成契機であったほどの根源的な問いである。したがって政策評価は、経済学の主要研究テーマの一つであり続けたといえるのかもしれない。しかしながら経済学界においても、日本の地方自治体や中央政府が講じる個別具体的な政策の効果効率に関する研究が盛んになったのは、1990年代以降のようである。猪木・大竹編（2001）では、最近まで日本の労働経済学では具体的な政策や法を取り上げてその効果を数量的に明らかにする研究は少なかったと指摘されており²⁵、玄田有史らによる1997年から1999年の労働経済学研究に関する学界展望では、政策評価が一つの柱として取り上げられていることが新しい労働研究の方向性を端的に示していると指摘されている²⁶。また金本良嗣は、2010年の日本経済学会の大会における日本の政策評価法施行後10年を振り返るパネル討論の中で、経済学者全体としてはまだ理論経済学をやる人が多く、経済学的知見を活かせる政策評価手法である費用便益分析を勉強している人が少ないことを指摘している²⁷。米国では連邦政府の職業能力開発プログラムや貧困対策等が早くから定量的な政策評価の対象とされて研究が進み、1981年からは連邦政府の主要な規制に関する規制影響評価が義務付けられて規制の経済分析が進展してきたことを鑑みると、やはり日本の経済学界においても行政学界と同様の状況があったのではないかと推察される。

²⁴ 村松（1999）第11章、特に pp. 249-256 を参照。1970年代から1980年代にも政策評価理論のブームはあったし（山谷 [2012] pp. 44-58）、科学の政策志向は既に明治期や第二次大戦直前から戦中戦後にかけての時期にも見られる（宮川 [1994] pp. 11-18）。しかし「政策評価論が注目され、かつ実務で試行錯誤が始まるのは1990年代半ばになってからであるし、それが学界一般の議論になるのはさらに遅れ」た（村松 [1999] p. 249 より引用）。

²⁵ 猪木・大竹編（2001）p. ii（猪木・大竹の共著による「はじめに」の部分参照）。

²⁶ 玄田ほか（2000）p. 25のうち、玄田の発言部分を参照。

²⁷ 大橋ほか（2011）pp. 165-166のうち、金本の発言部分を参照。なお金本は、日本の中央政府における政策評価制度の生成過程に深く関与した経済学者の一人である。

しかし 1990 年代に入ると、政策評価の研究も実務も急速に発展した。その研究の系譜は、3 つに大別することができる。第一の系譜は、政策評価手法に関する理論研究である²⁸。この研究群はバイアスの小さい政策評価を実践する上で参照しなければならない必須の知識を提供してくれるが、主たる研究関心は評価技術の高度化・精緻化であって、その技術が用いられる環境ではないため、どのような評価システムが日本におけるより良い自治体経営を促すかという制度的視点に欠ける。第二の系譜は、先進的な政策評価制度や政策評価事例に基づく政策評価の制度研究である²⁹。この研究群はより良い政策評価を実現するための制度について様々な示唆を与えてくれるが、推奨されている制度がより良い自治体経営に結びついていることを示す実証的証拠に乏しいため、推奨されている制度が日本の自治体経営の改善に本当に役立つのか疑問が残る。第三の系譜は、政策評価に関する定量的・定性的な実証研究である³⁰。この研究群は政策の効果効率等に関する科学的に信頼性の高い情報を与えてくれるが、主たる研究関心は政策の有効性・効率性というパフォーマンス要件に関わる部分に向けられており、目的要件やガバナンス要件に関わる分析評価は不足または欠落している場合が多い。

そこで本稿では、より良い自治体経営を促す政策評価システムについて、政策プロセスを通じた評価活動のマネジメントという視点及びデータによる実証を重視して考察することによって、上記に挙げた先行研究の限界を克服することを目指す。具体的には第一に、政策評価が充たすべき技術的な要件だけではなく、組織的・制度的な要件も重視する。とりわけ中央政府が定める法令のようなハードな制度ではなく、ストリートレベルの基礎自治体における業務プロセスや組織文化のようなソフトな制度に焦点を当てることで、先行研究の不足を補うだけではなく、自治体職員が現場発でボトムアップに改革を推進する際に役立つ知見を得ることを重視する。第二に、どのような制度を導入すべきかという制度論を展開するだけではなく、理想と実態とのギャップをデータに基づいて明らかにしたり、本稿で推奨する仕組みがより良い政策評価に結び付いているかを実証的に検証することに

²⁸ 宮川（1994）；常木（2000）；Ichimura and Taber（2001）；佐々木（2010）等。ここでは 1990 年代以降に公刊された日本の学界に基礎を置く研究者による研究業績を挙げたが、この研究群は基本的に、日本特有の政策評価手法ではなく汎用性のある普遍的な政策評価手法を研究対象としている。

²⁹ 上山（2002a）；上山・伊関（2003）；山谷（1997）；同（2006）；田中（2014）等。

³⁰ 金本ほか（2006）；松尾（2009）；Kuroda and Yamamoto（2012）；近藤（2014）；小野（2016）等。

も力を注ぐ。第三に、政策プロセス全体にわたる評価活動のマネジメントという視点に立って、自治体評価の文脈では軽視されがちであるニーズ評価や、目標を達成するための組織の在り方についても丁寧に検討する。

4. 本稿の構成

本稿の構成は以下の通りである。まず第1章「自治体ニーズ評価の理想と実態」では、ニーズ評価の理想的な手続きと現実の地方自治体によるニーズ評価とのギャップをデータに基づいて明らかにし、さらに、理想的なニーズ評価を行って政策問題を可視化することの意義を示す。政策問題は客観的に存在するものではなく、社会によって構築されるものである。したがって、データと評価者のバイアスを除去するよう系統的なニーズ評価を行い、その結果を可視化して社会に発信し、政策に反映することは極めて重要である。しかし地方自治体におけるニーズ評価は理想とは乖離しているという事実を、筆者が行った2種類の社会調査によって明らかにする。さらに、英国の貧困問題と日本の柔道事故の事例を用いて、ニーズ評価の社会的インパクトを確認する。

第2章「達成目標が充たすべき要件」では、文献研究に基づいて達成目標が充たすべき要件に関する仮説を立て、その仮説の妥当性を、潜在能力の拡大に貢献したと考えられる成功事例を用いて評価する。取り上げる事例は、岡山県総社市による「障がい者千人雇用事業」である。日本の労働研究には地方の労働市場に焦点を当てた実証研究が必ずしも多くないことや、日本の地方自治体には障害者就労支援策を主体的・積極的に行っている自治体が多くないこと等が影響して、管見によれば日本の地方自治体による障害者就労支援策が地域の障害者就業率に与える影響を定量的に明らかにした先行研究は存在しない。したがって本研究は、地方自治体の障害者就労支援策に関する貴重なインパクト研究になるのではないかと考えられる。

第3章「バイアスとエビデンス」では、地方自治体が目標に向かって正しい方向にやり抜くためにどうすればよいかを検討する。社会科学の推奨手続きに従ってできるだけバイアスのないように政策問題を定義し、良い達成目標を設定しても、政策選択を誤ってしまうことは十分にありうる。そのような事態を防ぐために、エビデンスに基づく医療の取り組みを参考にして、地方自治体がその時点で利用可能な最良のエビデンスを用いて政策を形成するためにどうすればよいかを論じる。さらに、行動科学の文献研究及び筆者が実施し

た基礎自治体での社会調査結果に基づいて、認知バイアスによる判断ミスを防いで最良の政策判断をするための組織的対策について述べる。最後に、自浄作用の不足によって地方自治体の政策判断が誤った方向に進んでしまうリスクを抑える方法について検討する。

第4章「目標を達成するまでやり抜く組織の構築方法」では、主にポジティブ心理学及び組織科学分野の文献研究に基づいて、失敗から系統的に学び、目標を達成するまでやり抜く組織の構築方法を検討する。挑戦には抵抗や失敗がつきものである。地方自治体が理想的なニーズ評価を行って良い達成目標を立て、その目標を達成するために適切な政策を立案したとしても、その政策が当初意図した通りに履行される保証はないし、当初予想した通りに機能する保証もない。それでも目標を達成するまでやり抜くために特に重要だと考えられるのは、失敗から組織的に学習することである。そこで、失敗から系統的に学ぶ能力と意欲を組織的に高める方法について検討する。

最後に終章で、本稿の結論を要約し、今後の研究課題及び謝辞を述べる。

第1章 自治体ニーズ評価の理想と実態

より良い自治体経営を実現するためには、的確な政策問題に対して正しい政策を適用することが必要である。政策問題を特定するための手続きを、評価学ではニーズ評価という。第1章では、日本の地方自治体によるニーズ評価の実態と課題を明らかにする。まずニーズ評価を含む政策評価の概念について簡単に説明した後、ニーズ評価の理想の手続きを整理する。次に自治体評価の理想と実態とのギャップを明らかにするための論証戦略を検討する。続いて、自治体ニーズ評価の理想と実態とのギャップに関する社会調査の結果を述べ、自治体ニーズ評価の課題を整理する。最後にニーズ評価の社会的意義を端的に示す事例について説明し、そこから得られる示唆をまとめる。

1. ニーズ評価とは何か

本節では、本稿で使用する基本用語の定義について、評価学の基本書の中でおそらく最も版を重ねて世界で読まれている文献の1つである Rossi et al.(2004)に依拠しながら述べる³¹。

1-1. 政策評価とは

本稿では「政策」を、社会問題の緩和ないし社会状況の改善を目的とする計画的な取組みのことで、Rossi et al.(2004)における社会的介入 (social intervention) ないし社会的プログラム (social program) のことと定義する。政府の計画、戦略、施策、事業、プログラム、プロジェクト等はいずれも政策でありうる。一方で「評価」とは、十分な根拠に支えられた (well-supported)、中立的で歪みのない (unbiased) 事実と価値基準に基づいて判断を下すことである³²。そして「政策評価」とは、社会調査法を系統的 (systematic) に用いて政策の実績や効果に関するデータを体系的に収集し、分析し、解釈する活動のことを指す³³。「社会調査法を系統的に用いる」とは、社会実験、標本調査、統計解析等の社会科学で発

³¹ Rossi et al. (2004) の初版は 1979 年公刊で、2018 年までに 8 回版が重ねられている。

³² 佐々木 (2010) 第 3 章を参照。

³³ Rossi et al. (2004) pp. 1-30 を参照。

達してきた調査分析手法を政策評価に体系的に応用することを意味する。

第3章で述べるように保健医療分野における系統的レビューの方法については業界内で標準とされる方法論が存在するが、政策評価は多様な政策について政策プロセスの各段階で適用される評価の総称なので、疫学研究のように方法論が確立していない。そもそも政策評価においては、科学的評価と実用的評価のどちらを重視すべきかを巡る論争に決着がつくことはないと言われている³⁴。大きな政府と小さな政府をめぐる政策論争の主流が時代と共に少しずつ変質しながら交互に揺れ動いてきたように、科学的評価と実用的評価のどちらを重視すべきかという価値観も、時代と共に揺らいできた。1960年代から1970年代は科学的評価が重要だとされ、米国では負の所得税や就学前教育の効果などに関する大規模な社会実験が行われた³⁵。だが、1970年代頃から小さな政府を支持する世論が高まると、科学的な厳密さよりも現場で使いやすい実用的な評価をすべきだという論調が大勢を占めるようになり、民間企業で1960年代頃から導入されてきた「目標による管理（Management by Objectives：MBO）」のような成果重視、顧客重視の業績管理制度を公共部門に取り入れる動きが盛んになっていった。日本の地方自治体及び中央政府が1990年代から2000年代にかけて導入した政策評価制度は、この流れを汲む。現在は第3章で述べるエビデンスに基づく医療の隆盛等の影響から、科学的評価を重視すべきだという考え方が再び主流になりつつある。

科学的評価と実用的評価のどちらを重視するにせよ、共通する考え方がある。それは、思いつきではなく系統的に良質な証拠を集めて多面的に分析することにより、評価のバイアスをできるだけ取り除くことが必要だという考え方である。例えば Rossi et al. (2004)

³⁴ 科学的評価と実用的評価をめぐる論争の歴史は Rossi et al. (2004) pp. 21-26；宮川 (1994) pp. 3-89；佐々木 (2010) pp. 85-96 を参照。古典的な論争は、Donald Campbell と Lee J. Cronback の見解の違いに表れている。Campbell は、バイアスのあるエビデンスに基づいて有効な政策を選び出すのではなく、重要な政策問題を選び出し、ランダム化比較試験や回帰分断モデルのような準実験的な方法で介入効果を確かめて何が最適か判断すべきだと主張した (Campbell [1969])。一方で Cronback は、評価は人々の判断に貢献する有益な情報を提供することによって正当化されるのであって、科学で標準とされる一つの型を当てはめることによって正当化されるのではないと指摘したのだった (Cronback [1980])。

³⁵ この時期は米連邦政府で「企画計画予算制度 (Planning-Programming-Budgeting System: PPBS)」が導入されて、科学的評価が重視された。同制度による科学的管理の過度な強調は1970年代初頭に頓挫したが、費用便益分析のような評価手法はそれ以降も定着し、現在も米連邦政府における規制影響評価等で用いられている。

はどちらかと言えば科学的評価を重視しているため、社会調査法の標準的手続きやランダム化比較試験のような厳格な研究デザインに基づくバイアス除去を重視する。他方で実用的評価を重視する Patton (2012) は、普遍的に最も頑健だと言える研究デザインは存在しないことから特定の研究デザインの適用は推奨しないが、体系的なフレームワークに基づいて系統的に評価することは推奨している。つまり、科学的評価と実用的評価のどちらを重視するとしても、理想的な政策評価とは、思いつきで場当たりの研究デザインを選んだり、都合のよいデータのみを分析評価に用いたりするのではなく、実践で許容される範囲内で社会科学の推奨手続きに則って実証的証拠を系統的に収集し利用することにより、バイアスをできるだけ小さくした上で価値判断を下す活動のことだと言えるだろう。逆に理想的ではない政策評価とは例えば、評価者やクライアントにとって都合の良い事実だけを選んで恣意的な方法で評価する活動を指す。

ただし本稿では政策評価という用語を、政策について事実に基づいて価値判断する活動という広い意味で捉える。したがって、理想的ではない政策評価も、ときに評価ではないとされるいわゆる業績測定 (performance measurement)³⁶に基づく目標管理型の実績評価も、政策に関する事実を価値判断している活動である限り政策評価とみなす。また、基本的に同じ評価技術を行政の内部事務評価にも適用できるし、行政の内部事務と政策との境界線は必ずしも明確ではないので、「政策評価」と「行政評価」は互換的な用語として用いる³⁷。「自治体評価」は特に地方自治体の政策に関する評価を指す用語として用いる。

1-2. 政策のロジック・モデルと5つの評価活動

Rossi et al. (2004) によれば、政策評価は①ニーズ評価、②セオリー評価、③プロセス評価、④インパクト評価、⑤コストパフォーマンス評価という5つの評価活動のうち、1

³⁶ 業績測定とは、あらかじめ定めた業績測定指標に基づいて政策等の業績をモニタリングする活動のことである。

³⁷ 用語の定義は、研究目的等によって多様である。例えば行政評価は、目標管理型の業績測定を指すと定義されることがある (上山 [1998])。政策、プログラム、プロジェクトという用語を区別し、業績評価、行政評価、政策評価はすべて異なる概念だと定義される場合もある (山谷 [2012])。政策評価とは、因果推論に基づいて政策の効果測定を行う研究領域だと定義されることもある (中室・津川 [2017])。

つ以上の評価活動から成る³⁸。①ニーズ評価とは、政策によって改善しようとする社会問題を特定して、政策の標的集団（target group）や政策に対する社会的ニーズを系統的に記述し、診断するための評価活動のことである³⁹。政策プロセスは一般に、政策形成（plan）、政策実施（do）、政策の見直し（see）の三段階に分けられる⁴⁰。そのうち政策形成プロセスは、政策問題を発見して定義することから始まる。そのステップをプログラム評価理論では「ニーズ評価」と呼ぶ。

②セオリー評価とは、政策が目指す社会状況の改善に至るまでの原因と結果の因果連鎖を描き、その妥当性を評価する活動のことである。セオリー評価では政策の論理的整合性を評価するために、政策体系の目的－手段関係に関する仮説群を図や表で表したロジック・モデルが作成されることがある⁴¹。ロジック・モデルは政策が最終的に目指す社会変化に至る道筋を描いたものなので、社会変革理論（theory of change）とも呼ばれる。遠位ないし長期アウトカムから近位ないし短期アウトカムまでを発生順に描いた部分を「インパクトパート」、政策効果が生じるまでの行政活動をインプット（投入）からアウトプット（産出）まで発生順に描いた部分を「プロセスパート」と呼ぶ。用途によって、一部分だけのロジック・モデルが構築されることもある。

③プロセス評価とは、政策が計画通りに進行し、意図した通りに標的集団にサービス等が届けられているかというプロセスパートの実績をモニタリングして評価する活動のことである⁴²。自治体評価では、プロセスパートで設定される業績指標を「活動指標」や「主要業績評価指標（Key Performance Indicators：KPI）」と呼ぶことが多い⁴³。

³⁸ Rossi et al. (2004) の枠組みは国際的にもよく使用されている。例えば、開発経済分野におけるエビデンスをつくり、つたえる活動を精力的に行っている「貧困アクションラボ」は、Rossi et al. (2004) に基づいて政策評価方法を説明している（同団体の「Introduction to Evaluations」を参照。Retrieved June 30, 2019, from <https://www.povertyactionlab.org/sites/default/files/resources/Introduction-to-Evaluations.pdf>）。

³⁹ Rossi et al. (2004) p. 54, p. 64 及び pp. 101-132。ある特定の政策に対するニーズ評価と、より幅広い地域ニーズを特定して新しい政策が必要かを判断するためのニーズ評価では規模も目的も異なるが、関連する方法はほぼ同じだとされる。

⁴⁰ 本章では国際的によく用いられてきた「Plan-Do-See」というフレームワークを用いているが、「See」の部分を「Check」と「Action」に分割しても議論の本質は変わらない。

⁴¹ Rossi et al. (2004) pp. 133-168；McLaughlin and Jordan (2004) を参照。ロジック・モデルを用いた電子行政の価値評価に関する理論的研究に、後藤（2007）がある。

⁴² Rossi et al. (2004) pp. 55-58 及び pp. 169-201 を参照。

⁴³ KPI や後述の KGI は、Norton and Kaplan (1992) のバランスト・スコアカードのフレ

④インパクト評価は政策の有効性評価のことで、ある政策によってその政策が目指す社会状況の改善が実現しているかを判断する活動のことである⁴⁴。インパクトパートの業績指標は、「アウトカム指標」と呼ばれる⁴⁵。自治体評価では、「成果指標」や「重要目標達成指標（Key Goal Indicators：KGI）」と呼ばれることもある。

アウトカム指標の業績測定を、「アウトカムモニタリング」という。日本の地方自治体で政策評価や行政評価と称されている活動は、アウトカムモニタリングであるか、または、アウトカム指標の実績値と目標値との間の乖離に基づいて実績を評価する「目標管理型の実績評価」であることが多い。業績情報はインパクト評価に役立つが、アウトカムモニタリングはインパクト評価とは異質な活動であり、アウトカムを測定するだけでは政策効果は分からないことに注意が必要である。インパクト評価とアウトカムモニタリングを分かちつ大きな違いの一つは、介入以外の要因がアウトカムに与える影響をできるだけ取り除いた上で介入効果の推定が行われているかという点にある。インパクト評価では介入によってアウトカムが変化したという因果関係を確認するために、研究デザインや解析手法の工夫が講じられる⁴⁶。もう一つの違いは、インパクト評価は評価活動なので、業績について何らかの価値判断がなされるという点である。

最後に⑤コストパフォーマンス評価とは、政策の費用対便益ないし費用対効果を評価して、インパクトをどれだけ効率的に高めるかを比較評価する活動のことである⁴⁷。第3章で述べるように、診療ガイドラインの作成時に用いられる GRADE システムでは、エビデンスの推奨の程度を決める際にコストを勘案することが標準手続きとされていない。その主な理由の一つは、コストは介入のタイミングや規模や地域のような外生的要因によって

ームワークで初めて用いられた用語とされる。なお KPI は、自治体評価ではインパクトパートの成果指標として用いられる場合もある。

⁴⁴ Rossi et al. (2004) pp. 58-59 及び pp. 203-232 を参照。

⁴⁵ 短期アウトカム指標としてよく用いられる指標に、利用者満足度がある。Gotoh (2009) では、事業者向け電子行政サービスの利用者満足度の決定要因を明らかにするため、後述する政策のロジック・モデルを用いてサービス品質の構成概念を整理した上で、共分散構造分析を行った。その結果、手続きの簡素化やサービス統合などのサービスプロセス改革を行わずに個別のサービス品質を改善しても、利用者満足度は高まらないことを明らかにした。

⁴⁶ 因果推論の方法については、Imbens and Rubin (2015) に詳しい。日本語の良質な入門書に、中室・津川 (2017) や伊藤 (2017) がある。

⁴⁷ Rossi et al. (2004) pp. 60-61 及び pp. 331-368 を参照。費用便益計算については Boardman et al. (2006) に詳しい。

大きく変動する不安定な要素だからである⁴⁸。だが地方自治体内の政策間あるいは国内の特定領域の政策間で比較する場合には、政策以外の要因がコストに与える影響は相対的に小さく、行政がコストをコントロールできる余地が相対的に大きいと考えられる。したがって自治体評価においては、限りある資源を有効活用するために、政策のコスト見合いの効果を比較評価する意義はより大きいだろう⁴⁹。

図1-1に、政策のロジック・モデルと上記①から⑤の評価活動との関係性を示した。ロジック・モデルでは「ニーズ」を明示しないことが多いし、自治体評価に関する文献でも政策現場でもニーズ評価を政策評価とみなさない例は多い⁵⁰。例えば、政策評価法⁵¹に基づく政策評価事務の全体を所掌しているのは総務省であるが、同省の「政策評価ポータルサイト」からアクセスできる「政策評価 Q&A（政策評価に関する問答集）」の冒頭には、「政策評価は、各府省が、自らその政策の効果を把握・分析し、評価を行うことにより、次の企画立案や実施に役立てるもので」あり、「このような政策評価の機能は、Plan（企画立案）、Do（実施）、Check（評価）、Action（企画立案への反映）という政策のマネジメントサイクルの働きとして考えることができ」とある⁵²。このように政策評価を「C（評価）」の段階で行う活動と捉える場合、Plan 段階のニーズ評価は政策評価とみなされないし、公共事業や規制の事前評価は、例外的な活動と位置づけられることになる。

⁴⁸ その他の重要な理由として、医療サービスの優先順位を効率性の優劣によって決めることは適切ではない場合が多いことや、機会費用の計算方法は専門家の間でも意見が分かれること等がある。GRADE システムについては、主に相原（2018）を参照。

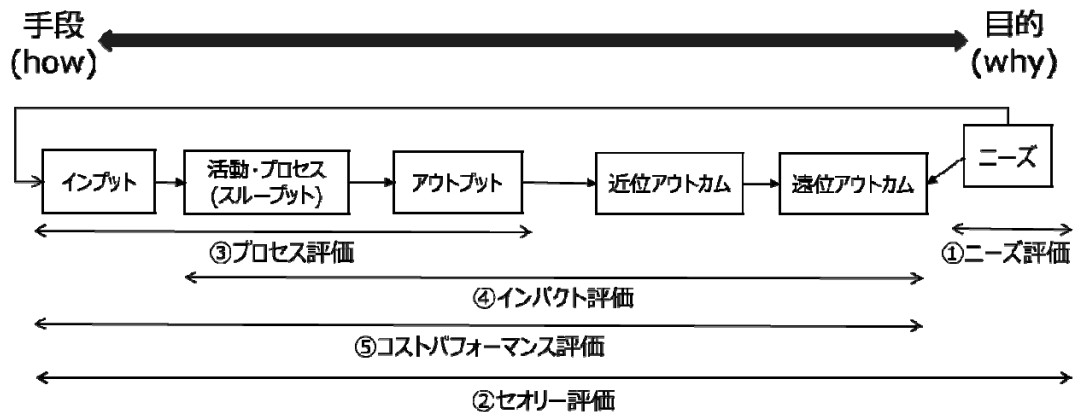
⁴⁹ 電子自治体の業績について、コスト見合いの効果ではないものの費用を勘案した業績指標を作成してパフォーマンス評価を行っている論文に、後藤（2009）がある。同論文では、日本の市区町村データを用いて、まず主成分分析によって「電子自治体の総合的な目標達成度」を表す総合的な業績指標を作成した上で、その業績指標に有意に影響を与えている組織的要因を重回帰分析によって評価し、投資対効果の事後評価に基づく政策改善の必要性や、投資対効果を高めるための組織的方策等を明らかにした。同論文は、日本の電子自治体投資に関する先駆的な定量研究であると評価されている（小野・倉橋 [2013]）。

⁵⁰ 例えば Wholey et al.（2004）に所収されている McLaughlin and Jordan（2004）は、ロジック・モデルにニーズを明記していない（同書、pp. 7-32 を参照）。

⁵¹ 正式名称は「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。最終改正は平成 27 年法律第 66 号）。

⁵² 総務省行政評価局「政策評価 Q&A（政策評価に関する問答集）」2017 年 10 月版、p. 1（Retrieved April 10, 2019, from http://www.soumu.go.jp/main_content/000359598.pdf）より引用。

(図 1 - 1) 政策のロジック・モデルと政策評価活動の関係



(注) 図中の①から⑤は本文中の説明順であり、評価手順を表しているものではない。

しかし政策のインパクトを評価するためにどれほど厳格に科学的な手続きを踏んでも、政策目的自体の誤りは分らない。正しい問題に誤った解を適用してしまった場合にはプロセス評価やインパクト評価によって政策の誤りを発見して正しい解を適用すればよいが、誤った問題を解いている場合には、どれほど政策を改善しても問題は解決されない。政策目的自体の見直しが必要である。政策目的の誤りや陳腐化に気付かないまま無駄な資源を投入したり、手段化してしまった目的を無益に追求しないようにするためには、ニーズ評価を政策評価の一部であると明確に位置づけ、政策目的を所与とみなさずにその妥当性を問い直す作業が不可欠である。

また、地方自治体によるニーズ評価は、近年重要性をより増している。第一次分権改革⁵³以降、地方自治体が各地域の政策問題を特定し、地域の実情を的確に反映した政策目的と政策手段を定めて運用しなければならない政策領域が拡大しているからである⁵⁴。したがって本稿では、政策形成段階で行う課題発見活動も政策評価に含まれると定義する。

⁵³ 第一次分権改革とは、1995 年成立の地方分権推進法に基づいて組織された地方分権推進委員会を中心に行われた、地方分権の推進のプロセスと成果とを指す。その最大の成果は、機関委任事務が全廃されていわゆる通達行政が廃止されたことである（大森 [2008] pp. 4-5 を参照）。

⁵⁴ 例えば障害者計画の策定は、2007 年 4 月から基礎自治体の義務になった（勝又 [2006a] を参照）。

1－3．政策プロセスと政策評価

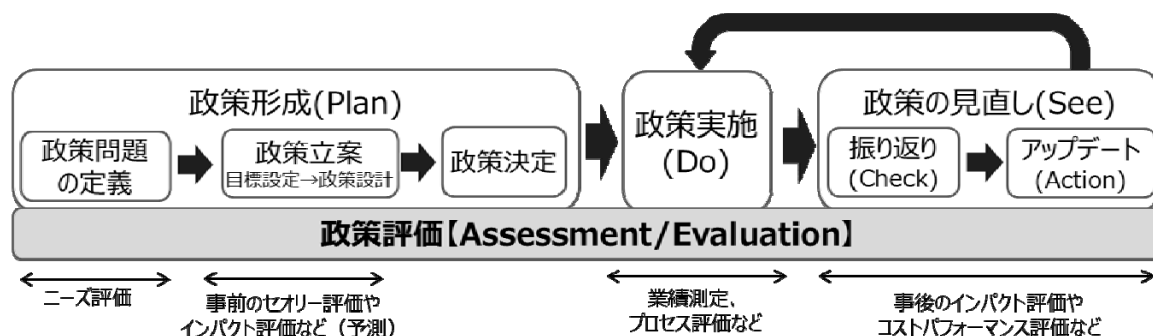
図1－2に日本の地方自治体が一般に想定していると考えられる政策プロセスと政策評価との関係を、図1－3に本稿の想定している政策プロセスと政策評価との関係を示した。

（図1－2）地方自治体で一般的に想定されている政策プロセスと政策評価の関係



（出典）総務省行政評価局「政策評価 Q&A（政策評価に関する問答集）」（平成29年10月版）p.1の図を参考にして筆者作成。

（図1－3）本稿の想定する政策プロセスと政策評価との関係



（注）「See」の部分で「Check」と「Action」に分割しても本質は変わらない。

図1－2と図1－3の主な違いは、2つある。1つ目は、政策評価を政策プロセスの一段階のみで行う活動と位置づけているか、政策プロセス全体を通して行うべき活動と位置づけているかという点である。本稿では、政策の見直し（See）段階で行われる事後評価だけでなく、政策形成（Plan）段階に行われるニーズ評価や事前評価もすべて政策評価と定義していることが、図1－3から読み取れると思う⁵⁵。

⁵⁵ 近年、日本の行政機関で急速に進展しつつある「証拠に基づく政策形成（Evidence-based policy making: EBPM）」の議論では、本稿と同じようにニーズ評価がエビデンスに基づく政策評価活動の一部だとみなされている。日本政府における EBPM 推進の動きは、

2つ目は、政策プロセスを、際限なく続く改善プロセスではなく、アップデートを何度か繰り返したら終了するプロセスとして描いている点である。地方自治体の予算は、基本的に単年度予算である。よって地方自治体は、毎年度新しい予算案を編成して議会に提案し、議会の議決を経た予算を執行して、当該年度が終わると決算書類を作成して議会に報告し、承認を得ている。毎年度の予算編成時において、常に政策体系が大きく見直されるわけではない。総合的な政策評価に基づいて政策体系を大きく見直すことができるのは、概ね3～5年の計画期間を持つ総合計画や分野別計画の策定時のみである。毎年度の予算編成時においては、基本的に計画に基づいて比較的簡単な振り返りとアップデートが行われるのみである。このような実態を反映して、図1－3では、図1－2とは異なる位置に矢印を配置した。政策を每期振り返ってアップデートする必要があることや、政策の終了を認識しやすくすることで、無益な政策の際限のない拡大を防ぐという意図もある⁵⁶。

2. 理想的なニーズ評価とは何か

日本の地方自治体及び中央政府は多くの場合、政策のPDCAサイクルの「C」の段階で行うアウトカムモニタリングないし目標管理型の実績評価を政策評価とみなしていると述べたが、現実の政策プロセスにおいてニーズ評価が行われていないわけではない。ニーズ評価は一般に政策評価と明確に位置づけられていないことが多いだけで、ニーズ評価自体は広く行われている。では、日本の地方自治体ではどのような方法でニーズ評価が実施されているのだろうか。理想的なニーズ評価と現実の自治体ニーズ評価とのギャップを把握するためには、理想的なニーズ評価とは何かを定義した上で実態を測定することが必要である。よってまず本節では、理想的なニーズ評価について、前節と同様に政策科学分野で版を重ねて用いられている定評ある基本書を用いて整理する。

「統計改革推進会議最終取りまとめ」（平成29年5月19日統計改革推進会議決定）を主たる契機としている（Retrieved April 10, 2019, from https://www.kantei.go.jp/jp/singi/toukeikaikaku/pdf/saishu_honbun.pdf）。同取りまとめは、統計等データを用いた事実や課題の把握、政策効果の予測・測定、評価に基づく政策改善のことを「EBPMサイクル」と称している。つまり本稿と同じように、課題把握、目標設定、政策効果の予測・測定・評価の全工程で、エピソードだけではなく、データに基づくエビデンスを活用するという意味でEBPMサイクルという用語が用いられている。

⁵⁶ 目標管理型の実績評価における目的と手段の倒錯リスクについては、例えば西尾（1990）第7章、特にpp. 285-286を参照。

前節で述べたように理想的な政策評価では、評価のバイアスをできるだけ小さくするために、時間やその他の資源制約上許容される範囲内で、社会科学の推奨手続きに則って実証的証拠を系統的に収集し利用することが求められるのだった。では、ニーズ評価における推奨手続きはどのようなものだろうか。Rossi et al. (2004) はニーズ評価を、①問題の定義、②問題の範囲の特定、③標的集団の特定、④標的集団の記述、⑤標的集団の行政サービスへのニーズの特定という5段階に分けて論じている。同書が参照している McKillip は 1987 年公刊の書物で、ニーズ評価のステップを、①利用者と利用の特定、②標的集団とサービス環境の記述、③ニーズの特定、④ニーズの重要性の評価、⑤結果の伝達、という5段階に分けている⁵⁷。他方で5版を重ねている政策分析の基本書で、どちらかと言えば実用的評価を重んじている Bardach and Patashnik (2016) は、①問題を定義し、②証拠を集めるという2つのステップをニーズ評価としている⁵⁸。この3冊を比較するだけでも分かるように、ニーズ評価のステップをどのように記述するかは文献によって異なる。しかし、それぞれのステップで理想的なニーズ評価が充たさなければならないとされる主な要件はあまり変わらない。それは一言でいえば、評価者や評価方法や評価対象等の偏りによって評価結果にバイアスが生じないようにしなければならないということである。社会問題は客観的な現象ではなく社会的に構築されるものなので、政策ニーズを状況固有の特徴から機械的に導出することはできない⁵⁹。何を解決すべき政策問題とみなすかについての最終的な決定は政治過程に委ねるしかないが、その決定を確かな証拠に基づいて公正に行うことができるように、できるだけ思い込みや恣意的な情報操作やデータの偏りによるバイアスを小さくすることが、ニーズ評価に求められることである。

理想的なニーズ評価が充たすべき要件は、大きく分けると以下の2つに大別できる⁶⁰。一つ目は、所与の制約の中で最大の成果を得るために、証拠を集める前によく考えて系統

⁵⁷ McKillip (1987) を参照。

⁵⁸ Bardach and Patashnik (2016) はプログラム評価ではなく政策分析の基本書であるためニーズ評価という用語を明示的に用いていないが、政策の代替案を分析してセオリー評価を行う段階を政策分析の3番目のステップとみなしているので、それに先立つ1番目と2番目のステップをニーズ評価とみなした。なお一般に政策分析 (policy analysis) とは、政策プロセスの Plan (政策形成) 段階におけるニーズ評価、セオリー評価、インパクトやコストパフォーマンスの事前評価を指す。

⁵⁹ Rossi et al. (2004) pp. 105-108 を参照。

⁶⁰ Rossi et al. (2004) pp. 101-132 並びに Bardach and Patashnik (2016) pp. 1-18 及び pp. 83-139 を参考に整理した。

的に探すことである。時間制約に基づく証拠や推論の質の低さは、思い込みや政治的動機によるバイアスと同じように事実や価値判断を歪ませうる。したがって、新しい情報を収集する前に、手持ちの知識と情報を駆使してよく考える必要があるとされる。具体的には、証拠集めをする前に、何が扱うべき政策問題なのか、問題の大きさはどのくらいか、何がその問題をもたらしているのか、誰が何を必要としているのか等について検討して暫定的な仮説を用意し、その仮説をどのように論証できるのかをよく考えなければならない。

証拠を集める前によく考えることが必要なのは、ニーズ評価の最初の段階だけではない。ニーズ評価のプロセス全体を通じて考えることと証拠を集めることを繰り返し、バイアスや無駄をできるだけ小さくするように努めなければならないとされる。ニーズ評価は学術研究ではないため、ユニークな発見をすることではなく、他の人々が過去に開発した優れたアイデアや過去の政策に関する業績情報、学術論文に掲載されているエビデンス等を発見し、解釈し、自分たちの課題分析に役立つように整理することこそが重要である。したがって独自の情報を得る前に既存のデータソースや社会指標を集めて分析すべきであるし、定量的な質問票調査を実施する前に、まずキーインフォーマント調査⁶¹などで探索的調査を行って、暫定的な問いとそれに対するありうべき答えを明確にすることが推奨される。なお本稿において「データ」とは統計や政治家の会話記録のような実証的事実のこと、「証拠」及び「エビデンス」は事実を証明する実証的根拠のこととする。証拠のうち因果関係に関する実証的根拠のみがエビデンスと定義される場合もあるが、本稿ではもっと広い定義を採用する。

データやエビデンスを系統的に収集するという要請は、多元的な情報源から証拠を得ることも意味する。バイアスを小さくするためには、人からの情報も文書からの情報も同じように重視すべきであるし、定量的な調査分析と定性的な調査分析の両方を駆使すべきだとされる。既存のデータソースには、統計、社会指標、学術研究、関連法令、他の自治体や他部局が実施した調査報告書や政策評価書等が含まれる。独自の調査を行う際には安易に全数調査や標本調査を行うのではなく、行政機関が保有する機関記録の活用や好事例（グッドプラクティス）調査なども行うべきだとされる。

理想的なニーズ評価に求められる 2 番目の要件は、政策の標的集団を代表する歪みの小

⁶¹ キーインフォーマントとは、その個人的または職業的地位によって、社会問題や標的集団の性質や範囲について見識をもっている人のことである（Rossi et al. [2004] p. 132, 訳 p. 123 を参照）。

さいデータを収集し、利用することである。独自の全数調査や標本調査に慎重でなければならない主な理由は、それが相対的に高度な専門性と高額な費用を必要とするからである。それでも問題の範囲や行政ニーズを特定するために質問票調査が必要なら、社会調査法や統計学の推奨手続きに従って、データのバイアスをできるだけ小さくすることが要請される。質的調査では、公聴会や地区住民懇談会では参加者が限られ政治的バイアスがかかりやすい点に注意する必要がある。

3. 自治体ニーズ評価の理想と実態のギャップの検証方法

理想的なニーズ評価が充たすべき要件は、①証拠を集める前によく考えて系統的に調べる、②政策の標的集団を代表するデータを利用する、という2つに整理できることが分かった。本節では、これら2つの点について日本の自治体ニーズ評価の実態を明らかにするための論証戦略について述べる。

3-1. 第一要件の検証方法

一番目の「証拠を集める前によく考えて系統的に調べる」という要件は、「証拠を集める前によく考えているか」と「系統的に調べているか」という2つの要件に分割できる。検証方法は複数ありうるが、前者は仮説検証型の質問票調査が行われているかを調べることによって、後者は独自の質問票調査を実施する前に、その機関が保有する過去の質問票調査結果を利活用したかを確かめることによって、少なくとも部分的な実態を把握することができる⁶²。独自の質問票調査を実施する前に既存のデータソースを用いた分析を行い、政策問題や行政ニーズに関する暫定的な仮説を立てた上で独自調査を行うべきだとされているのだから、独自の標本調査は原則として仮説検証型で行い、調査結果は調査担当部門以外が所掌する政策領域のニーズ評価でも有効活用すべきだと考えられる。では現実の自治体ニーズ評価において、仮説検証型の質問票調査が行われ、その結果が二次利用されているのだろうか。

⁶² 他の方法として例えば、好事例分析に用いられるデータのバイアスを確かめるという方法がある。Lundin et al. (2015) は、スウェーデンの自治体調査を実施して、地方自治体による成功事例の学習パターンとそのバイアスについて論じている。

この点を確認するため、本稿ではある基礎自治体（以下「X 自治体」という）が実施した住民向け質問票調査（以下「住民調査」という）の実態を調べて、理想的なニーズ評価とのギャップを確認する。調査単位を政策分野ではなく地方自治体に設定する一つの理由は、政策分野に焦点を当てて地方自治体による住民調査の実態を丁寧に分析評価した先行研究が存在するからである。具体的には大谷編（2002）が、2000 年末から 2001 年 4 月までの調査時点において最も新しい総合計画策定のための住民調査のうち、調査実施自治体に居住する住民のみを対象としており、かつ、調査内容が他自治体と同様のものを対象に分析評価を行っている。大谷編（2002）の主たる調査対象自治体は、大阪府内の全 44 市町村である。調査の結果、①多くの自治体が標本誤差をきちんと考慮せずに標本サイズを決めていたこと、②質問文や選択肢に問題があるために市民の意向や実態を反映したデータを得られない質問が過半数を超えること、③調査結果をホームページで公開していない自治体と調査結果データを電子媒体に保存していない自治体が共に 7 割を超えること、④殆どの自治体で調査の事実を全庁的に把握する仕組みがなく、1 回限りのただやっただけの調査になってしまっていること等が明らかにされている⁶³。上記①と②の結果は、理想的なニーズ評価が充たすべき第二の要件である「政策の標的集団を代表するデータを利用する」という要件が充たされていないことを示唆する。他方で③と④は、その結果が組織内外で利活用されていない可能性が高いこと、すなわち、「証拠を集める前によく考えて系統的に調べる」という理想的なニーズ評価が充たすべき第一の要件が充たされていないことを示唆する。

大谷編（2002）は自治体ニーズ評価の実態を自治体横断的かつ定量的に調べた貴重な先行研究だが、住民調査の実態を自治体単位で明らかにすることには成功していない。そこで本研究は、住民調査の実態を自治体単位で明らかにすることによって、大谷編（2002）では得られなかった知見を得ることを目指す。調査対象団体を複数団体ではなく一団体とした主な理由は、資源制約である。調査結果の外的妥当性を高めるためには一団体ではなく複数団体のデータを収集する方が望ましいが、自治体単位の信頼できる調査結果を得るためには調査対象自治体からの強力な研究協力が必要で、大きな交渉コストや調査コストがかかるため、断念した。したがって今回の調査は一自治体の事例分析に留まり、他の自治体にも通用するかは不確かであるという限界を有する。しかし一方で、政策単位ではなく自治体単位で実態を明らかにするという本研究のアプローチは、意欲のある地方自治体

⁶³ 大谷編（2006）。後藤（2017a）p. 46 より引用。

による自発的改革に繋がりやすいという利点がある。政策分野によっては同じような住民調査がほぼ同時期に複数自治体で実施されているので、複数自治体が協力して良質な住民調査を実施して政策問題の発見に活かすことは十分に実現可能で、かつ有益だと考えられるものの、自治体間の調整をボトムアップで実現するのは容易ではない。例えば次項で述べるように、男女共同参画計画を立案するためのニーズ評価では内閣府の世論調査を参考にした住民調査や事業所調査が多くの地方自治体で実施されている。そのため、調査項目を自治体間で調整して一斉に仮説検証型調査を実施し、その結果を自治体間で共有して二次利用すれば、各地域の政策問題をより容易に明らかにすることができると期待される。しかし、それを自治体間連携によってボトムアップ型で実現するのは、容易ではないだろう。一方で、自治体単位で理想的なニーズ調査を実施するための体制を整え、独自調査結果の庁内活用を促進することは、自治体間調整とは異なり業務命令で実行しうるので、意欲のある地方自治体による自発的な改革が相対的に容易なはずである。自治体単位でニーズ評価の実態と課題を明らかにするという本研究の研究アプローチは、自治体単位でのボトムアップ型の改革を促す効果を持つことが期待される。

調査対象団体には関東にある中核市クラスの X 自治体を選んだ。その主な理由は、統計担当部局職員がデータやエビデンスに基づく政策運営に強い関心を持っており、本研究への全面的な協力を得られたからである。大谷編（2002）が自治体単位での実態解明を断念した主な理由は、調査対象の基礎自治体には社会調査結果を一元的に把握する仕組みはなく、過去の住民調査の情報は各部局に分散していたため、調査実態を事後的に把握して集約することが難しかったからであった。X 自治体にも各課が実施している社会調査情報を一元的に把握する仕組みはなかった。そのため、X 自治体の各部局が過去に実施した住民調査の実態を把握するための質問票調査（以下「メタ調査」という）を実施することにした。調査単位は課の単位で出先機関は対象外とし、各課が調査時点までの過去 5 年間（2012 年度から 2016 年度）に実施した住民調査のうち標本サイズが最も大きい調査について、調査目的や調査方法等の基本情報、調査の主要な問いとありうべき答えに関する仮説、調査結果の公表方法、調査結果の保管実態及び調査結果の利活用状況等について尋ねた⁶⁴。

⁶⁴ 各課が過去 5 年間に実施した標本サイズの最大の調査について調べた理由は、調査対象部局の政策運営にとって最も重要な住民調査の標本サイズが最大であろうという想定に基づく。各課に複数の調査について詳細な回答を求めるのでは回答負担が大きすぎるので、課ごとに一つずつ住民調査の実態を尋ねた。

メタ調査の調査時点は 2017 年 3 月から 7 月で、調査票は X 自治体の統計担当部局職員に電子メールで配布・回収してもらい、未回収部局には電話や電子メールで催促してもらった。学術目的による調査データの利用や結果の公表に関しては明示的な同意を得たが、個人及び団体の特定を避けるため、地域が限定されうるものは一般的な表現で示す。

3-2. 第二要件の検証方法

二番目の要件である母集団を代表するデータを利用しているかについては、どのように調べればよいだろうか。データのバイアスとは、調査または推論の過程で諸事情により生ずる特定方向への歪みのことである。標本調査で注意しなければならない主なバイアスには、標本誤差、選択バイアス、測定バイアスがある。標本誤差とは全数調査ではなく無作為抽出による調査を行うことによって生じる、母集団の真の値と標本値のズレのことである。通常、関心のある属性ごとの標本サイズを一定規模以上にして、例えば 95% の信頼度で標本誤差を $\pm 5\%$ 以内に留めるように標本サイズが決定される。選択バイアスとは、調査対象に選ばれたものと選ばれなかったものとの間にみられる系統的な誤差のことである。標本調査で計画標本を母集団から無作為抽出で選んだとしても、調査や測定全体への無回答や不参加によって特定の属性をもったケースだけが選ばれてしまうことはよくある。そのときに選択メカニズムを無視して単純な推定を行うと、選択バイアスが生じる⁶⁵。無回答による系統誤差に対しては、無回答者への追跡調査、無回答者の属性に基づく欠損値の推計、事後的なウェイト付け等の対処方法があるが、不用意な推計はむしろバイアスを大きくしてしまう虞があるため、調査の企画・設計段階でできるだけ無回答バイアスが生じないような工夫を講じておく必要がある⁶⁶。最後に測定バイアスとは、測定法の違いによって同じ対象者の測定結果に違いが生じてしまうことを指す。例えば一般に、留置法よりも郵送法による質問票調査の方が、行動に関する調査よりも意識に関する調査の方が回答者の解釈のバラつきが大きくなるため、測定バイアスが大きくなる。

本稿では、中核市の男女共同参画計画において住民意識調査に基づく数値目標が設定されているかを調べ、住民意識調査に基づく数値目標が設定されている場合に、その数値目標を設定する際にデータのバイアスが考慮されていたかを調べる。さらに、男女共同参画

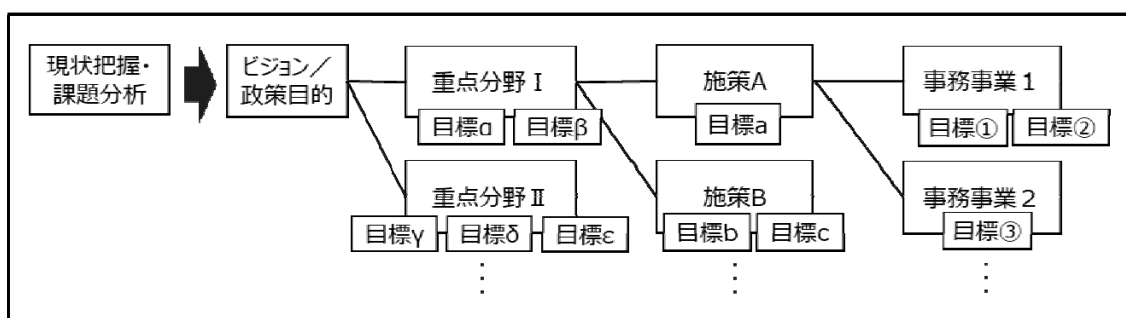
⁶⁵ 星野（2009） pp. 143-146 を参照。

⁶⁶ 本多・本川（2005） pp. 64-66 を参照。

計画で参照されている住民意識調査の有効回収標本サイズ、有効回収率、有効回収標本の男女比と母集団の男女比との差、欠落バイアスへの事後的対処の有無について調べることで、データのバイアスについて評価する。調査方法はインターネット調査で、2019年4月に各自治体の公式ホームページから最新の男女共同参画計画とそこで参照されている住民意識調査のデータを得た。

一般に地方自治体の各種計画は図1-4のような論理構造をしている。すなわち、現状把握・課題分析が行われた後に、ビジョン・ステートメントのような政策目的が述べられ、いくつかの重点分野ないし基本方針と基本目標が定められる。続いて、重点分野ごとに施策や事務事業が配置され、施策ごとあるいは事務事業ごとに業績目標が設定される。それらの目標は定性的なものも定量的なものもありうるが、一般に多くの自治体計画には定量的なアウトカム目標が設定され、その指標値の推移が定期的に測定評価される。そのような数値目標をバイアスの大きいデータに基づいて設定してしまうと業績の達成度合いを正しく把握することはできないから、基準値や実績は政策の標的集団を代表するデータに基づいて測定しなければならない。したがって、数値目標の設定時におけるデータのバイアスの考慮有無や、参照されている調査における有効回収標本サイズ等を調べることによって、政策の標的集団を代表するデータが得られているか、政策の標的集団を代表するデータが得られていない場合にはデータのバイアスを補正する何らかの工夫を講じた上で計画の目標値に用いられているかを明らかにし、もってニーズ評価の理想的な手続きとのギャップを確かめるわけである。

(図1-4) 地方自治体による各種計画の一般的な構造



調査対象として中核市を選んだ理由は、政令指定都市と並ぶ日本の大都市制度の一つで、一般的な市町村と比べれば職員や予算の規模が大きく、標本誤差等を小さくする工夫を講

じることが相対的に容易である一方で、政令指定都市よりは人口規模が小さいために住民の地元への帰属意識がより強く、他の条件が同一ならより高い回収率が見込まれると予想されるためである。中核市調査で仮に深刻なデータのバイアスがみられた場合、他の類型の地方自治体による標本調査ではより一層深刻なデータのバイアスがみられる可能性が高いのではないかと考えた。

調査対象の調査を男女共同参画計画策定のために用いられる住民意識調査にした第一の理由は、男女共同参画分野のニーズ評価段階で行われる住民意識調査は性別や年齢等によって大きな差があるような質問が多いため、選択バイアスを考慮せずに結果を単純に集計してしまうと、事実認識に大きな歪みが生じてしまう恐れが大きいからである。男女共同参画計画策定のためのニーズ評価で多くの地方自治体が住民意識調査を実施している理由の一つは、男女共同参画政策の推進を所掌する内閣府の指針等において、意識調査が重視されているからだと推察される。地方創生のための総合戦略その他多くの分野別計画の場合と同じように、男女共同参画分野も、中央政府が法令、行政文書、補助金等によって地方自治体の政策を誘導している典型的な政策分野の一つである。内閣府は、中央政府の男女共同参画計画を 2000 年 12 月に策定して翌年 2001 年 8 月に「市町村男女共同参画計画の策定の手引き」という行政文書を公表した⁶⁷。同手引きは基礎自治体が中央政府に準じた男女共同参画計画を策定できるようにするためのマニュアルである。同手引き 16 ページには、どのようなニーズ調査をすべきかについて細かく書かれている。具体的には、「男女共同参画計画策定に当たって、地域の実態を把握し、その上で政策課題を明らかに」するため、「基礎資料を収集・整備する」ことが重要であり、「新たに調査する場合は、調査項目に地域の特色を踏まえたものを加えつつも、国又は都道府県で実施している調査（＊6、原文ママ）と同じ項目を入れることによって、全国又は都道府県平均と地域との比較が可能となります。また、男女別、年齢別等により回答に大きな差が生ずるテーマが多い意識調査などは、集計の際クロス分析等ができるよう設計を工夫する必要があります。」と詳しく説明されている。引用文中の「＊6」で示したページ脚注には、内閣府が定期的に行っている「男女平等参画社会に関する世論調査」（以下「内閣府世論調査」という）の調査票から、「男女の地位の平等感」、「『男は仕事、女は家庭』という考え方について」、

⁶⁷ 内閣府男女共同参画局「市町村男女共同参画計画の策定の手引き」（2001 年 8 月），Retrieved April 11, 2019, from http://www.gender.go.jp/about_danjo/basic_plans/s-keikaku/keikaku-tebiki.html#tebiki。

「女性が職業を持つことについて」という3つのテーマに関する意識調査の質問文と選択肢が転記されている。この手引き 16 ページを素直に読むと、男女共同参画計画のニーズ評価のために独自調査を行う場合、内閣府が定期的実施している「内閣府世論調査」の調査票を参考にして、それに地域独自の調査項目を加えた意識調査を行うことが望ましい、と読めるのではないかと思う。

基礎自治体には男女共同参画計画の策定義務はないし、同手引きに従う義務もない⁶⁸。しかし市町村の男女共同参画計画の策定率は中央政府の男女共同参画計画の成果目標に設定されており、2017 年時点でほぼ全ての市区と過半数の町村が男女共同参画計画を策定している⁶⁹。基礎自治体による男女共同参画策定を後押ししている経済的誘因は、補助金である。「地域女性活躍推進交付金」を例にして、中央政府の補助金による地方の政策誘導についてみてみよう。男女共同参画分野では近年、「女性の職業生活における活躍の推進に関する法律（平成 27 年法律第 64 号。以下「女性活躍推進法」という）が施行され、中央政府の基本方針⁷⁰等を勘案して各地域の女性活躍推進計画を策定することが地方自治体の努力義務とされた。内閣府「地域女性活躍推進交付金」（平成 29 年度予算、平成 29 年度補正予算）の交付対象事業は各地方自治体の女性活躍推進計画の策定等で、男女共同参画計画を既に策定しているか策定することが確実なことがその補助事業実施主体の一つの要件とされていた。つまり、女性活躍推進法という新しい法律に基づく計画の策定を促すために中央政府の交付金が用意されており、その交付金による補助を受けるためには、男女共同参画基本法に基づく男女共同参画計画を策定しなければならないと規定されていた。以上のような制度的・経済的な誘因装置が強力な要因となって、多くの地方自治体は内閣府世論調査と同じような意識調査を行った上で男女共同参画計画を策定しているのだと考えられる。したがって中核市が男女共同参画計画のニーズ評価のために実施する標本調査で深刻な選択バイアスがみられるのなら、他の地方自治体でも同様の問題が生じている可

⁶⁸ 男女共同参画基本法第 9 条は地方自治体に対して、男女共同参画社会の形成促進に関して国の施策に準じた施策及びその他の地域特性に応じた施策を策定・実施する責務を課している。都道府県は国の男女共同参画基本計画を勘案して「都道府県男女共同参画計画」を定める義務が課されているが（同法第 14 条第 1 項）、市町村は国の男女共同参画計画及び都道府県の男女共同参画計画を勘案して市町村の男女共同参画計画を定める努力義務が課されているだけである（同法第 14 条第 3 項）。

⁶⁹ 正確には、市区の 96.4%、町村の 56.6%である（内閣府編 [2018a] p. 275）。なお市区とは市及び東京都特別区のことである。

⁷⁰ 正確には、「女性の職業生活における活躍の推進に関する基本方針」。

能性が高いと予想される。

調査対象の調査を男女共同参画計画のニーズ調査にしたもう一つの理由は、地方自治体による男女共同参画政策の評価実態及び課題の解明を目的とした先行研究が 2015 年に公開されたので、その先行研究との接合が期待されるからである。自治体評価の実態解明を目的とする学術研究は数多く存在するが、特定の政策分野における自治体評価の実態を定量的・横断的に調査した研究は必ずしも多くない⁷¹。まだ実態があまり明らかにされていない政策分野に関する自治体評価の実態解明に取り組むことも重要であるが、先行研究のある分野でその不足を補うことによって、一つの政策分野における自治体評価の全体像を把握できるようにすることにも大きな学術的意義があると考えられる。

本稿が着目した男女共同参画政策における自治体評価実態及び課題の解明を目的とした先行研究は、内藤・山谷編（2015）である。同書は、複数の研究者と実務家が文部科学省科学研究費補助金基盤研究（C）「男女共同参画政策の推進に向けた評価に関する調査研究」を共同で遂行し、その研究成果をまとめた書籍である。同研究プロジェクトでは、2011 年時点で男女共同参画社会基本法に基づく男女共同参画計画と公設の男女共同参画拠点施設の両方を持つ全 325 の地方自治体の男女共同参画所管課及び全 331 の男女共同参画拠点施設に対して、2011 年に質問票調査を実施している。うち自治体調査の有効回答率は 80.3%で、① 8 割以上の地方自治体が何らかの評価を実施していること、② 評価実施団体の 8 割以上は基本計画の目標管理・進捗管理のために評価を実施していること、③ 評価実施団体の 9 割以上の評価方式は担当者、担当部局による自己評価であること等が明らかにされている⁷²。自治体評価は総合計画や分野別計画の進捗管理のための実績評価が中心であることや、内部評価中心であることは先行研究でも指摘されてきたが、その実態が男女共同参画政策でもみられることを定量的に明らかにした点は重要な貢献であろう。一方で本稿の文脈で注目されるのは、内藤・山谷編（2015）は男女共同参画に関する自治体評価を主題としているが、ニーズ評価を主な調査分析対象に含めていなかったと考えられる点

⁷¹ 例えば日本評価学会の学会誌『日本評価研究』には自治体評価に関する論文が数多く掲載されているが、各号の目次に示された論文表題から推察する限り、2001 年 1 月の創刊号から 2019 年 1 月号（第 19 巻第 1 号）までの時点で特定の政策分野における自治体評価の実態を定量的・横断的に調査した研究論文は掲載されていないようであった。なお、自治体評価全般または特定自治体の政策評価の実態に関する定量研究は複数掲載されていた。また、特定の政策分野に特化した特集も複数回組まれていた。

⁷² 高橋（2015）を参照。

である⁷³。前掲図 1－2 で示したように日本の地方自治体では政策の PDCA サイクルの「C」の段階で行う評価活動のみを自治体評価とみなすことが多いので、同研究の視座は例外的なものではない。だが前述したように、良い政策を策定するためには社会問題を的確に診断して真に解くべき政策課題にアプローチしなければならない。よって先行研究に不足している男女共同参画計画のニーズ評価に焦点を当てて、自治体評価の実態と課題を明らかにする⁷⁴。

4. 自治体ニーズ評価の実態と課題

4－1. 第一要件の検証結果と考察

理想的なニーズ評価の第一要件は、「証拠を集める前によく考えて系統的に調べる」というものであった。本研究では X 自治体が 2012 年度から 2016 年度の 5 年間に実施した住民調査に関するメタ調査を実施し⁷⁵、現実の自治体ニーズ評価において仮説検証型調査が行われているか、調査結果が二次利用されているかを明らかにすることによって、理想と実態のギャップを確かめようとしているのであった。

まずメタ調査結果の概要について述べる。メタ調査は X 自治体の 88 部局に配布し、87 部局から回答を得ることができた。有効回答率は 98.9% と非常に高かったため、欠落バイアスによる問題は調査全体ではほぼ生じないと考えてよいだろう。有効回答の 87 部局のうち、過去 5 年間に住民調査を実施したのは 30 部局（有効回答の 34.5%）で、過去 5 年

⁷³ 内藤・山谷編（2015）所収の高橋（2015）に彼らの共同研究で実施された質問票調査結果が示されている。同論文から推察する限り、質問票調査には政策や予算や組織の有効性、効率性、答責性に関する調査項目はあったが、政策の必要性や目標の妥当性のようなニーズ評価に関係する調査項目はなかったようである。なおその理由は、自治体現場においてニーズ評価が自治体評価の主活動の一つだとみなされていないことにも起因するのではないかと推察される。

⁷⁴ なおニーズ評価においては、政策問題を引き起こしている真因を明らかにすることが求められる。そこで後藤（2019b）では、管理職到達度の規定要因を明らかにするため、地方都市居住者を対象にした調査データを用いてロジスティック回帰分析を行った。その結果、管理職到達度の男女差は、男性の年齢効果がプラスであるのに対して、女性の年齢効果がマイナスであることから生じていること等が分かった。

⁷⁵ 前述したように、住民調査とは住民向け質問票調査のことで、住民調査に関するメタ調査とは筆者が実施した X 自治体による住民調査に関する質問票調査のことである。

間に実施した住民調査の回数は、1回が14部局、2回が5部局、3回が3部局、4回以上が6部局、無回答が2部局であった。合計すると過去5年間で65件、年平均では13件の住民調査が実施されていた。大谷編（2002）によれば、1996年度から2000年度の5年間に大阪府内の3市が実施した市民意識調査件数は、一団体年平均3件強であった⁷⁶。それに比べるとかなり多くの住民調査が行われていたことになる。この違いが生じた主な理由は、地方分権化によるニーズ評価の必要性の増大、情報通信技術（information and communication technologies、以下「ICT」という）の進歩と普及による質問票調査の実施コストの低下、彼我の調査対象の違いに由来するのではないかと推察される。

X自治体で過去5年間に住民調査を実施した30部局に対しては、各部局が過去5年間に実施した住民調査のうち最も標本サイズの大きい調査について、調査目的等の詳細の回答を求めた。まず調査目的は、「計画策定のため」が13件、「実態やニーズ把握のため」が14件、「施策等評価のため」が3件であった。30件中27件、つまり9割の調査はニーズ調査またはニーズ評価のために実施されていたと考えてよいだろう。住民ニーズの把握または評価を目的としていた27件の住民調査について、調査方法を全数調査と標本調査に分けると、全数調査が8件、標本調査が19件で、標本調査の標本抽出方法は、無作為抽出法が10件、有意抽出法が8件、無作為抽出法と有意抽出法の組合せが1件であった。無作為抽出法の場合には住民基本台帳から無作為抽出するケースが多く、有意抽出法にはサービス利用者やインターネットモニターに任意で回答してもらう形式と、ランダムではない何らかの基準で母集団から標本を抽出する形式とがあった。調査票の配布・回収方法（複数回答）は、郵送法が配布では14件、回収では16件と最も多く、託送法が配布では7件、回収では6件で郵送法に続いた⁷⁷。計画標本サイズと回収標本サイズは100件未満から1万件以上までとバラつきが大きかった。

次に、ニーズ評価の理想的な手続き要件を充たす調査が行われていたかをみていく。ニーズ評価の理想的な手続きでは、独自の定量調査を実施する場合には既存のデータソースの分析評価や質的調査を実施して、政策問題の分布や標的集団の行政ニーズに関する暫定

⁷⁶ 大谷編（2002）p. 93を参照。

⁷⁷ 郵送法とは、調査票を調査対象者に郵送で配布または調査対象者から郵送で回収する方法のこと。託送法とは、調査票を既存の組織や集団を利用して配付または回収する方法のこと。その他の調査票配布回収法としては、留置法（調査員が調査対象者を訪問して調査票を配付または回収する方法）、窓口（窓口調査票を配置し回収箱で回収する方法）、オンライン（ウェブサイト調査票を掲載しオンラインで回収する方法）があった。

的な問いとそれに対するありうべき答えを明確にした上で仮説検証型の調査を実施することが推奨されているのであった。そこで、調査のリサーチ・クエスチョンが明確だったかどうかを尋ねたところ、「明確である」が16件、「やや明確である」が2件、「やや明確ではない」が2件、「明確ではない」が6件、「分からない」が1件だった⁷⁸。次に、主なりサーチ・クエスチョンに対して施策に直結する明確な仮説を立てたかを尋ねたところ、「明確で、施策に直結する仮説を立てた」が7件、「明確だが、施策に直結しない仮説を立てた」が2件、「仮説を立てていない」が16件、「分からない」が2件だった⁷⁹。つまり、大多数の住民調査は仮説検証型ではなかった。

佐藤（2015）は、どのような調査結果が得られるのかという見通しを立てずに「出たとこ勝負」で行われる質問票調査は、何がまだよく分かっていない事柄であるかについての検討が十分に行われていないために有益な知見を得ることが難しい場合が多いだけでなく、悪質な事後解釈を引き起こすと警鐘を鳴らす。調査で得られた結果のうち、何らかの意味で都合のよいものだけをつまみ食いの取り上げてもっともらしい解釈を事後的に付け加えることによって、事実が歪められてしまうリスクが大きいからである⁸⁰。だからこそニーズ評価の理想的な手続きでは、独自の量的調査を実施する前によく考えて系統的に調べることが推奨されているのである。調査資源を投下する価値のある調査結果を得て、それをバイアスが生じないように利活用するためには、筋のよいリサーチ・クエスチョンと筋のよい仮説を立ててから質問票調査を実施することが肝要であり、後付けバイアス⁸¹を防ぐために、事前に立てた仮説を事後的に確認できるよう記録として残して引き継いでいくべきであろう。

⁷⁸ 「どのような属性の女性の離職率が高いのか、それはなぜか。」を明確な問い、「働く女性の実態やニーズを知りたい。」を明確ではない問いと例示した上で尋ねた。

⁷⁹ 「子育て支援に熱心ではない職場で働く女性の離職率が高い。」を明確で施策に直結する仮説、「高収入の男性と結婚した女性の離職率が高い。」を明確だが、施策に直結しにくい仮説と例示した上で尋ねた。なお、回答の歪みが出ないように選択肢は仮説の明確さと施策との関連性の二点について対照的なものを用意したが、回答が0件だったものの説明は省いた。

⁸⁰ 佐藤（2015）pp. 113-151を参照。政策のために都合の良いデータやエビデンスを用いるという現象の発生原因と対策については、第3章第3節及び第4節で述べる。

⁸¹ 後付けバイアス（hindsight bias）とは、ある事象の事後的な結果をみて、その結果を事前に予想していたかのように思い込むバイアスのことである。Fischhoff（1975）；同（1977）；Roese and Vohs（2012）を参照。

次に二次利用、データの保管実態及び結果の公表状況についてみていく。X自治体で過去5年間にニーズ把握またはニーズ評価を目的として行われた27件の住民調査のうち、調査によって得られた個票データまたは集計データを直接的な調査目的以外に二次利用した実績または予定がある調査は3件しかなく、残り23件の調査では二次利用の予定がなく、4件は二次利用の予定が「分からない」という回答だった。大谷編（2002）は、「ほとんどの市民意識調査が、1回限りのただ〈やっただけの調査〉としておこなわれ、過去との比較や他調査との比較等の分析はまったくおこなわれてこなかった」⁸²と指摘していた。大谷らの調査から15年以上経過した時点でも、地方自治体による住民調査の利活用はあまり進んでいないことが示唆される。

ただし今回の調査では、条件さえ整えば調査データの有効活用を推進できることも分かった。大谷編（2002）によれば、大阪府内の基礎自治体による総合計画策定のための市民意識調査について、調査結果をデータ保存していた自治体は8団体しかなく、32団体ではデータが保存されていなかった⁸³。しかし今回のX自治体調査では、27件中23件の調査で集計データが、27件中21件の調査で個票データが、二次利用しやすいデータ形式で保管されていた⁸⁴。したがって、集計データや匿名加工データの二次利用を進めるためのルールと体制さえ整えば、多くの調査結果を再利用容易なデータ形式で公開することが物理的には可能である。データを再利用容易な形式で公開すれば、第三者が調査結果からより多くの知見を引き出すことが可能になるし、ニーズ評価の妥当性検証が容易になるため、より良質なニーズ評価を促す良いインセンティブにもなると期待される。個人情報の取り扱いには十分に注意しなければならないが、調査結果の公開やデータ開放を積極的に進めるべきであろう。

では、地方自治体は住民調査の結果を積極的に公開しているのだろうか。X自治体でニーズ調査またはニーズ評価のために実施された27件の住民調査について調査結果の公表方法を尋ねたところ、設問ごとの集計結果のような「結果の詳細な内容を公表」と回答したケースが8件、「結果の概要のみを公表」と回答したケースが6件、「結果を公表していない」と回答したケースが10件、「分からない」が1件、無回答が2件だった。ごく限ら

⁸² 大谷編（2002）p. 223 より引用。

⁸³ 大谷編（2002）p. 60 を参照。

⁸⁴ 調査データを二次利用する場合には個票データがないと目的に沿った分析評価ができないことが多いため、27件中6件の調査で個票のデジタルデータを保管していないか行方が分からない状態だったという実態も問題ではある。

れた対象者に対する調査の場合、調査結果から個人を特定できないように細心の注意を払ったとしても、母集団の傾向が明らかになること自体が憚られるケースがありうる。したがって全ての調査について結果の詳細を公表すべきだというわけではない。しかし、調査票その他の関連資料から筆者が判断する限りでは、「調査結果を公表していない」という10件の調査のうち、調査の客体側に起因する特別な事情があると思われる調査はごく僅かしかなかった。また「結果の概要のみを公表した」という6件の調査の多くは、より詳細な調査結果を公表しても調査の客体側に起因する問題は生じないと思われる内容だった。調査結果の詳細な内容を公表しているケースが27件中8件しかなかったという事実は、ニーズ評価プロセスの透明性の低さを示す深刻な問題ではないかと思われる。

4-2. 第二要件の検証結果と考察

理想的なニーズ評価の2番目の要件は、「政策の標的集団を代表するデータを利用する」というものであった。本研究では中核市による最新の男女共同参画計画とそこで参照されている住民意識調査に関するインターネット調査を実施し、男女共同参画計画の数値指標における住民意識調査データの使用有無、実績値や目標値におけるバイアスの補正有無、男女共同参画計画のための住民意識調査の有効回収標本サイズ、有効回収率、有効回収標本と母集団の性別構成比の差、及び、選択バイアスを考慮した分析が行われているかを明らかにすることによって、自治体ニーズ評価で母集団を代表するデータが用いられているかどうかを確かめようとしているのであった。

中核市は2019年4月1日時点で58団体あった。2019年4月に各団体の最新の男女共同参画計画とそこで参照されている住民意識調査の関連資料をインターネットで収集したところ、中核市における最新の男女共同参画計画は全ての団体について第二次以降の計画で、収集できた男女共同参画計画のための住民意識調査は全ての団体について二回目以降の調査だった。

58の中核市のうち、男女平等参画計画策定のために実施された住民意識調査のデータのバイアスについて検討する団体を、図1-5に示したスクリーニング・プロセスで抽出した。まず、2019年4月1日時点の中核市58団体のうち8団体では最新の男女共同参画計画で参照された住民意識調査の調査時点が中核市の指定を受ける前だったため、統合的分

析から除外した⁸⁵。次に、最新の男女共同参画計画で参照されている住民意識調査に関する必要な情報をインターネットで得られなかった9団体を統合的分析から除外した⁸⁶。さらに、標本抽出方法が有意抽出法の場合には選択バイアスが生じるのが当然であるため、分析から除外した⁸⁷。続いて、性別の回収標本サイズの情報を得られなかった団体を統合的分析から除外した⁸⁸。さらに、男女共同参画計画に数値目標が記載されていなかった団体を統合的分析から除外した⁸⁹。最後に、男女共同参画計画の数値目標に住民意識調査に基づく指標を用いていない団体を統合的分析から除外した⁹⁰。ただし、男女共同参画計画に住民意識調査に基づく数値指標を用いていない団体は、意識調査に基づく目標を用いる場合に避けがたい測定バイアスの影響を受けにくい指標を用いているわけであるから、母集団を代表するデータを用いる工夫を講じている団体とみなすことができる。

以上のスクリーニングの結果、男女共同参画計画策定のために実施された住民意識調査のバイアスに関する統合的分析の対象団体は、36団体となった（図1－5）。調査報告書等の資料に母集団サイズが書かれていないケースが多かったため、母集団サイズの記載がない場合には各調査の調査時点に最も近い住民基本台帳人口または推計人口をインターネットで調べてデータの欠落を補完し、母集団の男女比を推計した⁹¹。

⁸⁵ この理由で分析から除いたのは、山形市、福島市、川口市、福井市、甲府市、八尾市、鳥取市、松江市である。

⁸⁶ この理由で分析から除いたのは、青森市、宇都宮市、川越市、越谷市、寝屋川市、明石市、奈良市、高知市、那覇市である。

⁸⁷ この理由で分析から除いたのは、松山市である。

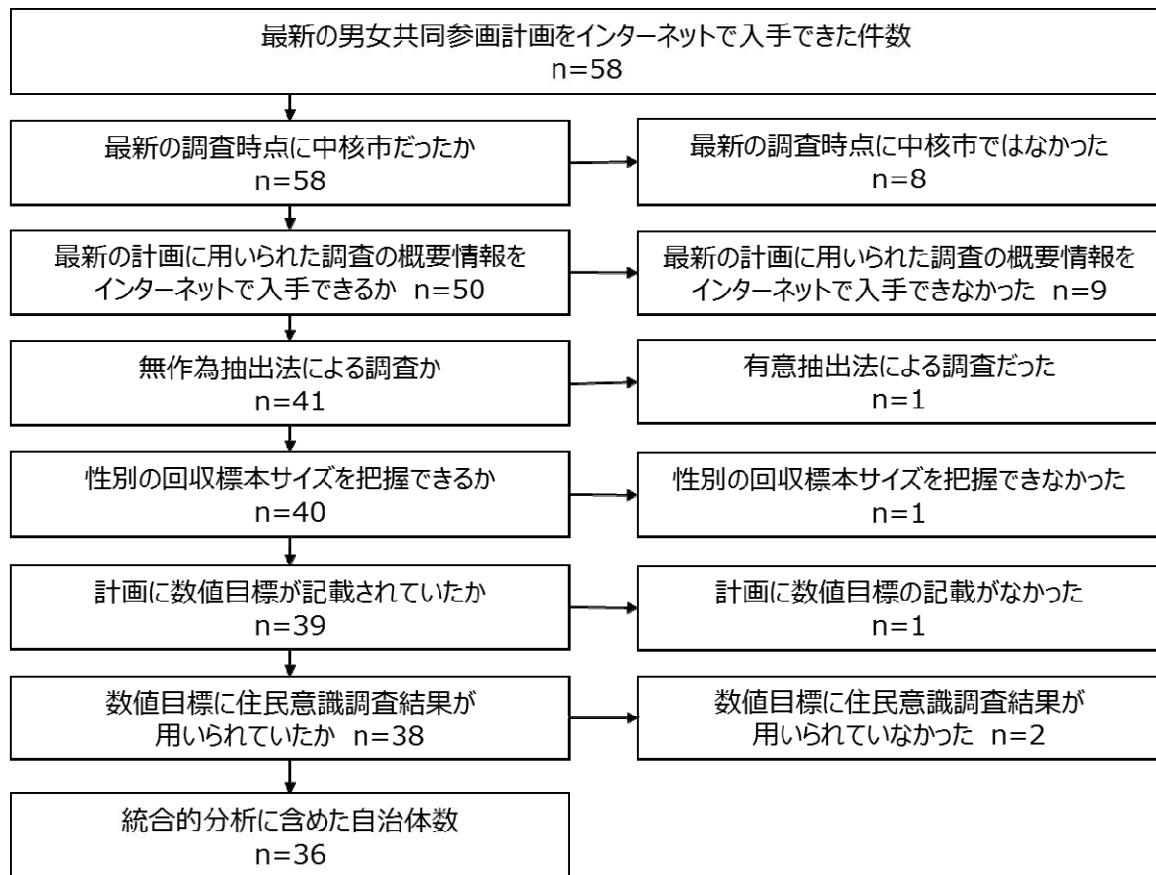
⁸⁸ この理由で分析から除いたのは、久留米市である。

⁸⁹ この理由で分析から除いたのは、和歌山市である。

⁹⁰ この理由で分析から除いたのは、旭川市と前橋市である。これらの団体は例えば「女性農業者の起業件数」や「LGBTの周知回数」のような、回答者の認知のあり様による値の変化が生じにくい指標のみを数値指標に用いていた。

⁹¹ 例えば調査対象が18歳以上なのに5歳刻みの人口データしか自治体公式ホームページで公表されていない場合、各年齢の人口が同数だと仮定して母集団人口の推計値を算出した。

(図 1 - 5) 中核市調査のスクリーニング・プロセス



(注) 調査時点は 2019 年 4 月である。

(出典) 第 3 章第 2 節で述べる PRISMA 声明のフローチャートを参考にして筆者作成。

男女共同参画計画に住民意識調査に基づく評価指標を用いていた 36 団体が母集団を代表するデータを得ていたかについてみていく。住民意識調査の概要を確認すると、新しい男女共同参画計画のニーズ評価を行うために独自の住民意識調査を実施した団体は 34 団体で、残り 2 団体は全庁的に実施している市民意識調査の調査項目の一部を男女共同参画に関する項目に充てていた⁹²。全団体とも調査対象は行政区域内の住民個人で、調査対象年齢は 20 歳以上が 21 団体、18 歳以上が 10 団体、その他が 5 団体であった⁹³。標本抽出法は層別または単純無作為抽出法で、全団体が自記式無記名の調査票を郵送で送付し郵送

⁹² 該当する 2 団体とは、盛岡市と高槻市である。

⁹³ なお調査年度は 2011 年度から 2017 年度に分散しており、最も多かったのが 2016 年度の 13 件、次に多かったのは 2014 年度の 10 件であった。

で回収していた⁹⁴。母集団サイズの平均値は 324,364、計画標本サイズの平均値は 2,760 で、最も多かったのは 3,000 の 13 団体、次に多かったのは 2,000 の 10 団体であった⁹⁵。有効回収標本サイズの平均値は 1,085 で、有効回収率の単純平均値は 39.5%であった⁹⁶。回収標本の女性割合は平均 58.8%⁹⁷、母集団の女性割合は平均 51.9%で、回収標本の女性割合と母集団の女性割合の差は平均 6.9%ポイントであった。回収標本の女性割合と母集団の女性割合に統計的に有意な差がある団体は 30 団体もあり、統計的に有意な差がみられなかった団体は 6 団体しかなかった⁹⁸。しかし回収標本と母集団の属性分布の差異を考慮した分析を行っていたのは 1 団体のみで、当該団体は性別ではなく年齢階層別構成比の差異を補正した集計値を算出していた⁹⁹。なお当該団体の性別構成比は回収標本と母集団との間に統計的に有意な差があるとはいえなかったため、性別構成比の歪みを補正する必要がないと判断したのではないかと推察される。

住民意識調査に基づいて男女共同参画計画の数値目標の実績値や目標値を設定する際に、何らかの方法でデータのバイアスが考慮されていたかを調べたところ、回収標本の女性割合と母集団の女性割合に統計的に有意な差がみられた 30 団体のうち、3 団体は基準値に男女別の値を用いていたが、残りの 27 団体では、特段の考慮をせずに住民意識調査の単純集計値に基づいて基準値と目標値が設定されていた。母集団と回収標本の年齢別構成比については自治体によって年齢階層区分が異なり統合的な分析が難しいため分析しなかったが、一般に質問票調査の回収率は、選挙の回収率と同様に若年層ほど低く高年齢層ほど高い傾向がある。それにも関わらず、年齢階層別の違いを考慮して基準値を設定していた団体は 1 団体しかなかった¹⁰⁰。また、実績値の標準誤差を考慮して基準値や目標値を設定

⁹⁴ オンラインによる回答を受け付けるケースもあった。

⁹⁵ 母集団サイズの最小値は 152,894、最大値は 508,331、計画標本サイズの最小値は 1,000、最大値は 5,000 であった。

⁹⁶ 有効回収標本サイズの最小値は 478、最大値は 2,099。有効回収率の最小値は 30.3%、最大値は 51.8%であった。

⁹⁷ 回収標本の女性割合は、性別無回答を除いた有効回答に占める女性の割合として計算した。

⁹⁸ なお内閣府「男女共同参画社会に関する世論調査(平成 28 年 9 月)」(Retrieved April 11, 2019, from <https://survey.gov-online.go.jp/h28/h28-danjo/index.html>) の計画標本(母集団の男女比の比例割当てと思われる)の女性割合と回収標本の女性割合は統計的に差があるとはいえなかった。

⁹⁹ 該当団体は、下関市である。

¹⁰⁰ 該当団体は、下関市である。

していた自治体は皆無だった。

前述したように、内閣府が 2001 年に示した「市町村男女共同参画計画の策定の手引き」は、男女共同参画分野の住民意識調査では回答者の性や年齢階層によって回答に大きな違いが生じるため注意して分析する必要があることを指摘していた。また、入手できた男女共同参画計画及びそのための住民意識調査は、全て第二次以降のものだった。したがって調査対象団体は、内閣府の手引きや前回調査の反省を踏まえて、データのバイアスの影響を小さくする工夫を講じることができたはずである。それにも関わらず、男女共同参画計画に数値目標を設定していた 38 団体のうち 36 団体が住民意識調査結果を数値目標に用いており、その 36 団体のうち 30 団体で母集団と回収標本の女性割合に統計的に有意な差があり、そのうち男女差を考慮して基準値を設定していた団体は 3 団体のみ、年齢階層別の違いを考慮して基準値を設定していた団体は 1 団体のみ、標準誤差を考慮して基準値を設定していた団体は一つもなかったという事実は、大多数の地方自治体がデータのバイアスに無自覚のまま、政策の標的集団を代表しているとはいえないデータに基づいて住民ニーズを把握し、計画を策定してしまっていることを示唆するのではないかと考えられる。

4－3．検証結果のまとめ

本節では、地方自治体による現実のニーズ評価と理想的なニーズ評価とのギャップを明らかにするため、①証拠を集める前によく考えて系統的に調べているか、②政策の標的集団を代表するデータを用いているか、という 2 つの点について、日本の地方自治体のデータを用いて検討した。①については、X 自治体で過去 5 年間に行われた住民調査が仮説検証型で行われていたか、その結果が二次利用されていたかを調べることで確かめたところ、ニーズ調査またはニーズ評価を目的とする住民調査実施部局の代表的な住民調査 27 件のうち、16 件の調査は仮説がないまま実施されていた。また、二次利用の実績または予定がある調査は 3 件しかなく、調査の詳細が公表された調査は 8 件しかなかった。仮説検証型ではなく思いつき程度で行われるニーズ調査は、深刻な後付けバイアスを引き起こす恐れがある。調査方法や調査結果がきちんと公開されていなければ、調査データの二次利用が行われなければ、第三者がニーズ調査やニーズ評価のバイアスに気づくことはできない。そのため、バイアスのあるニーズ評価に基づいて政策形成が行われている可能性が高いと懸念される。

②については、中核市による最新の男女共同参画計画とそこで参照されている住民意識調査の情報をを用いて、データのバイアスを考慮したニーズ調査やニーズ評価が行われているかを調べた。まず、最新の男女共同参画計画策定のための住民意識調査を実施した時点で中核市であった 50 団体のうち、10 の団体で住民意識調査の概要や回収標本の特徴に関する情報をインターネットで入手できなかった。また、男女共同参画計画に数値目標を設定していた 38 団体のうち 36 団体が住民意識調査の結果を数値目標の設定のために用いており、その 36 団体のうち、数値目標の設定に当たって母集団と回収標本の性別構成比の違いを考慮していた団体は 3 団体のみ、年齢階層別構成比の違いを考慮していた団体は 1 団体のみ、標準誤差を考慮していた団体は皆無だった。多くの地方自治体が政策の標的集団を代表しているとはいえないデータに基づいて住民ニーズを把握し、計画を策定しているのではないかと懸念される。

自治体ニーズ評価は、地方自治体が優先的に解決すべき政策問題は何か、誰が何に困っているのかを具体的に特定するために行うものである。バイアスのあるデータに基づいて政策問題を把握し、バイアスのある価値判断を下すのでは、公的介入を真に必要とする人に効果的・効率的かつ公平に政策を届けることはできない。ではどうすれば、自治体ニーズ評価の理想と実態とのギャップを小さくすることができるのだろうか。

5. 自治体ニーズ評価の理想と実態のギャップを埋めるために

5-1. ニーズ評価の理想と実態のギャップが生じている原因

米インテル社の最高経営責任者(Chief Executive Officer:CEO)だった Andrew S. Grove は、世界的ベストセラーとなった主著『High Output Management』において、人が仕事をしない理由は能力不足でできないか意欲不足でやろうとしないかのどちらかしかないと論じた¹⁰¹。地方自治体によるニーズ評価が理想的なニーズ評価と乖離している理由も、能力不足でできないのか、意欲不足でやる気にならないのかのどちらかだと整理できるだろう。

一般に、能力ないし資源不足と意欲不足との関係は双方向的で、ニワトリと卵のような関係にある。やる気がないと能力向上に資源を投じないので能力が高まらないし、能力や

¹⁰¹ Grove (1983) p. 157 (訳 p. 232) を参照。

資源が足りないと理想通りに活動できないので意欲が削がれる。しかしどちらがより根本的な問題かと問われれば、意欲不足の方がより根本的な問題ではないかと考えられる。能力ないし資源の不足によって理想的なニーズ評価ができないのなら、地方自治体は重要性の低い事業の予算を削って何らかの対策を講じることができるはずである。例えば、専門的スキルをもつ外部の専門家または専門業者にニーズ評価を依頼したり、社会調査法や統計学の専門知識とスキルをもつ職員を採用するか自前で育てて、その職員が全庁的にニーズ評価の質チェックを担う体制を整えることは、少なくとも中核市クラスの規模の自治体なら不可能ではないだろう。しかし X 自治体では、庁内で実施されている社会調査の実態を一元的に把握する仕組みはなく、調査データを二次利用する実績も予定も殆どなく、調査結果の公表も低調だった。中核市の男女共同参画計画策定のために行われる住民意識調査では、回収標本の性や年齢階層の歪みが深刻なバイアスを引き起こすことが知られているはずなのに、大多数の自治体はその対策を講じずに、歪みのあるデータに基づいて数値目標を定めていた。これらの調査結果は、ニーズ評価の理想と現実乖離がある根本的な原因は能力不足ではなく意欲不足のためであることを示唆しているのではないかと推察される。

もしそうであるなら、地方自治体が理想的なニーズ評価を行う意欲に欠ける理由は何だろうか。意欲不足の背景には、無知と意図的な無視がありうる。ニーズ評価の意義をよく知らないからやる気にならないのか、それらを知っているもののインセンティブに欠けるのでやる気にならないのかによって、解決策は大きく異なりうる。無知が原因なら、ニーズ評価の意義を理解し、バイアスのあるニーズ評価がもたらす問題を認知することで行動が変わり、より理想的なニーズ評価が実施されるようになるかもしれない。他方で意図的な無視が原因なら、理想的なニーズ評価の手続き要件を踏むことを予算措置の条件にするといったインセンティブ装置が必要だろう。後者の対策は第3章で述べるインパクト評価の方法論や第4章で述べる失敗からの学習意欲を高める方策と重なるため、次節ではニーズ評価の意義の理解を促すであろう事例について検討する。

5-2. ニーズ評価の意義（1）一貧困の発見と再発見

ニーズ評価の意義を端的に示す典型的な政策問題は、貧困問題である。貧困とは単なる事実を指す概念ではなく、社会が見過ごすことのできない状態とは何かという価値判断を

含む概念である。貧困を測定バイアスや推論や主観を伴わずに、客観的に定義することはできない。したがって貧困対策の歴史は、貧困の実態と原因について科学的に信頼できるデータとエビデンスを集め、それらに基づいて貧困問題を定義し直して対策を講じるという、ニーズ評価による問題発見と再発見が繰り返される歴史であった。岩田（2007）の言葉を借りれば、貧困の歴史は、「社会が責任をもって解決すべき状態と、個人や家族に委ねておけばよい状態との境界設定をどうするか」という「境界設定についての議論の歴史だ」といっても過言ではない¹⁰²。

産業革命が進む英国では、社会調査とジャーナリズムのおかげで 19 世紀半ば頃から、貧困が浪費という個人の責任よりも低賃金という社会の責任に起因する政策問題であると認知されるようになった。中でも貧困を科学的に測定評価したことによって政策や社会に大きな影響を与えたのは、Charles James Booth によるロンドンでの貧困調査と、Benjamin Seebohm Rowntree によるヨーク市での貧困調査である¹⁰³。Booth は貧困線という概念を初めて測定に用いて貧困者の生活実態を調査し、約 3 分の 1 のロンドン住民が生命を維持するギリギリの水準に定めた貧困線の所得水準を下回っていることを明らかにした¹⁰⁴。Rowntree の社会調査はその Booth の先行研究や政府統計、栄養生理学における当時最先端の科学的知見である栄養価計算などを取り入れて行われたもので、郊外の地方都市ヨークでも 3 割近い労働者が絶対的貧困状態にあることが明らかにされた¹⁰⁵。彼らの研究が示したのは、貧困は個人だけの問題でも大都市だけの問題でもなく、都市で働く賃労働者の賃金が低いという構造的な要因に起因するという事実である。この事実が科学的な方法で測定され、第三者が妥当性を検証できる形式で公開されたことによって、絶対的貧困は社会が解決すべき重大な政策問題であるという認識が社会に広がり、英国政府による公的年金制度や公的医療保険制度の導入に繋がったのであった¹⁰⁶。

Rowntree の社会調査から約 80 年後に行われた、Peter B. Townsend らによる英国における貧困の再発見は、Rowntree 型の貧困研究に欠落した視点を補い、貧困の概念を刷新す

¹⁰² 岩田（2007）p. 32 及び p. 35 より引用。

¹⁰³ 貧困研究の歴史については主に、ラヴァリオン（2018）pp. 15-168 及び岩田（2007）を参照。

¹⁰⁴ Booth（1892）を参照。

¹⁰⁵ Rowntree（1901）を参照。

¹⁰⁶ 戦後の英国政府の社会政策に大きな影響を与えたベヴァリッジ報告（Beveridge [1942]）における貧困の定義では、Rowntree の定義が採用された。

ることを目的として行われたものだった¹⁰⁷。Townsend は、その社会で普通とされている通常の社会生活を送ることが難しいという「相対的剥奪」(relative deprivation) の概念を用いて相対的に定義しなければ、貧困を客観的かつ首尾一貫したかたちで定義することはできないと主張した。Rowntree 流の「絶対的貧困」定義の核心は、最低限の肉体的能率を維持するための食糧等の入手費用によって最低限の生活水準を定義する点にあった。そのため、例えばあるコミュニティにとって紅茶が社会生活を営むために不可欠な生活必需品であるとしても、紅茶を購入するための費用は最低限の生活を送るために必要なコストだとみなされない。しかし、絶対的で科学的な貧困と、共同主観的で慣習的な貧困は異なる。ニーズとは個人がどのような社会—職業、学校、家族、地域コミュニティ等—to に属するかによって変化するものであり、慣習は単なるファッションではない。例えば子どもにとっての生活必需品は、教育制度や健康に関する社会規範等の慣習の影響を大きく受けるものである。それと同じように、同じ栄養価の食品のうち何を摂取するかは、社会の慣習や習慣によってかなりの程度規定されている。したがって食糧一つをとっても、社会的慣習を無視して貧困を客観的に定義することはできないはずである¹⁰⁸。このような考え方に基づいて、Townsend らは住環境、雇用状況、収入、健康と障害、生活様式等の 9 領域 39 ページにわたる調査票を用いて、英国の 3,000 以上の世帯と 1 万人以上の個人に対する相対的貧困調査を 1960 年代末に実施した。調査項目には、家にテレビや冷蔵庫や電話があるか、子どもの誕生日を友人を呼んで祝ったか、同年代の人々と比べて暮らし向きはどうか等の、所得では捕捉しきれない豊かさを測るための豊富な項目が用意されていた。そして調査分析の結果、所得と所得以外の相対的剥奪指標との間には統計的に有意な非線形の関係があり、ある一定の所得を下回ると相対的剥奪度合いが急激に増大すること等が明らかにされ

¹⁰⁷ 「貧困の再発見」とは、西側西洋諸国で「福祉国家の黄金時代」といわれる 1960 年代やそれ以降の経済的富の総体が増大している時代にも、豊かさに取り残された貧困者が数多く存在するという事実が、社会に広く知れ渡るようになった現象をいう。米国においてその社会認知に大きく貢献したのは John Kenneth Galbraith の『豊かな社会 (*The Affluent Society*)』(1958 年)や Michael Harrington の『もう一つのアメリカ (*The Other America*)』(1962 年)を嚆矢とする一般向け書籍で、戦後に充実してきた理論やデータに基づいて経済成長の陰で見えにくくなっていた貧困者の悲惨な実態が鮮やかに描かれたことが、世論や政治運動の高まりなどを引き起こしながら、米 Lyndon B. Johnson 政権の「貧困との戦い (War on Poverty)」と呼ばれる一連の社会政策の実現に繋がっていった (ラヴァリオン [2018] 第 2 章を参照)。

¹⁰⁸ Townsend (1979) pp. 31-60 を参照。

たのだった¹⁰⁹。

Townsend らの調査は、貧困を社会的剥奪度合いによって定義し、生活水準そのものを測る方法を提案した画期的なものであった。このアイディアに基づいて Townsend が開発した子どもの貧困の定義は、後に英国政府や EU (European Union : 欧州連合) で採用され、1990 年代末における英国政府による子どもの貧困に関する達成目標の設定や、2010 年英国子どもの貧困法 (Child Poverty Act) に繋がった¹¹⁰。日本でも、子どもの貧困問題の社会的認知に大きな影響を与えた阿部 (2008) や、日本の地方自治体による子どもの貧困実態調査で、Townsend 流の社会的剥奪調査が行われている¹¹¹。

1900 年頃の英国におけるワーキングプアも、1960 年代以降の英国における豊かさに隠された相対的貧困者も、まさにニーズ評価の手法によって発見されたものであった。すなわち、統計や先行研究を踏まえて社会問題を定義し、何を貧困とみなすかという価値基準を科学的方法で設定した上で系統的に独自調査を実施し、政策の標的集団を代表するようなデータを得て貧困の実態を把握し、貧困の原因を説明するエビデンスを作り出して詳らかに公開したことによって、貧困問題は社会に発見され、認知され、政策形成に繋がっていったのであった。

¹⁰⁹ Townsend (1979)。阿部 (2008) pp. 201-205 も参照。なお Townsend らの調査で、何を生活必需品とみなして相対的な剥奪度合いを調べるかを決定する際に用いられたエビデンスは事前に実施された質的調査結果であり、信頼性が高いとは必ずしもいえない定性的手法に基づいていた。その限界を克服して剥奪指標の信頼性を高める方法として、後に「合意基準アプローチ」という方法が提案された。合意基準アプローチとは、社会にとっての生活必需品を研究者が選ぶのではなく、無作為抽出された社会構成員に選んでもらい、回答者の過半数が「絶対に必要である」と答えた項目だけを社会的必需品とみなすという手法である。属性の異なる集団間の回答傾向の差を統計的に検討した上で社会的必需品項目を絞り込むため、社会科学的に信頼性の高い方法だとみなされている (阿部 [2008] pp. 180-184)。

¹¹⁰ Watts (2009) を参照。

¹¹¹ 阿部は、2003 年と 2008 年に、合意基準アプローチを用いて日本における子どもの社会的必需品調査を実施し、他の先進諸国と比べて日本の一般市民による子どもの必需品への支持率が大幅に低いことを明らかにしている (阿部 [2008] pp. 184-210)。地方自治体による調査事例については、例えば首都大学東京子ども・若者貧困研究センター (2017) 『東京都子供の生活実態調査報告書【小中高校生等調査】』 (Retrieved from August 11, 2019, <http://www.fukushihoken.metro.tokyo.jp/joho/soshiki/syoushi/syoushi/oshirase/kodomoseikatsujittaityousakekka.html>)。

5 - 3. ニーズ評価の意義（２）—柔道事故の可視化のインパクト

データやエビデンスがあれば、社会が変わるわけではない。しかし、信頼できるデータやエビデンスによって社会の構造的な問題を可視化することができれば、権力がなくても政策問題への社会の気づきを促し、社会を動かすことができる。そのインパクトを端的に示している事例は、貧困問題だけではない。日本の学校の部活動における柔道事故問題も、確かなデータとエビデンスを丁寧に集めて分析することによって浮かび上がってきた政策問題であった。

教育社会学者である内田良は、エビデンスに基づく学校事故・事件の実態解明に取り組み始めた 2006 年に『学校の管理下の死亡・障害事例と事故防止の留意点』という毎年刊行されている資料を探し当てた。その当時、学校の安全や子どもの安全は事故の実態に関するエビデンスによる検証がないままに理想論や精神論やエピソードベースで議論されており、学校事故の全体像を把握できるデータはなかったという。だが同資料をみれば、幼稚園や保育所から高等専門学校までを対象とする加入率 96%の災害共済給付制度によって把握された、学校管理下で起きた死亡・障害事例のデータを得ることができる。ただし、それを活用するためには紙媒体に印刷されているアナログデータを 1 つ 1 つデータベースに登録して、データを解釈して社会問題を発見していく必要があった。そこで内田らは、重大事故の防止が最優先という価値基準に基づき、1983 年から 2011 年に発生した死亡事例 118 件を 1 件ずつカードに貼り付け、カテゴリ別に分類するという作業を地道に行っていた¹¹²。

その集計・分析過程で内田が発見した重要な事実、柔道事故では死亡件数が多く死亡率も突出して高いことと、死亡事例の多くは中学校と高等学校の部活動中に発生しており、初心者が頭部外傷を負って死亡に至ることが多いという事実であった¹¹³。しかしその事実は、柔道の専門家集団にも社会にも殆ど認知されていなかった。柔道による死亡事故は主に頭部外傷対策を適切に行っていないことによって生じているという事実も、学校関係者や柔道の専門家は認知していなかった¹¹⁴。

¹¹² 内田（2013） pp. 16-18；内田（2015） pp. 27-34 を参照。

¹¹³ 内田（2013） pp. 19-57 を参照。

¹¹⁴ 内田（2011）；内田（2013） pp. 67-118 及び pp. 132-134 を参照。柔道による死亡事故の主な原因は、柔道という競技そのものの危険性ゆえではなく、頭部外傷の危険性が十分に認知されていないために頭部外傷後の休養が不十分だったり、投げ技で安全に着地させ

内田（2013）によれば、柔道事故が社会に認知されるようになったのは 2010 年頃からであり、そこで決定的な役割を果たしたのは、2012 年度からの中学校における武道必修化、2010 年の全国柔道事故被害者の会（以下「被害者の会」という）設立、そして内田の公表したデータや学術論文であった¹¹⁵。柔道事故の社会問題化にとってとりわけ重要だったのは、それまで総数すら把握されてこなかった柔道による死亡事故の情報を内田が 2009 年 9 月にインターネットでデータ公開し、柔道の死亡事故は個別の偶発的な事象ではなく社会の集合的な事象だと示したことにある。そのデータはすぐに柔道による死亡事故の被害者家族の人々の目に留まり、そのことが翌年 3 月の「被害者の会」設立を促し、中央政府やマスコミの注目を引き起こし、学校の部活動をはじめとする柔道の指導現場での頭部外傷対策の強化に繋がっていった。このようにして柔道事故が社会問題化したことがおそらく大きく影響して、学校における柔道による死亡事故の減少という現象がもたらされたのであった¹¹⁶。

6. 事例分析から得られる示唆

貧困問題と柔道事故の事例は、ニーズ評価が社会変革に向けた大きな原動力になりうることを示している。社会が責任をもって解決すべき政策問題とは何かについて明確な価値基準を設定した上で、データを丹念に積み上げて科学的に信頼できるエビデンスを作り出し、社会が対処しなければならない構造問題を可視化することができれば、それまで個別の特殊な問題として無視あるいは軽視されてきた問題が社会の対処しなければならない政策問題として認知され、社会を変える大きな力になりうる。これがニーズ評価のインパク

ることの重要性が周知されていなかったりするために生じるものであり、スポーツ科学に基づく安全対策が日本では不十分であったことに起因していたという。

¹¹⁵ 内田（2013）pp. 60-67 を参照。なお内田自身の影響については、山本（2013）pp. 26-33 を参照。

¹¹⁶ 内田（2013）pp. 9-13 及び pp. 135-139 並びに内田（2015）pp. 18-34 及び pp. 205-235 を参照。例年数件あった柔道による死亡事故が、2012 年から 2014 年には 0 件になったという。ただし 2015 年には、柔道による死亡事故が再び発生してしまった（「被害者の会」ホームページのニュース記事より。Retrieved January 10, 2020, from judojiko.net/category/news/page/5）。データに基づくニーズ評価による政策問題の可視化は問題解決の重要な契機になりうるが、政策問題を根絶するためには、問題解決に向けた体系的で継続的な取り組みが必要であることが示唆される。

トである。

ニーズ評価には、常にバイアスや不足が残るものである。例えば、Booth と Rowntree が英国の都市におけるワーキングプアの存在を知らしめた頃の英国では、彼らの調査研究の他には信頼できる社会調査データがなく、しかも両者が示す絶対的貧困者割合が3割前後と非常に近かったために、英国の人口の約3割が絶対的貧困者だというのが1900年頃の英国の通説だったという。しかし後の研究で、Booth が用いた貧困線と Rowntree が用いた貧困線を比べると後者の方が前者よりもかなり高く、したがって後者の絶対的貧困者割合は前者のそれよりも高く見積もられていたことが明らかにされている¹¹⁷。しかし完璧な評価が存在しないからといって、所与の資源制約の下でも努力すれば防げたであろうバイアスを放置して解決すべき政策問題を見逃したり、標的集団のニーズを誤って特定すべきではないだろう。

理想的なニーズ評価が充たすべき要件は、①証拠を集める前によく考えて系統的に調べる、②政策の標的集団を代表するデータを利用する、という2つに整理できるのだった。そのスキルを高めるために必要なことは政策分野や政策のステージや組織能力の成熟度等によって多様でありうるが、理想的なニーズ調査を系統的に実施するためのチェックリストや、既存の学術研究や政策の業績情報を一元的に得られるエビデンスプール、統計や行政記録を用いたデータ分析を容易にするデータ基盤等を整えることによって、地方自治体のニーズ評価能力は大きく高まると期待される。しかしまずもって必要なのは、手段を整えることではなく、行政が解決しなければならない政策問題をバイアスなく特定して社会を善くしなければならないという強い目的意識をもってニーズ評価に取り組むことではないかと思う。目的意識が明確であれば自ずと様々な工夫が講じられると考えるのは楽観的に過ぎるであろうが、手段をどれほど整えても目的達成に向けた努力がなされなければ、むしろ害ですらありうる。公的介入を真に必要とする人々が本当に求めている行政サービスを届けるために、ニーズ評価の理想と実態のギャップを埋める努力が地方自治体に求められている。

¹¹⁷ ラヴァリオン（2018）p. 93 を参照。

第2章 達成目標が充たすべき要件

ニーズ評価によって政策問題を定義し標的集団のニーズを明確化したら、政策体系を組み立てて実行計画を作成するプロセスに進む。その第一歩は、政策の使命と達成目標を定めることから始まる。使命とは組織あるいは政策の存在理由を簡明に表した定性的記述のことで、何をなぜ誰のためにするのかを簡潔に示したものである。達成目標とは、政策が実現を目指すアウトカムに関する定性的または定量的な記述のことである¹¹⁸。

目的地を知らずして、そこへ行き着くことはできない。達成目標が不明瞭だと政策の成否を判別することができず、政策の成否が分からなければ成功から学ぶことも失敗を改めることもできない。したがって適切な達成目標を定めることは非常に重要である。では、ニーズ評価に理想的な手続きがあったように、社会科学で推奨される達成目標の充たすべき要件はあるのだろうか。あるとすれば、それはどのようなものだろうか。

本章では、達成目標が充たすべき要件を検討する。第1節では、達成目標が充たすべき要件を文献に基づいて整理する。第2節では人々の潜在能力の向上に繋がったと考えられる事例を取り上げて、その事例で設定されていた達成目標が本稿の考える達成目標の充たすべき要件に合致していたかを検討する。最後に本章の結論を述べる。

1. 達成目標が充たすべき要件は何か

1-1. 達成目標が充たすべき要件（1）—政策科学の文献に基づいて

第1章でニーズ評価の理想的手続きを特定するために主に用いた文献は、Rossi et al. (2004) と Bardach and Patashnik (2016)であった。後者には達成目標が充たすべき要件に関するまとまった記述がなかったので、代わりに米国行政学会・行政経営センター (2001) を実用的評価を重視する代表的な基本書として用いる。同書は、米連邦政府機関

¹¹⁸ Rossi et al. (2004) ; Weiss (1998) を参照。最上位の目標を「goals」と呼び、「goals」を実際に測定評価できるように操作化した目標または「goals」より下位にある目標を「objectives」と呼ぶことが多い。ただし用語の定義は文献によって異なる。例えばハトリ（2004）は、「goals」という用語は特定の数値目標という意味で用いられることも多いという理由から、最上位目標を「objectives」と称し、その到達度に関する目標は、その他の業績指標ごとの目標と同じように定めればよいとしている。

にアウトカム重視の業績測定を義務づけた 1993 年米連邦政府業績成果法（Government Performance and Results Act of 1993：GPRA）に準拠した米連邦政府及び地方政府向けのマニュアルで、米国の行政現場で多用されたテキストである¹¹⁹。

Rossi et al. (2004) は達成目標が充たすべき要件として、以下の 4 要件を挙げている¹²⁰。

①標的集団に特有なニーズに合致しており、重要である。

②政策が説明責任を負うアウトカムの達成度を測定できるほどに、明確かつ具体的に記述されている。

③政策によって実際に達成できるほどに、現実的である。

④政策以外の外部要因がアウトカムに及ぼす影響は小さく、政策と直接的に関係がある。

上記①の要件を「重要性」、②の要件を「可測性」、③の要件を「現実性」、④の要件を「関連性」と呼ぶことにする。

米国行政学会・行政経営センター（2001）は、行政機関や政策が期待するアウトカムを説明した大まかな記述を「目的(goals)」と呼び、以下の 4 要件を充たすべきとしている¹²¹。

¹¹⁹ 米国行政学会・行政経営センター（2001）p. i を参照。GPRA 法は日本の中央政府及び地方自治体における業績測定型の政策評価に大きな影響を与えた制度で、Barack Obama 米前大統領期に、2010 年米政府業績成果法の近代化法（The GPRA Modernization Act of 2010：GPRAMA 法）へと改正されている。

¹²⁰ Rossi et al. (2004) pp. 143-154 に基づいて整理した。同書では「goals」と「objectives」が特に区別されずに、目標について一般的な説明がなされている。

¹²¹ 米国行政学会・行政経営センター（2001）p. 23 より引用した。ただし同書の用語法は本稿の用語法と異なるため、④の達成目標の後ろに「(objectives)」を追記した。

同書は「目的 (goals)」を数値で測定できるようにしたターゲットを「達成目標 (objectives)」と呼び、以下の要件を充たすべきだとしている（米国行政学会・行政経営センター [2001] pp. 24-26 を参照）。

①アウトカム志向・成果志向であって、プロセス志向ではない。

②使命と目的に一致している。

③SMART である。すなわち、(a)「Specific」（具体的で）、(b)「Measurable」（測定可能で）、(c)「Ambitious」（意欲的で）、(d)「Realistic」（現実的で）、(e)「Time-bounded」（期限がある）。

なお、「Time-bounded」は「時宜に叶った」と訳されていたが、「期限がある」という訳の方が適切ではないかと思われるので訳を変えた。SMART を異なる単語の頭文字だとして、「A」に「Achievable」（達成可能で）、「R」に「Relevant」（関連があり）、「T」

- ①行政機関の優先事項
- ②行政機関の将来の行動に対する明確な方向
- ③行政機関の仕事からどのような影響、またはアウトカムが生ずるか
- ④明確な達成目標（objectives）を策定するための基礎

上記4要件のうち、①の要件は「重要性」、③の要件は「関連性」、④の要件は「可測性」だと考えられるから、Rossi et al. (2004)と重複する¹²²。重複しなかった②の要件を「具体性」と呼ぶことにする¹²³。

これらの要件が、なぜ達成目標が充たすべき要件とされているのか確認しておこう。ニーズ評価の推奨手続きは、ニーズ評価に用いるデータやエビデンスの「バイアスをできるだけ小さくすべきである」という価値基準のもとで演繹的に求められる推奨手続きであった。達成目標が充たすべき要件のうち、「重要性」、「具体性」、「可測性」及び「関連性」は、解決すべき政策問題と目標とのズレや目標と手段とのズレという意味での「バイアスをできるだけ小さくすべきである」という価値基準のもとで演繹的に求められる要件だと考えられる。すなわち、「重要性」は、ニーズ評価によって特定された優先すべき政策問題とのズレのない重要な目標を定めるべきであるという規範的要請から求められる要件だと考えられる。「具体性」は、何を達成すべきなのかを明確に示すことで目的と手段の乖離を防ぐための要件であろう。「可測性」は、測定評価のブレを小さくするために必要な要件であろう。「関連性」は、政策以外の外部要因の影響の大きい目標を設定するほど介入による効果のみを取り出してインパクトを評価することが難しくなるから、インパクト評価のバイアスを小さくするための要件だと考えられる。他方で「現実性」という要件には、解決すべき政策問題とのズレを補正したり、評価の誤りを小さくする機能を期待し難いので、バイアスをなくすための基準ではなく、達成不能な目標を設定して無益な評価をしないようにするための基準ではないかと考えられる。よって、政策科学の基本書が規定する達成目標が充たすべき要件は、「現実性」という要件を除けば、評価のバイアスやバラつきを小さく

に「Timely」（時宜に叶った）という単語が当てられる場合もある。

¹²² 「④明確な達成目標を策定するための基礎」という要件を可測性とみなす理由は、前掲注で述べたように米国行政学会・行政経営センター（2001）は「達成目標（objective）」を「目的（goals）」を数値で測定できるようにしたターゲットと定義しているためである。

¹²³ この要件を「具体性」と称するのは、後述する Locke and Latham（1990）における specificity と同じことを意味していると考えられるからである。

するという価値基準のもとで必要性を演繹できる性質の要件ではないかと考えられる。

第1章で検討したニーズ評価の理想的手続きについては、政策問題の分布や標的集団の行政ニーズをバイアスなく特定するための手続きが「良い」と考えられるから、社会調査法や統計学に基づいて演繹的に推奨される要件を理想的な手続きだとみなせばよかった。しかし達成目標に求められるのは、ニーズ評価とのズレやインパクト評価のバイアスを小さくすることだけではない。介入によるインパクトを大きくするような目標であることも同時に求められる。では、業績を高める目標とはどのようなものだろうか。次に、達成目標が充たすべき要件を科学的なエビデンスに基づいて帰納的に導出している研究をみていこう。

1-2. 達成目標が充たすべき要件（2）—社会科学のエビデンスに基づいて

目標と業績の関係に関する代表的な研究は、過去 25 年間にわたって蓄積された約 400 ものラボ実験とフィールド実験のエビデンスに基づいて「目標設定理論 (Goal Setting Theory)」を構築した Locke and Latham の一連の研究であろう¹²⁴。Locke and Latham (1990) の最も重要な発見は、目標の難しさ (difficulty) と具体性 (specificity) が業績を高めるという事実であった。努力や業績に対する難易度の説明力は大きく、メタ・アナリシスの結果、努力や業績に対する難易度の効果量 (effect size) は 0.52 から 0.82 であったという¹²⁵。難易度が業績と正の相関関係にあるのは達成が難しい目標ほど努力する意欲を高めるからで、努力しても達成できないほどに難しすぎる目標を設定すると、逆に業績低下がみられるという。Locke and Latham (1990) のいう目標の「難しさ」は、本稿のいう「現実性」に関係する指標で、能力の及ばない目標は業績を低めるが、現状の延長で簡単に達成できるような目標よりも達成が困難な意欲的な目標がよいということなので、「現実性」という要件を「現実性かつ意欲性」と修正する。

他方で Locke and Latham (1990) のいう具体性は、それだけで業績を高める効果をもつわけではなく、難易度が的確である場合に目標が具体的であるほど業績を有意に高める

¹²⁴ Locke and Latham (1990) ; 同 (2002) ; 同 (2006)。

¹²⁵ メタ・アナリシスとは過去に行われた複数の独立な研究結果を統合するための統計解析のことである。効果量とはある検定における検定力の大きさを表す指標のことで、この場合には介入集団と対照集団の平均値の差を標準偏差で割って標準化した値のことである。メタ・アナリシスについては丹後 (2002) を、効果量については南風原 (2014) を参照。

効果をもつという。具体性が業績向上効果をもつのは、「最善を尽くせ。(Do one's best.)」のような抽象的目標では曖昧すぎて何をすればよいかよく分からないので行動変容に繋がりにくい、具体的な目標なら何をすればよいか明確なので行動変容が促されるからである。ただし、新しい複雑なタスクで、辛抱強く努力することより新しいスキルを習得しながら状況に適応することが求められる場合には、具体的な業績目標（performance goals：遂行目標）ではなく学習目標（learning goals）を設定の方がよいという。彼らのいう目標の「具体性」は、操作化しやすいという意味での具体性ではなく、行政機関が何をすべきか明確に分かるという意味での具体性なので、本稿でいう「具体性」と同一だと考えてよいだろう。

目標達成確率を高めるためには、シンプルで分かりやすい目標を設定すべきだとする研究も多い。例えば Soman and Zhao (2011)は、人々の貯蓄行動の変化に大きな影響を与えるのは「履行意思 (implementation intentions)」であり、単一の目標の方が複数の目標よりも行動変化を促す力が強い、目標達成に結びつくことを主張する¹²⁶。複数の目標を設定すると目標間で競合関係が生じてしまい、行動する前によく考えようという気持ちになってしまうからではないかと彼らは推論する。選択肢が増えると意思決定が麻痺して合理的な意思決定ができなくなるという現象は、外科手術をするか投薬を試すかという医師の意思決定でもみられる。ある実験によれば、外科手術をしないで済む新薬が1種類提示されたときは72%の医師が外科手術ではなく投薬を試すと答えたのに、新薬が2種類提示されたときには外科手術ではなく投薬を試すと答える医師は53%に下落したという¹²⁷。多すぎる選択肢がセルフコントロールを消耗させ、目標達成を阻むことを示すエビデンスは数多く存在する¹²⁸。本稿のいう「具体性」は行政機関が将来取るべき行動に明確な指針を与えるという意味なので、シンプルで分かりやすい目標を設定すべきであるという要請は「具体性」に含まれるとも解釈できるが、一方で目標の中身が具体的であることと目標の数が少なく何を優先すべきかが明確であることは異なると整理することもできる。後者の要件を「簡潔性」と称することにする。

目標と業績との関係が最も強まるのは、目標達成へのコミットメントがあるときである

¹²⁶ 「履行意思 (implementation intentions)」については、Gollwitzer (1999)；Gollwitzer and Sheeran (2006) も併せて参照。

¹²⁷ Redelmerier and Shafir (1995) を参照。

¹²⁸ 自制していると自制心が消耗して熟慮が難しくなるという現象を「自我消耗 (ego depletion)」という。Baumeister et al. (1998) 及び Baumeister et al. (2007) を参照。

という。コミットメントがあるとは、単なる口約束ではなく、実効性を高める措置が講じられているという意味である。目標達成の実効性が高まるのは、目標が重要である場合と、目標達成が可能だと信じられるときだとされる¹²⁹。ここでいう目標が重要である場合とは、目標達成が首長の公約であるとか、目標に納得している等の理由で目標達成に強く動機づけられているという意味なので、本稿でいう「重要性」とは異なる。また、目標達成が可能だと信じられるとは、「有能感」(self-efficacy)という主観的認知に関する概念なので、能力に比して可能かどうかという「現実性」とは異なる。よって目標達成へのコミットメントがあるという要件を「実効性」と呼ぶことにする。

まとめよう。これまでの考察を総合すると、達成目標が充たすべき要件について次の命題を得ることができる。

現実的だが意欲的かつ具体的で、実効性を伴う簡潔な目標が業績を高める。それに加えて、重要性が大きく、政策によって直接的に達成できる測定可能な目標であれば、目標設定や評価のバイアスを小さくすることができる。

良い達成目標が充たすべき要件を列挙するに当たっては、手段の目的化を防ぐために実質要件から形式要件の順に配置する方が望ましいと考えられる。よって、厳密ではないが①重要性、②具体性、③現実性かつ意欲性、④関連性、⑤簡潔性、⑥実効性、⑦可測性という順に付番しておく。

個々の要件は数多くの実証的エビデンスに支持されているか、「目標設定や評価のバイアスとバラつきを小さくする」という価値基準のもとで必要な要件である。しかし、これら7つの要件を備えた達成目標が本当に日本の地方自治体による政策の業績を高め、政策評価のバイアスを小さくするかどうかは、実証研究で明らかにしなければならない研究課題である。その検証方法は、例えば達成目標を設定する際に上記7つの要件をチェックリストとして用いた場合と、異なる要件をチェックリストとして用いた場合と、何のチェックリストも用いなかった場合で政策効果に違いが生じるかをランダム化比較試験で調べるといった対照群のある方法が望ましいであろうが、そのような実験を行うのは容易ではない。そこで本稿では、潜在能力の拡大という目的に合致する政策で注目すべき成果を上げたと考えられる事例において、上記に挙げた7つの要件を充たす目標が設定されていたか

¹²⁹ Locke and Latham (2002) p. 707.

を定性的に評価することによって、7つの要件の妥当性を検証する。取り上げるのは、日本の自治体障害者就労支援策において優れた成果を上げたと思われる、岡山県総社市役所¹³⁰の「障がい者千人雇用事業」¹³¹（以下「千人雇用事業」という）である。

2. 岡山県総社市「障がい者千人雇用事業」の事例分析

千人雇用事業は、2011年度から2017度にかけて実施された、総社市役所による障害者就労支援策である¹³²。管見によれば、おそらく本稿が千人雇用事業の初めての定量的なインパクト評価研究になる¹³³。したがって、最初になぜ千人雇用事業を取り上げるのかを説明し、次に、同事業の概要と主な成果を整理する。続いて、同事業のインパクト評価方法と評価結果を述べ、介入効果があったと確認できた場合に、同事業の目標が本稿の考える達成目標の7要件を充たしていたかを検討する。

2-1. 千人雇用事業を取り上げる理由

千人雇用事業を取り上げる第一の理由は、障害者就労の実態は総計の量的把握も十分に なされていないため、データに基づいて実態を明らかにすることの社会的意義が大きいと 考えられることにある。日本を含む先進7か国の障害者雇用制度は、機会平等型、義務雇 用型、公的保証型に大別できる¹³⁴。日本は義務雇用型で、一定規模以上の事業主に対して

¹³⁰ 正式名称は「総社市」だが、行政区域名と行政機関名が同一だと地域を指しているのか行政機関を指しているのか判別するのが難しく、行政機関名を指すことを明示する方がよいと思われる場合には、「総社市役所」と称する。

¹³¹ 総社市役所は「障がい者」という表記法を用いているが、国の法令上は「障害者」という表記法が用いられている。本稿では、総社市役所による障害者就労支援施策に関連する場合には「障がい者」という表記法を、一般用語として用いる場合には「障害者」という表記法を用いる。

¹³² 千人雇用事業の事業開始年度は2011年度、当初の終了予定年度は2015年度であった。しかし後述するように、障害者雇用1,000人を達成するまで当該事業は継続され、2017年9月に「障がい者千五百人雇用事業」へと発展的に解消した（千人雇用事業の終了時期については、2020年1月に筆者が総社市福祉課職員に対して実施した集団面接聴取結果に基づく）。

¹³³ 千人雇用事業について分析評価している学術論文はある（例えば薬師寺・清水[2013]）。

¹³⁴ 遠山（2001）を参照。比較対象国は、ドイツ、デンマーク、フランス、日本、スウェ

従業員の一定割合（以下「法定雇用率」という）以上の障害者を雇い入れることを義務付けるという障害者雇用率制度を障害者雇用対策の核としている¹³⁵。障害者雇用率制度の実態については毎年6月1日時点の実績が厚生労働省によって調査されているが、その測定対象は障害者のうちどれだけの人が働いているかを表す障害者就業率ではなく、法定常用労働者のうちどれだけの人が障害をもつ常用労働者かを表す障害者雇用率である。障害者雇用率の算定に当たって潜在失業者は考慮されないし、重度身体障害または重度知的障害をもつ常用労働者は1人ではなく2人働いているとみなされ、短時間で働く障害者は1人ではなく0.5人働いているとみなされる¹³⁶。そのため、障害者実雇用率に関する調査結果をみても、何人の障害者が働いているのかを直ちに把握することはできない。

割当雇用中心の障害者雇用政策であることが一因となって、中央政府の「障害者基本計画」における成果目標は当事者目線ではなく行政目線で設定されている。例えば最新の「障害者基本計画（第4次）」（2018年度～2022年度）における障害者就労関係の成果指標には、政策実施段階の活動指標に該当する業績指標ばかりが並んでいる¹³⁷。そのため、障害者のうちどのくらいの人かが働いているのか、障害者就労支援策から漏れている障害者はどのくらいいるのか、障害者と非障害者との格差はどのくらいあるのかといった実態を推し測ることは難しい。その事実を、男女共同参画計画の成果指標と対比すると、より鮮明になる。「第4次男女共同参画基本計画」（2016年度～2025年度）における雇用に関する成果指標は逆に、女性のうちどのくらいの人かが働いているのか、育児休業を取得している男性はどのくらいいるのか、女性と男性の格差はどのくらいあるのかを示す、当事者目線のアウトカム指標しか設定されていない¹³⁸。

ーデン、英国、米国である。

¹³⁵ 民間企業における雇用率設定基準は、障害者雇用率＝{(身体障害者、知的障害者及び精神障害者である常用労働者の数)＋(失業している身体障害者、知的障害者及び精神障害者の数)}／(常用労働者数＋失業者数)。2019年4月現在の法定雇用率は、民間企業が2.2%、特殊法人、中央政府及び地方公共団体は2.5%、都道府県等の教育委員会は2.4%である。

¹³⁶ 短時間で働く重度身体障害者及び重度知的障害者はダブルカウントとハーフカウントの結果、1人働いているとみなされる。

¹³⁷ 例えば「就労移行支援の利用者数」、「43.5人以上規模の企業で雇用される障害者数」、「公共職業安定所における就職件数（障害者）」等の成果指標が設定されている。

¹³⁸ 例えば、「25歳から44歳までの女性の就業率」、「男性の育児休業取得率」、「本省課室長相当職に占める女性の割合」等の成果指標が設定されている。

障害者基本計画において障害者全体を母集団とした成果指標が設定されていない理由の一つは、障害の多様性を無視して障害者を一括りにして政策目標を設定することが難しいからかもしれない。これは障害者目線の理由だが、そもそも日本には障害者数を直接把握できる政府統計が存在しないため、障害者全体を母集団にして成果目標を設定したくても技術的に難しいことも強調しておかなければならないだろう。内閣府編『平成 30 年版障害者白書』に掲載されているのは三障害（身体障害者、知的障害者、精神障害者を指す。以下同じ）別の障害者数で、障害者総数のデータは掲載されていない¹³⁹。障害者総数については、2011 年と 2016 年に行われた厚生労働省「生活のしづらさなどに関する調査（全国在宅障害児・者等実態調査）」（以下「生活のしづらさ調査」という）の結果の概要が記載されているウェブページに PDF 形式の 1 枚分の参考資料へのリンクが貼られており、そこに政府推計値が公表されている¹⁴⁰。その推計方法は、三障害別の調査に基づく三障害別障害者数を単純に合計するという方法なので、三障害以外の障害者を含まず、かつ三障害について重複障害を持つ人を二重ないし三重に計上した粗い推計値でしかない。

障害者就労者数については、厚生労働省「障害者雇用実態調査」結果をみれば、常用労働者 5 人以上を雇用している民営事業所で雇用されている三障害別障害者数及び発達障害者数の推計値を産業別や職業別等に把握することができる。しかし調査頻度は 5 年ごとで最近の調査時点は 2013 年と 2018 年であり、他方で障害者数に関する最近の政府推計値は

¹³⁹ 内閣府編（2018b）を参照。障害者を定義し測定する方法は、2 通りある。1 つは自己評価による主観的方法、もう 1 つは行政管理上の定義に基づく客観的方法である（OECD [2010] p. 42）。前者では一般に、世界保健機関（World Health Organization : WHO）の国際生活機能分類（International Classification of Functioning, Disability and Health : ICF）のように障害中立的な方法で障害が定義され、「日常生活の活動を制限する長期の健康上の問題または障害の存在」に関する質問への回答が推計に用いられる。ここでいう長期とは、通常 6 カ月ないし 12 カ月以上の継続する期間を指す。先進諸国の多くはこの自己評価による主観的方法で障害者数を把握している。他方で行政管理上の定義に基づく客観的な方法では、障害年金の給付要件や障害者手帳の交付要件などで障害が定義され、その受給者数ないし手帳所持者数などで障害者数が測定される。日本の地方自治体及び中央政府による障害者数の把握方法は、後者にに基づく。日本の障害者を含む政府統計調査の全体像は、勝又（2006b）に詳しい。また、2001 年 5 月に採択された国際生活機能分類については、障害福祉研究会編（2002）に日本語訳がある。

¹⁴⁰ 2011 年調査に基づく推計値は Retrieved July 14, 2019, from https://www.mhlw.go.jp/toukei/list/dl/seikatsu_chousa_b_02.pdf。2016 年調査に基づく推計値は Retrieved July 14, 2019, from https://www.mhlw.go.jp/toukei/list/dl/seikatsu_chousa_b_h28_01.pdf。

上述の通り 2011 年と 2016 年のものだから、障害者就業率の推計には難がある。その他に障害者就労者数を把握できる調査には、厚生労働省「障害者の就業実態把握のための調査」と厚生労働省「生活のしづらさ調査」がある。前者では生産年齢人口の三障害別手帳所持者の就業者数を就業形態別等で把握することができるが、最新の調査時点は 2011 年で、それ以降 2019 年 4 月現在までに調査が一度も実施されていない。後者では在宅の障害者のうち就労している者の数を障害の種類別や年齢階級別に把握できる。施設入所者は収入のある仕事をしていないと仮定すれば全国の障害者就労者数を推計することができ、この調査を用いて障害者数の政府推計が行われているから障害者就業率の推計も可能であるが、障害者就労者数の政府推計値は公表されていないので、独自に算出する必要がある。つまり障害者就労の実態は、柔道事故の場合と同じように、総計さえもよく把握されていない状況なのである。したがって本稿で千人雇用事業のインパクトを確かめるために障害者就業率を推計することの社会的意義は大きいと考えられる。

千人雇用事業を取り上げる 2 番目の理由は、障害者就労の実態に関する透明性を高めて、ガバナンスの改善を促す必要性が高いと考えられるからである。2018 年 8 月、行政機関における障害者雇用率のデータが水増しされていたことが明らかになった¹⁴¹。障害者雇用率の水増しは統計手続きのミスや不正ではなく、本来障害者としてカウントしてはならない労働者を障害者と誤ってあるいは偽って申告した等の申告ミスや虚偽申告という問題であり、義務雇用型の障害者雇用制度をもつ日本において障害者就労支援制度の核となる障害者雇用率のデータが過大申告されていたという深刻な社会問題である。

表 2-1 の「達成機関の割合」をみると、独立行政法人等よりも市町村の機関及び市町村教育委員会の方が、市町村の機関及び市町村教育委員会よりも都道府県の機関及び都道府県教育委員会の方が、都道府県の機関及び都道府県教育委員会よりも国の機関の方が過大申告を行っていたことが分かる。ルール形成においてより強い権限を持つ上位の行政機関ほど法令遵守の規範意識が強くないということなのだろうか。その原因は明確には分からないが、もし障害者就労の実態に関する様々なデータがより再利用しやすい形態で広く公開されていれば、外部の第三者がデータの信憑性を他のデータによって確かめて公に議論することが容易になるため、データの改ざんや手続きの逸脱等を発見し是正することが

¹⁴¹ 厚生労働省ほか「国の行政機関における平成 29 年 6 月 1 日現在の障害者の任免状況の再点検結果について」（2018 年 8 月 28 日公表、Retrieved July 19, 2019, from <https://www.mhlw.go.jp/content/11601000/000347582.pdf>）。

より容易になると期待される。

(表 2 - 1) 障害者雇用率の水増し問題^(注)

	障害者雇用者数 ^(注1)		実雇用率 ^(注2)		達成機関の割合 ^(注3)		
	当初	精査後	当初	精査後	当初	精査後	差(当初-精査後)
国の機関	7,593.0人	3,711.0人	2.50%	1.17%	88.1%	16.7%	71.4%ポイント
都道府県教育委員会	12,782.0人	10,592.0人	2.23%	1.83%	78.7%	31.9%	46.8%ポイント
都道府県の機関	8,633.0人	7,951.5人	2.65%	2.36%	97.4%	68.4%	29.0%ポイント
市町村教育委員会 ^(注4)	1,862.0人	1,693.0人	2.14%	2.01%	88.0%	74.6%	13.4%ポイント
市町村の機関 ^(注4)	26,412.0人	25,643.0人	2.44%	2.29%	88.2%	76.0%	12.2%ポイント
独立行政法人等	10,276.5人	10,224.0人	2.40%	2.38%	78.3%	76.6%	1.7%ポイント

(注) 当時の「国の機関」、「都道府県の機関」、「市町村の機関」、「独立行政法人等」の法定雇用率は 2.3%、「都道府県教育委員会」及び「市町村教育委員会」の法定雇用率は 2.2%。

(注 1)「障害者雇用者数」は、障害者雇用率制度に基づいて重複障害者をダブルカウント、短時間労働の障害者をハーフカウントした数値。

(注 2)「実雇用率」は、法定雇用率適用事業所における常用雇用者に占める障害をもつ常用雇用者の実際の割合。

(注 3)「達成機関の割合」は、法定雇用率適用機関に占める法定雇用率達成機関の割合。

(注 4) 市町村教育委員会のうち、中学校、高等学校及び中等学校に置かれる教諭、助教諭及び講師の任命権者であるものの以外は、市町村の機関に含まれる。

(資料) 厚生労働省ほか「平成 30 年 8 月 28 日に公表した「国の行政機関における平成 29 年 6 月 1 日現在の障害者の任免状況の再点検結果について」及び同年 9 月 7 日に公表した「立法機関及び司法機関における平成 29 年 6 月 1 日現在の障害者の任免状況の再点検結果について」の訂正について」(2018 年 10 月 22 日公表); 厚生労働省「都道府県の機関、市町村の機関、都道府県等の教育委員会及び独立行政法人等における平成 29 年 6 月 1 日現在の障害者の任免状況等の再点検結果について」(2019 年 10 月 22 日公表)¹⁴²

千人雇用事業を取り上げる 3 つ目の理由は、地方自治体による障害者就労支援策の効果

¹⁴² 中央政府の資料は Retrieved July 19, 2019, from https://www.mhlw.go.jp/stf/newpage_01958.html. 地方自治体の資料は Retrieved July 19, 2019, from <https://www.mhlw.go.jp/content/11704000/000463282.pdf>.

を定量的に明らかにすることの学術的意義が大きいと考えられるからである。障害者就労支援策に限らず、日本では地域労働市場政策に関する定量研究が多くは行われてこなかった¹⁴³。おそらく主要労働統計の地域別表章単位が地域ブロックや都道府県のように広域である場合が多いこと及び雇用政策の分権化があまり進んでいないことが影響して、地域差を考慮したインパクト評価研究は少なく¹⁴⁴、管見によれば日本の基礎自治体による障害者就労支援策が地域の障害者就業率に与える影響を定量評価した研究は存在しない。したがって本稿で千人雇用事業のインパクト評価を行うことができれば、当該事業に関する初のインパクト評価研究になるだけでなく、日本の地方自治体による障害者就労支援策に関する貴重なインパクト評価研究になるのではないかと考えられる。

千人雇用事業を取り上げる最後の理由は、同事業は人々の潜在能力の拡大という本稿の考える自治体経営の目的要件に合致する政策で、その目的に資する達成目標を掲げた先進的な事例だからである。自治体評価では測定の容易さという技術的な理由によって成果指標が設定されてしまうことが多いが、千人雇用事業では中央政府が推奨する成果指標ではなく、地域に住む障害者やその家族にとって追求する価値のあるアウトカムが成果指標として設定された。そして目標設定後に開発された測定方法に基づいて、アウトカムの実績値が月に1回の頻度でタイムリーに公表され続けてきた。したがって千人雇用事業のインパクト評価を行うことで、政策実施前のベースラインデータや比較に適する他自治体のアウトカムデータがなくても、政策によって期待される重要なアウトカムの業績を定期的に測定し公開し続けられれば、完璧ではないかもしれないが第三者がインパクト評価を行うことができることを示したい。

2-2. 千人雇用事業の概要と主要な成果

千人雇用事業を分析対象にする理由を説明し終えたので、千人雇用事業の概要と主要な

¹⁴³ 太田（2017）を参照。

¹⁴⁴ 労働研究で定評のある学術誌の1つである労働政策研究・研修機構『日本労働研究雑誌』に2010年1月から2017年12月までに掲載された特集論文、査読付き研究論文および研究ノート408本（特別号を除く）の系統的レビューを行ったところ、標題および要旨から判断する限り、国内の労働市場の地域差や特定地域の労働市場について考察している実証研究は22本しかなかった（後藤〔2019b〕p. 30）。そのうち、地域雇用政策の効果に関する量的研究は1本もなかった。

成果について述べていく。岡山県総社市は人口約7万人ほどの都市で、いわゆる平成の大合併で1市2町が合併して2005年に現在の行政区域になった。県庁所在地である岡山市や観光名所である倉敷市のベッドタウン的な役割を果たしており、日本全体の人口が減少する中で、2019年現在に至るまで趨勢的に人口が増加している人口増加都市の一つである。市長は片岡聡一で、長く橋本龍太郎元首相の秘書を務めた後、2007年に総社市長に初当選し、2019年4月現在で3期目を務めている。千人雇用事業は片岡が2011年に再選を果たしたときの公約の一つで、障害者が就労を通して、生きがいを感じながら安心して暮らすことのできる地域社会の実現に寄与することを目的として、2011年度から2015年度までの5年間で総社市で働く障害者数を1,000人にすることが目指された¹⁴⁵。目標年度までには目標を達成できなかったが、その後も事業は継続され、2017年5月に総社市で働く障害者就労者数は目標の1,000人を超えた。その後、目標は上方修正され、現在は「障がい者千五百人雇用」が推進されている。

総社市役所の公式ホームページで公開されているデータを用いて、総社市における障害者就労者数の推移をみてみよう（図2-1参照）。最初の公表時点である2011年11月に420人であった障害者数は、2015年4月に852人へとほぼ倍増した。千人雇用事業の当初の達成目標時期は2015年度末で、2016年4月1日時点の障害者就労者数は909人であったから、あと約90人で目標達成というところまで到達した。総社市役所は当初の目標年度を過ぎても同事業を継続し、2017年5月には遂に障害者就労者数1,000人を突破した。その後、若干の増減を経験しつつ、2019年4月現在まで総社市の障害者就労者数は趨勢的に増大し続けている。

この障害者就労者数には、一般就労者だけではなく福祉的就労者数も含まれている。福祉的就労者とは、通常の事業所に雇用されることが困難なために就労継続支援事業所や地域活動支援センター等で働く障害者のことである。就労継続支援事業所には、雇用契約を締結して利用し、労働に対して原則として最低賃金以上の給料が支払われるA型事業所と、雇用契約を締結せずに利用し、授産的な活動に対する工賃が支給されるB型事業所がある。

¹⁴⁵ 千人雇用事業を立ち上げるきっかけとなったのは、岡山県立支援学校誘致の失敗であった。市有地を無償提供して誘致しようとしたがうまくいかず、倉敷市に建設することが2010年12月に決まったため、「『支援学校を卒業した後の、働く場所は総社市が担う』という考えにシフトした」のだという（総社市公式ホームページより引用。Retrieved July 14, 2019, from http://www.city.soja.okayama.jp/fukushi/shogaisha/senninkoyou/syuuga_isya_koyou_toha.html）。

全国的にみると福祉的就労者の大半は B 型事業所の利用者で¹⁴⁶、2015 年度における A 型事業所の月平均工賃は、全国平均では 71,513 円、総社市平均では 63,528 円であったのに対して、同時点における B 型事業所の月平均工賃は、全国平均では 15,033 円、総社市平均では 13,252 円に過ぎなかった¹⁴⁷。賃金は、労働者の自立を支える重要な要因の一つである。障害者就労者数が増えても、低工賃の B 型事業所で働く就労者ばかりが増えるのでは、障害者の社会的包摂が果たせたとはいえない。したがって千人雇用事業では、単に障害者就労者数を増やすのではなく、一般の労働法規が適用され、より開かれた職場で働く障害者を増やすことも目標とされていた。具体的には、障害者就労者数に占める一般就労者の割合は 5 割が目標とされ、実際には 4 割程度になるのではないかと予想されていた¹⁴⁸。図 2－1 から分かるように、障害者就労者数が増えても、一般就労者割合は 50%を下回ることなく、概ね 50%台後半で推移している。

総社市役所が総社市で働く障害者就労者数としてカウントしているのは、①総社市内の事業所において就労している障がい者、②総社市外の事業所において就労している総社市民の障がい者、③就労支援ルームを通じて就労するなど総社市役所の取組に基づき就労している障がい者である¹⁴⁹。「就労支援ルーム」とは、障害者の福祉と就労を一体的に支援するためにハローワーク総社¹⁵⁰内に設置された施設である。設置時期は総社市役所が千人雇用事業を開始した年度である 2011 年度の 7 月で、地方自治体の職員がハローワーク¹⁵¹に常駐して障害者に対する福祉支援と就労支援をワンストップで実施するという取組みが全

¹⁴⁶ 厚生労働省「社会福祉施設等調査」結果によれば、就労継続支援事業所利用実員数に占める就労継続支援 B 型事業所利用実員数は、2011 年で 89.0%、2016 年で 78.8%であった。

¹⁴⁷ 全国の数値は、厚生労働省「平成 28 年度工賃（賃金）の実績について」に基づく（Retrieved from July 14, 2019, <https://www.mhlw.go.jp/file/06-Seisakujouhou-12200000-Shakaiengokyokushougaihokenfukushibu/0000199264.pdf>）。総社市の数値は、筆者が 2017 年 3 月に総社市福祉課から提供を受けた資料に基づく。比較年度を 2015 年度にした理由は、総社市については 2015 年度以降のデータしか入手できなかったためである。

¹⁴⁸ 2017 年 7 月に総社市総合文化センターで開催された「総社市障がい者千人雇用事業 1,000 人達成記念フォーラム」における片岡市長の発言に基づく。

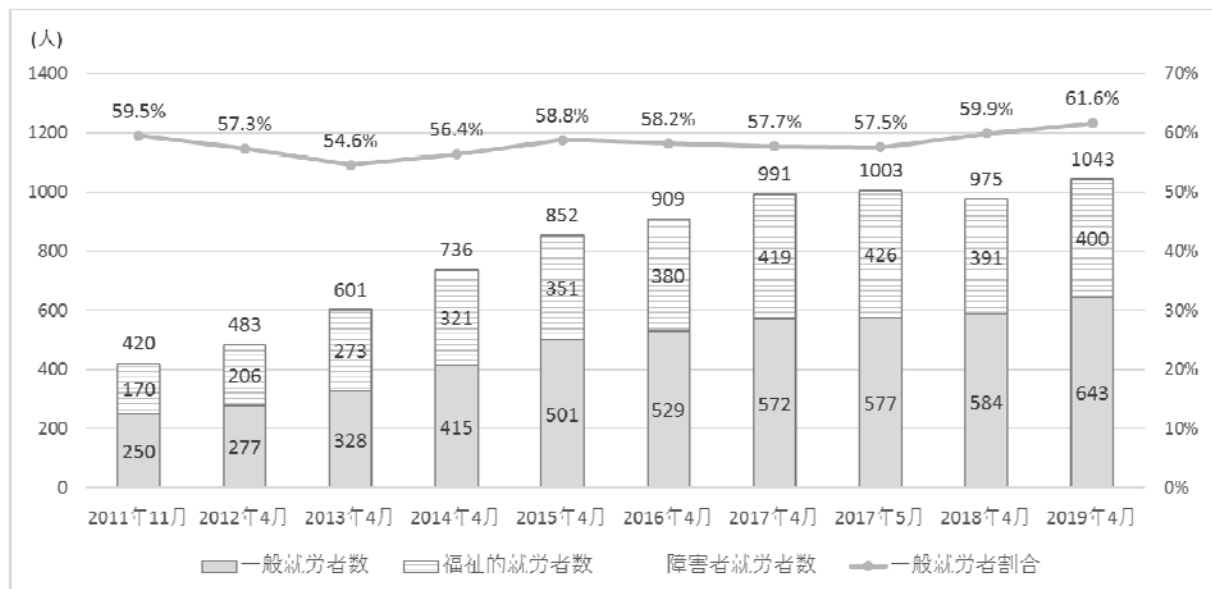
¹⁴⁹ 総社市公式ホームページより引用（Retrieved July 14, 2019, from http://www.city.soja.okayama.jp/fukushi/shogaisha/senninkoyou/syuugaisya_koyou_toha.html）。

¹⁵⁰ 正式名称は「倉敷中央公共職業安定所総社出張所」。

¹⁵¹ 正式名称は「公共職業安定所」だが、本稿では通称である「ハローワーク」を用いる。

国に先駆けて実現されたのだった¹⁵²。

(図 2 - 1) 総社市における障害者就労者数(左軸)及び一般就労者割合(右軸)の推移



(注) 調査時点は各月 1 日。

(出典) 総社市役所公式ホームページのデータに基づき筆者作成。

ここで雇用政策の分権化について、少し触れておく必要があるだろう。1999 年に成立した地方分権一括法¹⁵³の施行に伴う雇用対策法¹⁵⁴の改正によって、地方自治体にも雇用施策を実施する努力義務が課されるようになり、2003 年には職業安定法の改正で地方自治体が届出により無料紹介事業を行うことができるようになった。だがハローワークの有する求人求職情報のオンライン提供や雇用保険の認定・給付等の事務・権限の見直しはなかなか進まず、近年における雇用政策の分権化は、中央政府から地方自治体への権限移譲という

¹⁵² 障害者に対する福祉支援と就労支援をワンストップで一体的に実施するという取り組みは、埼玉県志木市が総社市よりも 10 日ほど早い 2011 年 6 月 20 日に、市役所庁舎内に一体的支援施設を設置する形態で開始していた。しかし、ハローワークに市役所職員が常駐するかたちで障害者福祉と雇用の一体的支援体制を実現したのは、総社市が全国初である(2017 年 3 月に筆者が総社市福祉課職員に対して行った集団面接聴取結果とそのときに提供を受けた資料に基づく)。

¹⁵³ 正式名称は「地方分権の推進を図るための関係法律の整備等に関する法律」。

¹⁵⁴ 正式名称は「労働施策の総合的な推進並びに労働者の雇用の安定及び職業生活の充実等に関する法律」。2018 年 7 月に「働き方改革を推進するための関係法律の整備に関する法律」によって改正された。

より、中央政府と地方自治体の連携強化が推進されてきた¹⁵⁵。中央政府と地方自治体の連携を強化する目的で2011年6月から行われている事業の一つが、一体的実施事業である。これは希望する地方自治体において、ハローワークが行う無料職業紹介等と地方自治体が行う福祉的支援の相談等業務とをワンストップで一体的に実施する事業である。2018年3月末時点で47都道府県のうち33道府県(約70.2%)、1741市区町村のうち140市区町(約8.0%)の計173自治体が一体的実施事業を行っている。障害者に対する福祉と就労の一体的実施を行っている基礎自治体は、埼玉県志木市、滋賀県湖南市、そして岡山県総社市の3団体だけである¹⁵⁶。

総社市役所が一体的実施に踏み切った理由は、障がい者の就労機会を増やすためには、無料職業紹介事業を所掌し、豊富な求人求職情報をもつハローワークとの連携が不可欠だったからである¹⁵⁷。ただし、総社市役所が千人雇用事業を開始して真っ先に立ち上げたのは、ハローワーク総社内の就労支援ルームではない。就労支援ルームの設置に先立って、2011年5月に「総社市障がい者千人雇用委員会」(以下「千人雇用委員会」という)が設置され、障害者雇用の場の創設や就労の安定化に向けた施策に、官民が協力して取り組むための議論が行われた¹⁵⁸。そこで千人雇用事業の理念や具体的な戦略が検討され、ハロー

¹⁵⁵ 最近になって、ハローワークの行う無料職業紹介業務等の地方分権化も進み始めている。中央政府が所管するハローワーク業務を試験的に都道府県へ移管する取組みとして、2012年度から埼玉県及び佐賀県で「ハローワーク特区」が実施された。その成果検証を受けて、2016年5月成立同年8月施行の第6次地方分権一括法で、地方自治体による無料職業紹介事業の届出の廃止、ハローワークの有する求人求職情報のオンライン提供の法定化等が行われた。地方版ハローワークの設置事業所数は2016年8月末時点で552か所だったが、2018年3月末時点で734か所に増大している。ハローワーク特区の成果検証については全国知事会「ハローワーク特区等の成果と課題の検証について」(2015年6月30日付, Retrieved July 14, 2019, from https://www.cao.go.jp/bunken-suishin/doc/kaigi21shiryou_09.pdf)を、地方版ハローワークの設置状況については厚生労働省職業安定室公共職業安定所運営企画室「雇用に関する国と地方公共団体の連携状況等について」(平成30年7月公表, Retrieved July 14, 2019, from <https://www.mhlw.go.jp/content/1160000/0/000349937.pdf>)を参照。

¹⁵⁶ 内閣府の「公共職業安定所(ハローワーク)の一体的実施」に関するウェブページより (Retrieved July 14, 2019, from <https://www.cao.go.jp/bunken-suishin/jimukengenijoh/hellowork.html>)。なお滋賀県湖南市は、「知的障害者福祉の父」と呼ばれる糸賀一雄らが創設した近江学園のある地である。

¹⁵⁷ 坂之上(2014)を参照。

¹⁵⁸ 現在は、「総社市障がい者千五百人雇用委員会」と名称変更されている。

ワーク総社との連携協定の締結（2011 年 5 月）、就労支援ルームの設置（2011 年 7 月）、総社商工会議所及び倉敷平成病院との、障害者雇用の推進に関する協定締結（2011 年 11 月）、障がい者千人雇用推進条例の制定（2011 年 12 月）等の体制や制度が迅速に整えられた¹⁵⁹。さらに翌年の 2012 年 4 月には総社市独自の「総社市障がい者千人雇用センター」（以下「千人雇用センター」という）が市役所隣の総社市社会福祉協議会内に設置された¹⁶⁰。同センターは、全国にある「障害者就業・生活支援センター」と同等の機関で、雇用、保健福祉、教育等の関係機関が連携して、障害者に対して就業面と生活面の一体的な相談支援を行うための支援機関である。「障害者就業・生活支援センター」は 2019 年 5 月現在、全国に 334 か所設置されている¹⁶¹。市区町村の数は 1,741 であるから、約 2 割の自治体に設置されている計算になる。総社市の場合、隣接する倉敷市に障害者就業・生活支援センターが設置されているので総社市に新設することは難しいとされたが、倉敷市の同センターは職員数に比して管轄地域が広いため、総社市独自に設立して、総社市の障害者のことをきめ細かく支援できるようにしたのだという¹⁶²。

なぜ総社市役所は、千人雇用事業を開始して約 1 年で同事業を支える核となる組織体制を整え、事業開始から 6 年ほどで、就労年齢の障害者の 8 割超が働くという目標を達成することができたのだろうか。2017 年 3 月に筆者が行った総社市役所職員への聞き取り調査によれば、以下の 3 つの要因が主な成功要因であった。

1 つ目は、片岡市長が強力なリーダーシップを発揮して、千人雇用事業の目標達成に向けて市役所職員や市議会をはじめとする関係者に強く働きかけたことである。同事業を打ち出した当初は市職員の抵抗が強く、できない理由ばかりを挙げたり、雇用はハローワークに任せればよいという意見もあったが、片岡市長が同事業に強くコミットし、機会があればいつでも真っ先に同事業のことを話し、具体的な数字を出して PR してきたので、次第に市役所内外での支援者や協力者が広がっていったのだという¹⁶³。

¹⁵⁹ 2017 年 3 月に筆者が総社市福祉課から提供を受けた資料に基づく。

¹⁶⁰ 現在は「総社市千五百人雇用センター」と名称変更されている。

¹⁶¹ 数字は厚生労働省の資料に基づく（Retrieved July 19, 2019, from [https:// www.mhlw.go.jp/content/000525099.pdf](https://www.mhlw.go.jp/content/000525099.pdf)）。

¹⁶² 2017 年 3 月に筆者が総社市福祉課職員に対して実施した集団面接聴取結果に基づく。

¹⁶³ 「できない理由ばかりを挙げる」という点については、坂之上（2014）を参照。なお、「千人雇用事業」担当課に、内閣府からの出向者で中央政府との連携調整等に長けた職員、総務課経験をもち自治体業務全体を見渡せる職員、障害福祉サービスの経験を有する職員等がバランスよく配置されたことも、関係各所への説明を重ねて理解者を増やしながらス

2つ目は、障害者雇用に関わる重要な支援機関が千人雇用事業に協力的で、関係機関が一体となって同事業を推進する体制を初期に構築できたことである。まず、障害者の福祉的就労に関する豊かな知見と経験と人脈をもつ社会福祉法人旭川荘の江草安彦名誉理事長（当時）が千人雇用委員会の会長に就任してくれたため、総社市役所や千人雇用センターが障害をもつ当事者や障害者就労支援に携わる人々との信頼関係を築き、障害者就労支援のノウハウを習得することが容易だったという。また、ハローワーク総社も千人雇用事業に協力的で、他の連携協定に先駆けて 2011 年 5 月に「福祉から就労」支援事業に関する協定を締結できたことで、福祉と雇用のギャップを埋める足場と他機関との連携によって千人雇用事業を進めるという体制を早期に確立することができた。

3つ目は、地の利である。総社市は岡山市や倉敷市のような大都市ではなく、関係機関が地理的に近接しているため、関係者と顔が見える関係になることが容易であった。例えば、ハローワーク総社は総社市役所から 650m ほどの距離にあり、歩いて 10 分もかからない。千人雇用センターと総社市社会福祉協議会が入る総社市総合福祉センターは、総社市役所に隣接している。そのため、何か問題が生じたらすぐに対面で相談することができ、相互に人材交流するのも抵抗が小さかったという。このように物理的な距離が近く人口規模があまり大きくなかったことが、障害者を含む市民、企業、障害者支援機関そして行政機関が一丸となって千人雇用事業を推進するという風土を醸成するのに役立っているのではないかとのことであった。

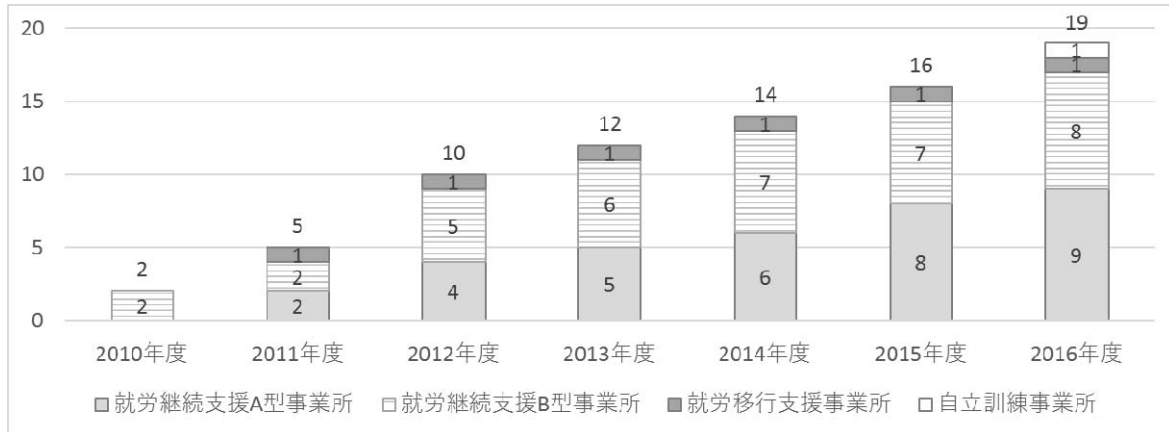
千人雇用事業の実績は、障害者就労者数の伸びだけではなく、市内の障害者福祉に関わる事業所数の伸びや、障害をもつ当事者やその家族から寄せられる声の変化、障害をもつ納税者の増加等によって把握することもできる。図 2-2 に示したのは、総社市内における障害者就労訓練系事業所数の推移である。千人雇用事業を始める前に市内にあった障害者就労訓練系事業所は就労継続支援 B 型事業所 2 カ所だけであったが、その後関連機関は増え続けてきたことが分かる。福祉的就労支援機関が増え、福祉的就労者が増えると、障害者就労に係る経費も増える。総社市では、2010 年度に 8,202 万円だった千人雇用事業関係経費が、2011 年度には 1 億 2,816 万円に、2015 年度には 3 億 6,046 万円に膨らんだ¹⁶⁴。

ピード感をもって事業を推進するために重要だったという。その職員配置にも、片岡の強い思いが表れていると考えられる。

¹⁶⁴ 単市事業である千人雇用センター委託料（2012 年度 900 万円、2015 年度 1,556 万円）を含む。単市事業以外の市の負担額は 4 分の 1 である（2017 年 3 月に筆者が総社市福祉課から提供を受けた資料に基づく）。

しかし市議会の反対はなく、市が一丸となって千人雇用事業を推進する体制は維持され続けているという。

(図 2 - 2) 総社市内における障害者就労系障害福祉サービス事業所数の推移



(注)「就労継続支援 A 型事業所」: 通常の事業所に雇用されることは困難だが、雇用契約に基づく就労が可能な障害者に対して訓練等を行う支援機関。「就労継続支援 B 型事業所」: 通常の事業所に雇用されることは困難であり、雇用契約に基づく就労が困難な障害者に対して訓練等を行う支援機関。「就労移行支援事業所」: 通常の事業所に雇用されることが可能と見込まれる障害者に対して、利用期間を限定して訓練等を行う支援機関。「自立訓練事業所」: 知的障害者や精神障害者に対して、利用期間を限定して、自立した日常生活を営むために必要な訓練や生活相談・助言等を行う支援機関。

(出典) 2017 年 3 月に総社市福祉課から提供を受けた資料に基づいて筆者作成。

量的な把握は難しいが貴重な成果としては、例えば市報や講演会で障害者が顔を出して、障害者がいきいきと働ける場が増えたことを PR してくれるようになったことや、庁内の他部局による政策を千人雇用事業に関係づけることで相乗効果を発揮できるようになったこと等が挙げられる¹⁶⁵。「第 2 期総社市障がい者計画」(2012 年度から 2017 年度)の総括には、障害者の市民税納税者が 2012 年度の 235 人から 2016 年度の 254 人へと 4 年間で 8.1%増加し、障害者の給与収入総額が 2012 年度の約 10 億 9,700 万円から 2016 年度の約 12 億 4,700 万円へと 4 年間で 13.7%増加したことも成果として挙げられている¹⁶⁶。

¹⁶⁵ 2017 年 3 月に筆者が総社市福祉課職員に対して実施した集団面接聴取結果に基づく。

¹⁶⁶ 総社市「第 3 期総社市障がい者計画」(2018 年度～2013 年度) p. 27 (Retrieved July

2-3. 千人雇用事業のインパクト評価方法

千人雇用事業は、当初の目標年度である 2015 年度末までには、総社市で働く障害者数を 1,000 人にするという目標を達成することはできなかったが、目標達成を諦めずに事業を継続し、2017 年度に目標を達成したわけであるから、目標を達成したという意味では成功事例とみなしてよいだろう。しかし第 1 章で述べたように、業績測定とインパクト評価は異なる。目標を達成したからといって、千人雇用事業が有効であったとみなすことはできない。では、千人雇用事業のインパクトをどのように評価すればよいのだろうか。

アウトカム指標から検討していこう。千人雇用事業で最も重視されてきた成果指標は、障害者就労者数であった。その目標は、2011 年 4 月当時における就労年齢の障害者数が約 1,200 人いたため、そのうち 1,000 人が働けるようにするという趣旨で設定されていた¹⁶⁷。この成果目標は活動目標ではなく当事者目線のアウトカム目標なので、千人雇用事業の主要アウトカム指標に適している。しかし過去の年齢階層別障害者就労者数を把握することはできなかったため、就労年齢の障害者就業率ではなく、全年齢の障害者就業率を、1 つ目のアウトカム指標に用いる。

千人雇用事業の目的は、単に障害者の就労機会を量的に増やすことではなく、障害者の社会的包摂にあった。つまり千人雇用事業は、障害者一人ひとりが地域社会を構成する一員として社会経済活動に参加し、その個性や能力を発揮できる機会が与えられなければならないという考え方に基づいて推進されていた¹⁶⁸。したがって障害者就労者に占める一般就労者の比率（以下「一般就労者割合」という）を、2 つ目のアウトカム指標として用いる。この指標は、障害者就労の質を表す指標と捉えることができるだろう。地方自治体による障害者就労支援策の成果指標としては、その他にも賃金水準、職場定着率、仕事満足度のような指標がありうるが、それらの数値は統計や機関記録で把握されておらず、新た

19, 2019, from http://www.city.soja.okayama.jp/fukushi/shogaisha/sien_keikaku/syougai_fukushi_keikaku_2.html).

¹⁶⁷ 前掲注の総社市公式ホームページに掲載されている数字に基づく。多くの自治体と同様に、総社市でも障害者手帳交付台帳登載者数によって障害者数を把握している。就労年齢の障害者数とは、身体障害者と知的障害者は 18 歳以上 65 歳未満の人数、精神障害者は 20 歳以上 60 歳未満の人数である。障害者数は障害者手帳所持者数によって把握されているため、身体障害者と知的障害者のダブルカウントがある。

¹⁶⁸ 総社市「障がい者千人雇用推進条例」の第 1 条及び第 2 条を参照。

に調査をして遡及的に把握することは非常に難しいと考えられる。したがって、障害者就業率及び一般就労者割合の2つを千人雇用事業のアウトカム指標に用いる。

次に、評価デザインについて検討する。インパクトをできるだけ歪みなく評価するためには、対照集団を設定してマッチング・デザインのように属性が同一の対照集団のある手法で比較評価することが望ましい。しかし管見によれば、行政区域における障害者就労者を定期的に正確に把握して公開している基礎自治体は、総社市以外に存在しない¹⁶⁹。対照集団のないインパクト評価手法のうち、信頼性は低いものの一般的に適用しやすい手法は二時点の単純な事前事後比較デザインであるが、千人雇用事業の場合には事業開始が2011年4月で、現在と同じ方法で障害者就労者数が把握されるようになったのは2011年11月である。事業開始前のデータは把握されておらず、2011年4月の事業開始時点で把握されていた180人という障害者就労者数からは、多くの一般就労者が漏れていたことが分かっている¹⁷⁰。そのため、単純な事前事後比較法を適用することもできない。

しかし一方で、前述したように厚生労働省「生活のしづらさ調査」を用いれば全国に在る在宅の障害者就労者数を推計することができるので、施設に入所している障害者は就労していないと仮定すれば、2011年度と2016年度時点の障害者就業率を求めることができる。よって本稿では、一般指標デザイン（generic control model）に基づく差の差法（difference-in-difference design）を用いて、千人雇用事業のインパクト評価を行う。

一般指標デザインとは、全国平均や広域平均のような自己の所属する母集団の一般指標

¹⁶⁹ 総社市についても、ハローワーク総社の管轄区域には総社市だけではなく倉敷市の旧・真備町が含まれていることと、就労支援ルームを通して就労するなど総社市役所の取組に基づき就労している障害者は居住地域や就労地域に関わらず総社市の障害者就労者にカウントされていることから、総社市民または総社市で働いている人以外の障害をもつ一般就労者が総社市の障害者就労者とみなされている。なお、障害者就労に関する一体的実施を行っている埼玉県志木市及び滋賀県湖南市は総社市と同じくらいの精度で障害者就労者数を把握している可能性があるが、公表データを得ることはできなかった。また、ハローワークの管轄区域と自治体の行政区域が一致している場合に自治体とハローワークの両方に調査をすることで障害者就業率を過去に遡って把握できるのではないかと考え、2018年度に聞き取り調査と質問票調査を試みたが、十分なデータを得ることはできなかった。

¹⁷⁰ 2011年4月時点で把握できていたのは主に総社市役所が機関記録から把握できる福祉的就労者数で、ハローワークから情報提供を受けるようになった2011年11月に初めて一般就労者数を現在と同じ精度で把握できるようになったのである（2017年3月に筆者が総社市福祉課職員に対して実施した集団面接聴取結果に基づく）。

値を比較に用いてインパクトを評価する方法である¹⁷¹。実験デザインやマッチング・デザインと比べれば内的妥当性は低いとされるが、対象地域が受けた影響とある程度同じ影響を一般指標値が受けていると想定できるため、単純な事前事後デザインに比べれば外部要因による影響を除去することができる。一方で差の差法とは、二時点間における介入群のアウトカム値の変化と、同期間における対照群のアウトカム値の変化との差分を介入効果とみなすインパクト評価法のことである¹⁷²。具体的には、2011年度の総社市のアウトカム値をA1、2016年度の総社市のアウトカム値をA2、2011年度の全国のアウトカム値をB1、2016年度の全国のアウトカム値をB2としたとき、千人雇用事業のインパクトを次式によって計算する。

$$\text{千人雇用事業のインパクト} = (A2 - A1) - (B2 - B1) \quad \dots (1)$$

この式(1)を用いて算出される値を2011年度から2016年度までの千人雇用事業のインパクトとみなすために、主に次の3つの仮定をおく¹⁷³。1つ目は、平行トレンド仮定である。これは差の差法で一般的に前提条件とされる仮定で、仮に千人雇用事業が行われなかった場合、総社市のアウトカムの変化は全国のアウトカムの変化に一致するという仮定である。2つ目は、共通ショック仮定である。これも差の差法で一般的に前提条件とされる仮定で、観察期間中に生じた介入とは無関係の予期せぬイベントが総社市と全国のアウトカムに与える影響は似通っているという仮定である。3つ目は、全国の指標値を対照群の指標値とみなしてよいという仮定である。これは一般指標デザインを用いるために必要になる仮定で、総社市のアウトカムの変化が全国のアウトカムの変化に与える影響は、無視できるほど小さいという仮定である。

次に、使用するデータとアウトカムの算出方法を説明する。総社市の障害者数は、2011年度については「第4期総社市障がい福祉計画」に記載されている2011年度末時点の三障害別障害者手帳所持者数のデータを、2016年度については「第3期総社市障がい者計画(第5期総社市障がい福祉計画・第1期総社市障がい児福祉計画)」に記載されている2016

¹⁷¹ 龍・佐々木(2000) p. 76; 佐々木(2010) p. 64を参照。

¹⁷² 山口(2016) pp. 117-122を参照。

¹⁷³ 他にも、アウトカムの算出に用いたデータについて、総社市と全国との調査期間の違いは、アウトカムの差の差に影響しないという仮定をおいている。データの制約上、この仮定の検証は難しいと考えられるため、本文での説明を省略した。

年度末時点の三障害別障害者手帳所持者数のデータを用いる。総社市の障害者就労者数は、障害者数のデータと時点を合わせるため、2011年度分は2012年4月1日時点のデータを、2016年度分は2017年4月1日時点のデータを、図2-1と同様に総社市公式ホームページから得る。

全国の障害者就労者数は、一般就労者数については厚生労働省「生活のしづらさ調査」結果を、福祉的就労者数については厚生労働省「社会福祉施設等調査」結果を用いる。2011年の「生活のしづらさ調査」問21及び2016年の同調査問25は、複数の選択肢から複数回答で選ぶ形式で在宅の障害者等に対して日中の過ごし方を尋ねている。両調査とも、選択肢のうち4つは収入のある就労につながる活動を含み、うち3つは「正社員として働いている」、「正職員以外（アルバイト、パート、契約職員、派遣職員、日雇い等）として働いている」、「自営業をしている（家の仕事を手伝っている）」という一般就労に関する選択肢である。よって、これら3つの選択肢を選んだ障害者手帳所持者を一般就労者とみなす。これら3つの選択肢を重複して選択した回答者がいるかもしれないが、その事実を知る術はないため、単純集計によって一般就労者数を把握する。

日中の過ごし方について尋ねている設問における、収入のある就労に繋がる活動を含む4つの選択肢のうち、残り1つは「障害者のための通所サービスを利用している」という福祉的就労を含む選択肢である。2016年の「生活のしづらさ調査」結果では障害者のための通所サービスのうち福祉的就労をしているのか、療養介護や生活介護等の福祉的就労以外のサービスを利用しているのかを把握できるが、2011年の「生活のしづらさ調査」結果では、障害者のための通所サービス利用者のうち福祉的就労者のみを取り出して把握することができない。他方で福祉的就労者数については、厚生労働省「社会福祉施設等調査」結果によって、各年9月における障害福祉サービス利用実員数を把握することができる。したがって、福祉的就労者数は「社会福祉施設等調査」結果を用いて把握する。総社市の場合には、地域活動支援センターⅢ型利用者、就労移行支援事業利用者、就労継続支援A型事業利用者及び就労継続支援B型事業利用者を、福祉的就労者とみなしているという¹⁷⁴。「社会福祉施設等調査」で把握できる就労移行支援事業利用者、就労継続支援A型事業利

¹⁷⁴ 2020年1月に筆者が総社市福祉課職員に対して実施した集団面接聴取に基づく。地域活動支援センターとは、働くことが困難な障害者の日中の活動をサポートする通所型の福祉施設で、一般に、創作的活動、生産活動、社会との交流の促進等の事業を実施している。総社市では、授産的作業を行っているⅢ型利用者を福祉的就労者に含めているとのことであった。

用者及び就労継続支援 B 型事業利用者を、全国の福祉的就労者とみなす。

「社会福祉施設等調査」は全数調査なので福祉的就労者数の実数を直接把握できるが、「生活のしづらさ調査」は標本調査なので、一般就労者数は推計値を出す必要がある。まず 2011 年の「生活のしづらさ調査」問 21 及び 2016 年の同調査問 25 に回答した障害者手帳所持者数のうち、「正社員として働いている」、「正職員以外として働いている」または「自営業をしている（家の仕事を手伝っている）」と回答した人の占める割合とその 95%信頼区間を求める。これを仮に、『生活のしづらさ調査』に基づく障害者一般就業率」とその 95%信頼区間と呼ぶ。障害者手帳所持者数については「生活のしづらさ調査」結果の報告書に推計値が掲載されているのみで集計値が記載されていないため、95%信頼区間を求めることはできない。よって障害者手帳所持者数は報告書に掲載されている数値をそのまま用いて、それに『生活のしづらさ調査』に基づく障害者一般就業率」を乗じて、障害者一般就労者数の推計値を求める。

全国の障害者数については、前述したように厚生労働省「生活のしづらさ調査」に基づく政府推計値が公表されている¹⁷⁵。政府推計によれば、2011 年調査に基づく推計障害者数は 787.9 万人で、人口の約 6.2%であった。また 2016 年調査に基づく推計障害者数は 936.6 万人で、人口の約 7.4%であった。しかしこの政府推計値は総社市を含む一般的な地方自治体が用いている障害者数とは算出方法が異なるため、地方自治体との比較には適さない。具体的には、総社市を含む一般的な地方自治体は基本的に、障害者数を三障害の障害者手帳所持者数によって把握している。他方で政府は、在宅の身体障害者数及び知的障害者数は個人向けの標本調査に基づいて障害者手帳所持者数によって把握し、施設入所の身体障害者数及び知的障害者数は事業所調査によって把握し、精神障害者数は厚生労働省「患者調査」という病院調査によって把握している。特に精神障害者数の把握方法の違いの影響は大きい。「患者調査」では調査日に層化無作為法で抽出された医療施設に通院等している者が障害者とみなされるが、精神科医療機関に通院しているからといって必ずしも日常生活や社会生活に長期にわたって相当の制限があるとは限らないから、手帳所持によって定義されている在宅の身体障害者及び知的障害者と比べれば、在宅の精神障害者数は多めに見積もられている可能性がある。他方で精神障害者保健福祉手帳を所持することで利用で

¹⁷⁵ 2011 年調査に基づく推計値は Retrieved July 14, 2019, from https://www.mhlw.go.jp/toukei/list/dl/seikatsu_chousa_b_02.pdf。2016 年調査に基づく推計値は Retrieved July 14, 2019, from https://www.mhlw.go.jp/toukei/list/dl/seikatsu_chousa_b_h28_01.pdf。

きる福祉サービスは相対的に貧弱であるし、精神障害を持つ者は手帳非所持でも条件を満たせば障害福祉サービスを受けられるため、手帳所持の誘因が相対的に小さい。同手帳の取得については本人及び家族が抵抗を示すケースが少なくないともいわれる。したがって障害者手帳制度に基づく認定障害者のみを障害者とみなすことも、妥当とはいえないと思われる。しかし、「患者調査」の表章単位は病院の入院は二次医療圏、それ以外は都道府県であるため、政府推計と同じ方法で総社市の障害者数を推計することは難しい。

政府推計値以外に用いることのできる障害者数に関する主なデータは、2種類ある¹⁷⁶。1つ目は厚生労働省「福祉行政報告例」及び同「衛生行政報告例」（以下、まとめて「報告例」という）で把握される障害者手帳交付台帳登載数である。身体障害者手帳交付台帳登載数及び療育手帳交付台帳登載数は「福祉行政報告例」に、精神障害者保健福祉手帳交付台帳登載数は「衛生行政報告例」に掲載されている。地方自治体による障害者数の把握は基本的には障害者手帳交付台帳登載数によってなされているし、「報告例」は全数調査のため選択バイアスがないから、「報告例」に基づくデータは地方自治体との比較に適すると考えられる。しかし、2011年の厚生労働省「生活のしづらさ調査」の調査結果は東日本大震災等の影響等で岩手県、宮城県、福島県、仙台市、盛岡市、郡山市、いわき市及び大阪市を除いた数値であるが、「福祉行政報告例」に基づく療育手帳交付台帳登載数及び「衛生行政報告例」に基づく精神障害者保健福祉手帳交付台帳登載数には都道府県及び指定都市のデータしかないため、盛岡市、郡山市及びいわき市の障害者数を把握して除外することができない。一方で厚生労働省「生活のしづらさ調査」は標本調査のため推計値の標本誤差や欠落バイアスがありうるが、障害者就労者数の推計に用いることのできる近年の政府調査は同調査しかないため、障害者数も同じ調査結果を用いることに合理性がある。したがって本稿では、「生活のしづらさ調査」結果を用いて障害者数を把握する。「生活のしづらさ調査」の調査結果には、全国における在宅の三障害別障害者手帳所持者数の推計値が示されている。それに同調査結果を用いた政府推計で用いられている三障害別施設入所者数を加えて合計すれば、総社市の障害者数の把握方法と類似の方法で全国の障害者数を求めることができる。以上の方法で総社市及び全国の障害者就労者数及び障害者数を算出し、さらにそれぞれの障害者就業率及び一般就労者割合を推計して比較し、千人雇用事業のインパクト評価を行う。

¹⁷⁶ 後藤（2019a）を参照。

2-4. 千人雇用事業のインパクト評価結果

総社市及び全国の障害者就業率及び一般就労者割合を求めた結果を、表 2-2 に示した。

(表 2-2) 障害者就業率及び一般就労者割合に関する総社市と全国の比較

		障害者就労者数			障害者数 (D) 注	障害者就業率 (A/D)	一般就労者 割合 (B/A)
		総計 (A=B+C)	一般就労者数 (B)	福祉的就労 者数(C)			
総社市	2011年度	483人	277人	206人	3,190人	15.1%	57.3%
	2016年度	991人	572人	419人	3,014人	32.9%	57.7%
全国	2011年度	844,777人	717,020人	127,757人	5,568,100人	15.2%	84.9%
	(95%信頼区間)	(810,843人- 878,711人)	(683,086人- 750,954人)	-	-	(14.6%- 15.8%)	(84.2%- 85.5%)
	2016年度	1,758,414人	1,406,686人	351,728人	6,596,000人	26.7%	80.0%
	(95%信頼区間)	(1,687,055人- 1,829,775人)	(1,335,327人- 1,478,047人)	-	-	(25.6%- 27.7%)	(79.2%- 80.8%)

(注) 障害者数は三障害別障害者数の単純合計値のため、重複障害者を二重ないし三重にカウントした粗い推計値である。

(資料) 総社市役所「第 4 期総社市障がい福祉計画」、同「第 3 期総社市障がい者計画」、同公式ホームページにおける「障がい者の就労者数データ」¹⁷⁷、厚生労働省「平成 23 年生活のしづらさなどに関する調査（全国在宅障害児・者等実態調査）結果」、同「平成 28 年生活のしづらさなどに関する調査（全国在宅障害児・者等実態調査）結果」、同「社会福祉施設等調査」、障害者数の政府推計値が掲載されている資料¹⁷⁸。

2011 年度時点においては、総社市の障害者就業率は 15.1%、全国の同指標値は 15.2% で、ほぼ同じ水準だった。しかし 2016 年度には、総社市の障害者就業率は 32.9%、全国の同指標値は 26.7%と差が開いた。式(1)を用いて千人雇用事業のインパクトを求めると、千人雇用事業は、2011 年度から 2016 年度の期間に、総社市の障害者就業率を 6.3%

¹⁷⁷ Retrieved July 19, 2019, from http://www.city.soja.okayama.jp/fukushi/shogaisha/senninkoyou/syougaisya_syurousyasuu.html.

¹⁷⁸ 2011 年調査に基づく推計値は Retrieved July 14, 2019, from https://www.mhlw.go.jp/toukei/list/dl/seikatsu_chousa_b_02.pdf。2016 年調査に基づく推計値は Retrieved July 14, 2019, from https://www.mhlw.go.jp/toukei/list/dl/seikatsu_chousa_b_h28_01.pdf。

ポイント高める効果を有していた¹⁷⁹。2016 年度における総社市の障害者就労者 991 人のうち、約 190 人分の就労機会は千人雇用事業によってもたらされたことになる。

次に、障害者就労の質の代理指標である、一般就労者割合をみる。2011 年度の総社市の同割合は 57.3%、全国の同割合は 84.9%で、一般就労者割合は総社市よりも全国の方がかなり高かった。その傾向は 2016 年度にも変わっていないが、総社市の場合には 2011 年から 2016 年にかけて 0.4%ポイントと僅かながら一般就労者割合が上昇したのに対して、全国の同割合は同期間に 4.9%ポイント低下した。千人雇用事業のインパクトを求めると、千人雇用事業は、2011 年度から 2016 年度の期間に、総社市の一般就労者割合を 5.3%ポイント高める効果を有していた¹⁸⁰。

以上みてきたように、一般指標デザインに基づく差の差法を用いて 2011 年度から 2016 年度の期間における千人雇用事業のインパクトを評価した結果、千人雇用事業は、障害者就業率及び一般就労者割合というアウトカムに対して正のインパクトをもつ有効な政策であったという結果を得た。そのインパクトの大きさをどのように捉えるかは議論あるところであろうが、多くの地方自治体が積極的に行っているとはいえない障害者就労支援策に意欲的に取り組むことによって、5 年間に約 190 人もの総社市の障害者が働いて社会参加できるようになり、全国の傾向と比べて福祉的就労から一般就労への移行もより進展したという成果は、貴重な成果ではないかと思われる。

2-5. 千人雇用事業のインパクト評価結果の頑健性評価

では、これらの変化の全てが、千人雇用事業の効果だといえるのだろうか。データの制約上、千人雇用事業のインパクトを正確に推定することは難しいが、次に述べる観点から結果の頑健性評価を試みる。まず、一般指標デザインに基づく差の差法を適用するに当たって前提とした、1) 平行トレンド仮定、2) 共通ショック仮定、3) 全国の指標値を対照集団の指標値とみなしうるという仮定について、妥当性を検証する。続いて、評価に用いたデータに歪みはないかという観点から、4) 総社市のアウトカム値と、5) 全国のアウトカム推計値について、バイアスの方向性や程度を検討する。

¹⁷⁹ 95%信頼区間は、4.6%ポイントから 8.0%ポイントであった。

¹⁸⁰ 95%信頼区間は、3.9%ポイントから 6.7%ポイントであった。

1) 平行トレンド仮定の妥当性検証

平行トレンド仮定は、観察期間中のアウトカムの経時変化は似通っていたという、差の差法で一般におかれる前提条件であった。一般指標デザインでは、景気変動や突発的事態のような外部要因による影響をある程度取り除くことができるとされるが、仮に景気変動の労働市場への影響等に地域差があり、地域の労働市場全体の就業率の変化がその地域の障害者就業率の変化と相関するなら、労働市場全体の変化トレンドの違いが障害者就業率の変化の差の一部を説明している可能性がある。

その可能性について、労働市場全体の就業率の変化トレンドを総社市と全国で比較することによって検討してみよう。総社市の就業率のデータを得られるのは、総務省統計局「国勢調査」結果である。よって、2010年と2015年の同調査結果を用いて、就業率の水準と変化を総社市と全国で比較する。もし2010年における総社市の就業率が全国とほぼ同じ水準で、2015年にかけて総社市の就業率の伸びが全国の伸びを上回っていれば、2011年度から2016年度における総社市の障害者就業率の伸びの一部は、千人雇用事業の効果ではなく、総社市全体の就業率の伸びが全国と比べて高かったために生じた可能性がある。

表2-3に、2010年と2015年における総社市及び全国の就業率を示した。2010年の就業率は、総社市が47.3%、全国が46.6%で、ほぼ同じ水準であった。2015年には総社市の就業率が2010年に比べて微増したのに対して、全国の実業率は同じ期間に微減した。就業率の変化の方向は総社市と全国とは逆で、総社市の就業率の伸びが全国の実業率の伸びを上回っていた。したがって平行トレンド仮定を前提とした本稿のインパクト評価結果は、千人雇用事業のインパクトを過大評価している可能性がある。ただし同期間における就業率の差は、0.6%ポイントと僅かであった。よって、総社市と全国における労働市場全体の就業率の変化トレンドの違いが本稿のインパクト評価結果に与える影響は、あったとしてもあまり大きくないと考えてよいだろう。

(表2-3) 就業率の水準と変化に関する総社市と全国の比較

		就業者数(A)	人口(B)	就業率(A/B)	変化幅
総社市	2010年	31,317人	66,201人	47.3%	0.4%ポイント
	2015年	31,878人	66,855人	47.7%	
全国	2010年	59,611,311人	128,057,352人	46.6%	-0.2%ポイント
	2015年	58,919,036人	127,094,745人	46.4%	

(出典) 総務省統計局「国勢調査」結果に基づき筆者作成。

2) 共通ショック仮定の妥当性検証

共通ショック仮定は、観察期間中における介入と無関係の予期せぬイベントがアウトカムに及ぼす影響は、総社市と全国で似通っていたという仮定であった。観察期間中に生じた大きな共通ショックは、2013 年度の法定雇用率引上げであろう¹⁸¹。その影響は、総社市と全国で似通っていたとみなしうるのだろうか。

総社市は人口約 7 万人の小都市で、障害者雇用率制度の適用事業所が多くない。総社市内にある法定雇用率適用事業所の全てが法定雇用率を達成したと仮定しても、2017 年度時点で 100 人強の一般就労機会にしかないという¹⁸²。ちょうど 100 人の障害者が 2016 年度に法定雇用率制度を通じて総社市で一般就労していたと仮定すると、障害者雇用率制度を通じた一般就労者は、2016 年度における総社市の障害者の約 3.3%、一般就労者の約 17.5%に相当する。他方で厚生労働省「平成 28 年障害者雇用状況の集計結果」によれば、全国にある民間の法定雇用率適用事業所の全てが法定雇用率をちょうど達成した場合、2016 年 6 月 1 日時点で 493,004 人の障害者が法定雇用率適用事業所で一般就労していたはずであった。それは本稿の推計値に基づく、2016 年度における全国の障害者の約 7.5%、一般就労者の約 35.0%に相当する。全国にある民間の法定雇用率適用事業所における現実の雇用実人員は 386,606 人だったので、2016 年度における全国の障害者の約 5.9%、一般就労者の約 27.5%が、障害者雇用率制度を通じて一般就労していたことになる。障害者のうち障害者雇用率制度を通じて働く一般就労者の割合は総社市よりも全国の方が大きいと考えられるから、2013 年度における法定雇用率の引上げが障害者就業率の伸び及び一般就労者割合の伸びに与えた影響は、総社市よりも全国の方が大きく、その分だけ千人雇用事業のインパクトは過小評価されている可能性がある。ただし、2013 年度に行われた法定雇用率 0.2%ポイントの引上げが総社市と全国の障害者就業率及び一般就労者割合の変化ト

¹⁸¹ 民間の法定雇用率適用事業所の場合、法定雇用率は 1.8%から 2.0%に引き上げられた。

¹⁸² 2017 年 7 月 9 日に総社市総合文化センターで開催された「総社市障がい者千人雇用事業 1,000 人達成記念フォーラム」における片岡市長の発言に基づく。全国的にみると、法定雇用率適用事業所のうち、法定雇用率を達成している事業所は 2016 年 6 月 1 日現在で 48.8%と過半数に満たず、法定雇用率適用民間事業所における実雇用率は 1.92%と、民間事業所の法定雇用率 2.0%に届かなかった（厚生労働省「平成 28 年障害者雇用状況の集計結果」, Retrieved July 19, 2019, from <https://www.mhlw.go.jp/file/04-Houdou-11704000-Shokugyouanteikyokukoureishougai koyoutaisakubu-shougaisakoyoutaisakuka/0000146180.pdf>).

レンドに与えた影響は、さほど大きくないと推察される¹⁸³。観察期間中に法定雇用率の引上げ以外の大きな共通ショックはなかったと思われるので、共通ショック仮定は概ね妥当だったと考えてよいであろう¹⁸⁴。

3) 全国の指標値を対照集団の指標値とみなしうという仮定の妥当性検証

一般指標デザインに基づく差の差法は、総社市を介入群、全国を対照群とみなして、それぞれのアウトカム指標値の差の差を介入効果とみなす研究デザインである。それが妥当性を持つためには、総社市のアウトカムの変化が全国のアウトカムの変化に与える影響は無視できるほど小さいという仮定が成立している必要があると考えられる。

表2-2のデータを用いて、2011年度及び2016年度における総社市の障害者就労者数が全国の障害者就労者数に占める割合と、総社市の障害者数が全国の障害者数に占める割合を計算してみると、それらは共に0.1%にも届かないほど小さい。そのため、2016年度における総社市の全障害者が一般就労者であったと仮定しても、全国の障害者就業率及び一般就労者割合は殆ど変化しない。総社市のアウトカム値の変化が全国のアウトカム値の変化に与える影響は無視できるほど小さいという仮定は、成立していたと考えてよいだろう。

¹⁸³ 2011年度から2016年度にかけて、法定雇用率適用事業所の常用労働者数に変化がなく、雇用実人員でみた総社市及び全国の障害者雇用率は2011年度に1.8%で、2013年度の法定雇用率引上げにより、2016年度には2.0%に上昇したとする。また、2016年度に総社市では10人、全国では386,606人の障害をもつ一般就労者が法定雇用率適用事業所で働いていたとする。このとき、2013年度の法定雇用率引上げによって増えた障害者一般就労者数は、総社市が10人、全国が約38,661人になる。したがって2013年度に法定雇用率引上げが実施されなかった場合、2016年度の障害者就業率は、総社市では約32.9%から約32.5%に、全国では約26.7%から約26.1%に微減する。その結果、障害者就業率の差の差は、6.3%ポイントから6.5%ポイントに増大する。一般就労者割合は、総社市でも全国でもそれぞれ表2-2に示した推計値より0.4%ポイント減少し、差の差は変わらない。

¹⁸⁴ なお、2011年3月に起きた東日本大震災は、観察期間中ではなく観察期間前の事象である。その総社市及び全国のアウトカムの変化トレンドへの影響は、共通トレンド仮定の妥当性検証によって間接的に検証されたと考えることができるだろう。

4) 総社市のアウトカムのバイアスについて

総社市の障害者就労者数及び障害者数のデータは、本稿で推計した値ではなく行政の機関記録に基づく。障害者数は重複障害をもつ者についてダブルカウントしているため実際よりも多く見積もられているが、全国の障害者数も同様の方法で算出したため、差の差法に基づくインパクトを大きく歪めるような影響はないと考えてよいだろう。障害者就労者数については、ハローワーク総社との連携によって初めて正確な把握が可能になったとのことであった。したがって、ハローワーク総社や千人雇用センターによる支援を受けずに働いている障害者が無視できないほど多くいた場合、総社市の障害者就業率及び一般就労者割合は実際よりも低く見積もられていることになる。しかしデータの制約のため、総社市役所が把握し切れていない障害者数が2011年度から2016年度にどのように変化したのかを把握し、その影響を推し量ることは、少なくとも現時点では難しい。

5) 全国のアウトカム推計値のバイアスについて

最後に、全国のアウトカム値の推計に用いたデータの歪みについて検討する。全国の障害者就労者数及び在宅の障害者数は標本調査に基づく推計値なので、バイアスのリスクが大きい。表2-2には全国のアウトカム値について95%信頼区間も示しており、区間推定値を用いてインパクトを評価しても、千人雇用事業は障害者就業率及び一般就労者割合に対して正のインパクトを有していたという結果になる。よって、ここで検討すべきバイアスは選択バイアスや測定バイアスであろう。

まず、福祉的就労者とみなした対象の違いの影響について検討する。本稿は、総社市が福祉的就労者に含めている地域活動支援センターⅢ型利用者を、比較可能なデータを入手できないという理由で全国の福祉的就労者には含めなかった。そのことは本稿における千人雇用事業のインパクト評価に、どのような影響を与えた可能性があるだろうか。2016年「生きづらさ調査」における日中の過ごし方を尋ねた問25の選択肢には「地域活動支援センター」がある。地域活動支援センター利用者の全てが工賃を得ている福祉的就労者ではないが、「地域活動支援センター」を選んだ障害者手帳所持者を福祉的就労者とみなして、2011年度と2016年度の福祉的就労者数の全国推計値を求めてみる。そのために、2016年「生きづらさ調査」において「地域活動支援センター」を選択した障害者手帳所持者に対する、「就労移行支援」、「就労継続支援A型」または「就労継続支援B型」を選択した

障害者手帳所持者の比率が、本稿で福祉的就労者数の算出に用いた 2011 年度と 2016 年度の「社会施設等調査」でも同一であったと仮定する。そうすると福祉的就労者数の全国推計値は、2011 年度は 138,450 人に、2016 年度は 381,168 人に増える。それに伴い、障害者就業率の全国推計値は、2011 年度は 15.4%に、2016 年度は 27.1%に上方修正される。他方で一般就労者割合の全国推計値は、2011 年度は 83.8%に、2016 年度は 78.7%に下方修正される。その結果、千人雇用事業の障害者就業率に対するインパクトは 6.3%ポイントから 6.1%ポイントに下方修正され、千人雇用事業の一般就労者割合に対するインパクトは 5.3%ポイントから 5.5%ポイントに上方修正される。ここで措いた仮定の妥当性は不確かであるものの、福祉的就労者とみなした対象の違いが千人雇用事業のインパクトに与える影響が大きいということは確認できたと考えてよいだろう。

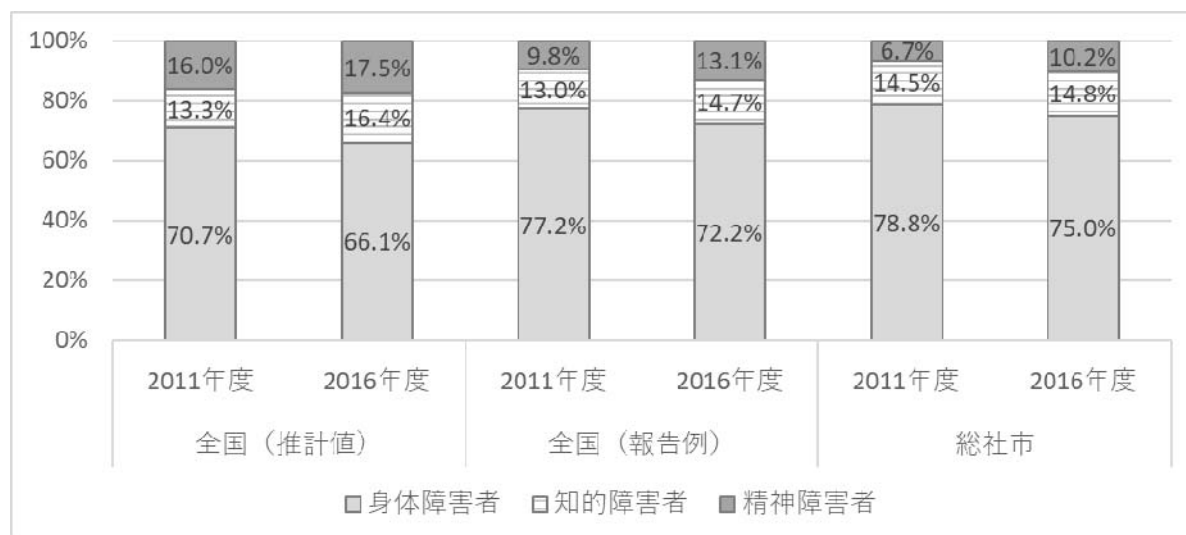
次に、標本調査の選択バイアスと測定バイアスの影響について検討する。全国のアウトカム値の推計に用いた「生きづらさ調査」は、原則自記式の質問票調査であった。したがって回収率を比べると、非就労者よりも就労者の方が、福祉的就労者よりも一般就労者の方が高かった可能性があるのではないかとと思われる。また、一般就労者数の推計に当たっては、日中の過ごし方に対する複数回答形式の設問への回答を単一回答形式の設問への回答とみなして推計しているため、一般就労者数を多く見積もっている可能性がある。なぜなら、例えば 2011 年度の同調査で「障害者のための通所サービスを利用している（地域活動支援センターや作業所等を含む）」を選択した人は福祉的就労者だと考えられるが、もしその人が工賃を得ているという理由で「正職員以外（アルバイト、パート、契約職員、派遣職員、日雇い等）として働いている」という選択肢も選んでいた場合、本稿ではその人を一般就労者とみなしているからである。

では、全国推計に用いた標本調査の選択バイアス及び測定バイアスによって、本稿における全国の障害者就業率及び一般就労者割合は、本当に過大推計になっているのだろうか。仮に障害者就業率及び一般就労者割合を過大に見積もっているなら、本稿の全国推計に基づく三障害別構成比は、実際の三障害別構成比と比べて、障害者就業率及び一般就労者割合が相対的に高い障害種別の構成比が相対的に大きいのではないかと考えられる。前述したように、三障害別障害者手帳所持者数は厚生労働省「報告例」によって把握することができる。また、2016 年「生きづらさ調査」結果を用いれば、在宅の障害者手帳所持者の三障害別障害者就業率及び一般就労者割合を求めることができる。それらのデータを用いて、本稿の全国推計では障害者就業率及び一般就労者割合が相対的に高い障害種別の構成比が

本当に大きくなっているのかを確認してみよう。

図 2－3 に、全国推計、「報告例」及び総社市の障害者に占める三障害別構成比を示した¹⁸⁵。標本調査に基づく全国推計値は、機関記録に基づく「報告例」や総社市と比べると、身体障害者の割合がやや小さく、精神障害者の割合がやや大きい。

(図 2－3) 障害者に占める三障害別構成比の比較



(注) 表 2－2 と同じく、障害者数の総計には三障害別障害者数の単純合計値を用いた。

(資料) 厚生労働省「平成 23 年生活のしづらさなどに関する調査 (全国在宅障害児・者等実態調査) 結果」、同「平成 28 年生活のしづらさなどに関する調査 (全国在宅障害児・者等実態調査) 結果」、同「福祉行政報告例」、同「衛生行政報告例」、障害者数の政府推計値が掲載されている資料¹⁸⁶、総社市役所「第 4 期総社市障がい福祉計画」、同「第 3 期総社市障がい者計画」

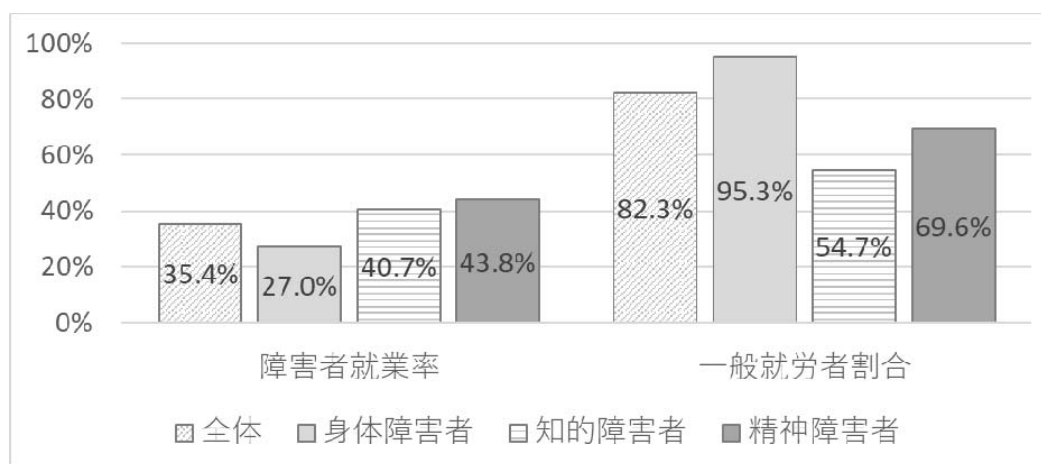
では、身体障害者の割合がやや小さく、精神障害者の割合がやや大きいというデータのバイアスは、全国の障害者就業率及び一般就労者割合の水準に、どのような影響を及ぼすのだろうか。2016 年「生きづらさ調査」問 25 への回答を用いて、在宅の障害者について

¹⁸⁵ 前述したように、「報告例」では指定都市以外の基礎自治体単位のデータが公表されていないため、2011 年のデータには、2011 年「生きづらさ調査」では含まれていない盛岡市、郡山市及びいわき市の障害者を含めたデータを用いている。

¹⁸⁶ 2011 年調査に基づく推計値は Retrieved July 14, 2019, from https://www.mhlw.go.jp/toukei/list/dl/seikatsu_chousa_b_02.pdf。2016 年調査に基づく推計値は Retrieved July 14, 2019, from https://www.mhlw.go.jp/toukei/list/dl/seikatsu_chousa_b_h28_01.pdf。

三障害別の障害者就業率及び一般就労者割合を計算した結果を図 2－4 に示した。障害者就業率は精神障害者、知的障害者、身体障害者の順に高く、一般就労者割合は身体障害者、精神障害者、知的障害者の順に高かった。そのため、全国推計値にみられる「身体障害者の割合がやや小さく、精神障害者の割合がやや大きい」というデータの歪みによって、全国推計の障害者就業率は実際よりもやや高く、一般就労者割合は実際よりもやや低く見積もられているのではないかと推察される。

(図 2－4) 2016 年度における全国の在宅の三障害別障害者就業率及び一般就労者割合



(注) 福祉的就労者数のデータに、「社会福祉施設等調査」結果ではなく「生きづらさ調査」結果を用いているため、全体の障害者就業率及び一般就労者割合は表 2－2 の値と異なる。

(資料) 厚生労働省「平成 28 年生活のしづらさなどに関する調査（全国在宅障害児・者等実態調査）結果」、障害者数の政府推計値が掲載されている資料¹⁸⁷

予想は、標本調査の選択バイアスと測定バイアスによって、全国の障害者就業率及び一般就労者割合は共に過大評価になっているのではないかというものであった。データに基づく検討結果からは、障害者就業率は確かに過大評価である可能性があるものの、一般就労者割合はむしろ過小評価ではないかという示唆を得た。

最後に、図 2－3 に示した全国推計の三障害別構成比の「報告例」と比べた歪みが、本稿における千人雇用事業のインパクト評価にどのような影響を与えるのかを検討する。全国推計と「報告例」の三障害別構成比について、2011 年度から 2016 年度にかけての変化を比べると、変化の方向はいずれも同じで、かつ、身体障害者割合は概ね同じくらい低下

¹⁸⁷ 前掲注と同じ。

している。他方で、知的障害者と精神障害者については変化の大きさが若干異なり、全国推計の方が「報告例」と比べて、精神障害者割合の変化がやや小さく、知的障害者割合の変化がやや大きい。したがって、2011年度から2016年度にかけて、もし精神障害者の方が知的障害者と比べて障害者就業率の伸び及び一般就労者割合の伸びが大きければ、本稿における全国のアウトカム推計値は過小評価だった可能性が高いと考えられる。

では、2011年度から2016年度にかけて、三障害別障害者就業率及び一般就労者割合はどのように変化しただろうか。2011年「生きづらさ調査」では福祉的就労者数を把握できないため、在宅の障害者手帳所持者の三障害別一般就業率を2011年度と2016年度で比較することによって、その答えに接近してみよう。2011年度における在宅の三障害別障害者一般就業率を「生きづらさ調査」問21への回答を用いて計算すると、身体障害者は15.3%、知的障害者は12.3%、精神障害者は13.1%であった。2016年度における同指標値を同調査問25への回答を用いて計算すると、身体障害者は25.4%、知的障害者は21.5%、精神障害者は28.7%であった。2011年度から2016年度にかけての変化幅を計算すると、身体障害者は10.1%ポイント、知的障害者は9.2%ポイント、精神障害者は15.6%ポイントになる。全国推計の方が「報告例」と比べて、精神障害者割合の変化がやや小さく、知的障害者割合の変化がやや大きかったのであるから、全国推計の障害者就業率の変化幅は、実際よりも過小評価されていた可能性がある。すなわち、全国推計における三障害別構成比の歪みによって、本稿における千人雇用事業が障害者就業率に与えたインパクトは過大評価になっていた可能性がある。ただし、障害者の多くは身体障害者であるし、三障害別構成比の変化の差は大きくないので、全国推計における三障害別障害者構成比の歪みが本稿のインパクト評価結果に与えた影響は小さいと推察される。

6) 頑健性評価結果のまとめ

頑健性評価の結果をまとめよう。第一に、平行トレンド仮定については、総社市と全国における労働市場全体のトレンドを比較して妥当性を検証した。その結果、千人雇用事業のインパクトを過大評価した可能性が示唆されたが、その影響は大きくないと推察された。第二に、共通ショック仮定については、2013年度における法定雇用率の引上げの影響を試算して妥当性を検証した。その結果、千人雇用事業のインパクトを過小評価した可能性が示唆されたが、その影響はやはり大きくないことが分かった。第三に、全国の指標値を対

照集団とみなしうるという仮定について、総社市の障害者就労者数の変化が全国のアウトカム値に与える影響を試算することによって検証した。その結果、この仮定は成立していたとみなしてよいことが明らかになった。第四に、総社市のアウトカム値のバイアスについては、データ入手方法について論理的に検討した。データの漏れのために千人雇用事業のインパクトを過小評価した可能性が示唆されたが、データの制約上、その影響を推し測ることはできなかった。第五に、全国のアウトカム推計値のバイアスについては、福祉的就労者とみなした対象の違いによる影響並びに標本調査の選択バイアス及び測定バイアスの影響について、機関記録に基づく三障害別構成比との比較等によって試算した。その結果、千人雇用事業が障害者就業率に与えたインパクトはやや過大評価であった可能性が示唆された。他方で千人雇用事業が一般就労者割合に与えたインパクトについては、やや過小評価であった可能性が示唆された。ただし、全国のアウトカム推計値のバイアスが千人雇用事業のインパクトに与えた影響は大きくないと推察される。

以上みてきたように、頑健性評価の結果によれば、本稿における千人雇用事業のインパクトは、過大評価または過小評価されている可能性がある。ただし、その過大評価または過小評価の程度は小さく、2011年度から2016年度にかけて総社市において全国と比べてより多くの障害者の就労機会が創出され、特により多くの障害者の一般就労機会が創出されたというアウトカム指標値の良い変化の多くは、千人雇用事業のインパクトだったとみなしてよいのではないかと考えられる。

2-6. 千人雇用事業の目標と達成目標が充たすべき7要件

前項までの考察により、千人雇用事業は、障害者就業率と一般就労者割合に正のインパクトを与えた成功事例だとみなしうることが分かった。では千人雇用事業の目標は、本章第1節でまとめた達成目標が充たすべき7要件を備えていたといえるのだろうか。まず①重要性については、障害者が働いて社会参加できる機会を増やすという目標は、標的集団である障害者のニーズに合致している重要な目標だと考えられるから、重要性要件は充たされていたと考えられる。②具体性については、2015年度末までに一般就労と福祉的就労を合計した障害者就労者数を1,000人にするという目標は、きわめて明確で具体的だったといえるだろう。

③現実性かつ意欲性についてはどうだろうか。現実性については、就労年齢の障害者数

を上回る障害者就労者数の達成を目指すというような非現実的で物理的に不可能な目標ではなかったという意味で、現実性要件を充たしていたと解釈できる¹⁸⁸。意欲性については、千人雇用事業は 2015 年度末までに就労年齢の障害者数の 8 割を超える 1,000 人が働いて社会参加する状態を目指すという目標であった。2011 年度当時の障害者就業率で計算すると 31.5%程度の障害者就業率を目指すという目標で、全国の障害者就業率は 2016 年度においても 26.7%に留まっていたのであるから、千人雇用事業の目標は意欲的だったといえるだろう。

④関連性については、目標変数とされた障害者就労者数は総社市の障害者就労支援策だけではなく、日本全体の景気や全国的な制度変更等の外部要因の影響を受ける可能性があると考えられる。しかし、例えば女性就労者数と比べれば、障害者就労者数に対する自治体施策の影響力は強いと考えられる。また、中央政府の障害者雇用対策は専ら障害者雇用率制度に、多くの地方自治体による障害者対策は専ら教育と福祉に偏っている状況下では、ある地方自治体が当事者目線の目標を掲げて独自の障害者就労支援策を展開したとき、その施策はその地域の障害者就労者数に大きな影響を与える可能性があると考えられる。したがって、千人雇用事業の目標は関連性要件を充たしていたと考えられる。

「⑤簡潔性」については、千人雇用事業の主要達成目標は 2015 年度末までに障害者雇用 1,000 人を達成するという非常にシンプルで明確な目標だったので、簡潔性要件を充たしていたと考えられる。「⑥実効性」については、千人雇用事業は自治体経営のトップである首長の公約で、首長が強い思いをもって強力に進めてきた事業であるから、Locke and Latham (2002) のいう実効性要件を充たしていたと言えるだろう。最後に「⑦可測性」については、千人雇用事業では期限付きの明確な数値目標が設定されていたので、理論的には可測性があったと言える。以上のことから、千人雇用事業は本稿の考える達成目標の 7 要件をすべて満たしていたと評価できると考えられる。

¹⁸⁸ 2017 年度から総社市役所が推進している「障がい者千五百人雇用事業」では、総社市における就労年齢の障害者数を上回る障害者就労者数の目標値が設定されている。そのため、現実性要件を充たしていないと解釈するのが適当ではないかと考えられる。同事業の今後の推移が注目される。

3. 小括

第2章では、政策の達成目標が充たすべき要件について検討し、政策科学の基本書と社会科学のエビデンスに基づいて、達成目標が充たすべき7つの要件を導出した。先行研究によれば、現実的だが意欲的かつ具体的で、実効性を伴う簡潔な目標が業績を高める。それに加えて重要性が大きく、政策によって直接的に達成できる測定可能な目標であれば、目標設定や評価のバイアスを小さくすることができることが分かった。

達成目標が充たすべき7つの要件を構成する個々の要件は比較的確かな根拠に基づいて導かれたものであるが、7つの要件を備えた達成目標が本当に日本の地方自治体による政策の業績を高め、政策評価のバイアスを小さくするのかは、実証研究で明らかにする必要がある。そこで、岡山県総社市「障がい者千人雇用事業」の事例を取り上げて、同事業の達成目標が本稿の考える達成目標の充たすべき要件をすべて備えていたかを検討した。具体的には、総社市及び全国の障害者就業率及び一般就労者割合を推計して、一般指標デザインに基づく差の差法を用いてインパクト評価を行い、千人雇用事業が成功事例であることを確認した後に、当該事業の達成目標が本稿の考える達成目標が充たすべき7つの要件を備えていたかを評価した。分析の結果、千人雇用事業は確かな成果を上げた成功事例で、その達成目標は本稿の考える達成目標が充たすべき7つの要件すべてを備えていたと考えられることが分かった。

本章で行った達成目標が充たすべき要件に関する検証方法は、好事例に基づく定性評価であった。したがって今後、7つの要件の妥当性を異なる研究デザインや研究対象を用いて評価することによって、仮説を追試する必要があるだろう。千人雇用事業のインパクト評価についても、より正確なデータを用いて、より頑健な結果を出すことが求められる。特に、総社市と全国の一般就労者割合の差が、2011年度においても2016年度においても20%ポイント以上あった要因を明らかにする必要があるだろう¹⁸⁹。

¹⁸⁹ 考えられる理由としては、第一に、共通ショック仮定の妥当性を検討する際に述べたように、法定雇用率適用事業所の分布の地域差が影響している可能性がある。第二に、総社市のアウトカムのバイアスを検討する際に述べたように、ハローワークを経由しない一般就労者が把握されていない等の理由によって、総社市の障害者一般就労者は実際よりも少なく見積もられている可能性がある。第三に、全国のアウトカム推計値のバイアスを検討する際に述べたように、全国の障害者数及び障害者就労者数の推計にバイアスがあり、一般就労者割合を実際の割合よりも高く見積もっているのかもしれない。より正

千人雇用事業以外の障害者就労支援策との有効性の比較評価や、千人雇用事業の効率性の評価も、今後の研究課題として残されている。一般に、差別解消を主目的とする政策の効率性評価は、便益の金銭価値換算が容易ではないため、難しいものである。特に障害者就労支援策のように、基礎的なアウトカムさえ正確なデータを得ることが難しく、他に比較可能なインパクト評価研究が見当たらない状況では、政策の効率性評価は極めて困難である。しかし、人口減少により地方自治体の予算制約が厳しくなっている中で、効率性を無視して目的の善さや有効性のみで政策の是非を論じることは、決して望ましいことではない。今後の研究課題としたい。

それでもなお、基礎自治体の障害者就労支援策が地域の障害者就業率及び一般就労者割合に影響することをおそらく初めて定量的に明らかにし、活動目標ではなく当事者目線のアウトカム目標を定めて、その指標値を定期的にモニタリングして公表し続ければ、第三者によるインパクト評価が可能であると示せたことは、日本の障害者就労支援策の在り方を問い直す上でも、より一般的に当事者目線の意欲的な目標を定めてデータを公開することの意義を示す上でも、学術的・社会的に貴重な貢献ができたのではないかと思う。日本の障害者就労支援策では当事者目線のアウトカム目標が設定されている例は少なく、管見によれば殆どの地方自治体は行政区域内の障害者就労者数さえ把握していない。そのような状況が大きく変わり、より良い目標が設定され追求されるようになることが望まれる。

障害者就業率及び一般就労者割合は、障害者の社会的包摂度合いや、障害者と健常者との格差を測るための基礎的な業績指標である。それらを把握するために、特定自治体の例外的努力やバイアスのある標本調査に基づいて推計しなければならないという現状が改善され、政府統計調査によって地域別・性別・年齢階層別等の障害者就労実態が定期的に把握され、公表されるようになることを強く期待する。

確な推計を行うために、データの入手方法や分析方法をさらに工夫していきたい。

第3章 バイアスとエビデンス

目的の正しさは、手段の正しさを保証するわけではない。社会科学の推奨手続きに従ってできるだけバイアスのないように政策問題を定義し、良い達成目標を設定しても、政策選択を誤ったり、実施方法が悪くて目標を達成できないことは十分にありうる。しかも政策がうまくいっているときほど、手段が正しいと過信するリスクも大きくなる可能性が高い。人間には自信過剰バイアスがあるため、結論ありきで論理やエビデンスよりも自分の直感に頼ってしまう傾向があるからだ¹⁹⁰。

本章では、地方自治体が目標に向かって正しい方法でやり抜くためにどうすればよいかを検討する。ここでいう「正しい」とは、その時点で利用可能な最良のエビデンスを用いて、その時点で最良の政策判断をするという意味である。最良のエビデンスをつかい、つくり、つたえるための取組みが最も進展している政策分野の一つは、保健医療である。よって第1節と第2節では、エビデンスに基づく医療の特徴を整理し、地方自治体が効果の確かな政策をつかい、つくり、つたえるための方法を検討する。続く第3節では、行動科学の文献研究及び筆者が実施した基礎自治体での社会調査結果に基づいて、認知バイアスによる判断ミスを防いで最良の政策判断をするためにどうすればよいかを論じる。最後に第4節で、自浄作用の不足によって地方自治体が誤った選択をしてしまうリスクを抑える方法について検討する。

1. エビデンスをつくり、つたえ、つかうーエビデンスに基づく医療の取組み¹⁹¹

政策問題を定義して達成目標を設定したら、政策を組み立てて実行計画を作成するプロセスに進む¹⁹²。達成目標を実現するための政策は、解決すべき政策問題に対して確かな介

¹⁹⁰ 自信過剰バイアスについては、Kahneman (2011) Chapter 3 を参照。

¹⁹¹ 本節のタイトルは津谷 (2000a) から表現を借りた。津谷 (2000a) によれば、エビデンスを「つくる」臨床試験と、エビデンスを「つかう」臨床家や行政官、消費者等の間をつないで、エビデンスを「つたえる」のが大事だと力説したのが Archiebald Cochrane であり、それを実現したのが後述する「コクラン」(旧「コクラン共同計画」)であった。

¹⁹² 政策分析は一方向に単線的に行うものではない (Bardach and Patashnik [2016])。例えば、達成目標の検討段階で政策問題のより詳細な定式化が必要になったり、政策の検討段階で目標の再考が必要になったりするものである。

入効果を期待できるものであることが望ましい。そこで近年、勘や経験ではなく効果が実証されている政策や実践を行っていかうとする動きが世界のあちこちでみられるようになっている。そのような世界的潮流を牽引してきたのは、「エビデンスに基づく医療（evidence based medicine：EBM）」の取組みである。

1－1．エビデンスに基づく医療とは

エビデンスに基づく医療とは、系統的な研究から得られた利用可能な最良のエビデンス（best available evidence）と個々の臨床的専門知識を用いて、個々の患者の治療について判断することを意味する¹⁹³。エビデンスに基づく医療の現代的起源は、「エビデンスに基づく医療の父」と呼ばれる David L. Sackett らが臨床医向けに医学論文の読み方をまとめた知見を『カナダ医師会雑誌（*Canadian Medical Association Journal*）』に公表した 1981 年にあるとされる¹⁹⁴。エビデンスに基づく医療という用語が急速に広まったのは 1990 年代のことで、Sackett et al. (2000)によれば、その理由は臨床医が次に述べる事実に気づいたからである。すなわち、保健医療に関する信頼性の高い情報に対するニーズが高まっているのに、従来の情報源からは時代遅れで誤りの多い、効果が不確かな情報しか得ることができない。しかも医師は多忙を極めており、目の前の患者のためにエビデンスを参照したり、読書や研究のために時間を割くことが難しい。そのために臨床経験を積んだベテラン医師ほど最新の情報から取り残され、適切な臨床判断がしにくくなるという事態が生じていた。その事実多くの医師が気づくようになったことが、エビデンスに基づく医療が急速に進展する原動力であった¹⁹⁵。

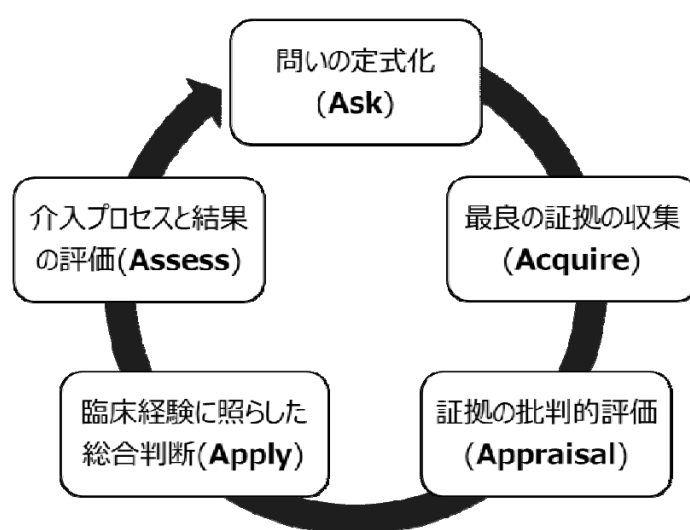
¹⁹³ Sackett et al. (1996) ; Sackett et al. (2000) p. 1.

¹⁹⁴ Smith and Rennie (2014)。Smith らの原稿の元になったイベントの様態とインタビューは、Retrieved May 29, 2019, from <http://ebm.jamanetwork.com>。Sackett らが 1985 年に公刊した『臨床疫学（*Clinical Epidemiology*）』という臨床研究の方法を詳解した書物は、「エビデンスに基づく医療のバイブル」と称されているという。

¹⁹⁵ Sackett et al. (2000) pp. 2-3。津谷 (2000b) は、エビデンスに基づく医療という用語が世界的に用いられるようになったのは、特に Sackett らが『エビデンスに基づく医療—EBM を実践し教える方法（*Evidence Based Medicine: How to Practice and Teach EBM*）』という書籍を 1997 年に公刊してからのことだと述べている。なお初めて「エビデンスに基づく医療」という用語を用いたのは Gordon Guyatt で、1992 年のことである (Smith and Rennie [2014])。

エビデンスに基づく医療の実践とは、①問題を定式化して答えられる問いを立てる、②定式化された問いに答えるための最良のエビデンスをできる限り効率的に見つけ出す（そして最良のエビデンスの二次情報源を増やす）、③そのエビデンスの妥当性、有効性、有用性を批判的に吟味する、④評価結果を自らの臨床の専門知識及び患者の状況や価値観と統合して臨床的判断を下す、⑤介入プロセスとアウトカムの効果効率を評価して次に活かす、という自主学習プロセスを生涯にわたって続けることを意味する（図3－1参照）¹⁹⁶。

（図3－1）エビデンスに基づく医療実践の5つの（5A）ステップ



（出典）Sackett（1997）；Sackett et al.（2000）；Johnson（2008）を参考にして筆者作成

ここでいう「最良のエビデンス」とは、バイアス（bias）とバラつき（variance）の小さい、利用可能な最新のエビデンスのことである¹⁹⁷。忙しい医師が、限られた時間内に特定の臨床的な問いに答えるためのエビデンスを膨大な情報の中から漏れや偏りなく選び出し、その全てを批判的に吟味することは非常に困難である。医療現場で利用可能な最良のエビデンスが臨床的判断に用いられるようにするためには、医師に代わって最新のエビデンスを系統的に収集し、それらを批判的に吟味した上で最良のエビデンスを生産し、必要ときに必要な情報を素早く見つけ出せるようにする仕組みが必要である。その実現を目指し、世界におけるエビデンスに基づく医療を牽引してきた代表的な活動が、「コクラン」（かつ

¹⁹⁶ Sackett（1997）；Sackett et al.（2000） pp. 1-12。

¹⁹⁷ 津谷（2000b）。

ては the Cochrane Collaboration：コクラン共同計画、現在は Cochrane）である。

1－2．最良のエビデンスをつくり、つたえるーコクランの5つの特徴

コクランとは、研究を通じて生み出された膨大なエビデンスを保健医療に関する意思決定に役立てるために活動している、グローバルな非営利組織である。そのミッションは、良質で価値の高いアクセス容易な系統的レビュー（systematic review）をはじめとする総合的な研究エビデンスをつくり、つたえることによって、エビデンスに基づく医療を促進することである¹⁹⁸。保健医療に関する「系統的レビュー」とは、単なるレビューとは異なり、バイアスを減らすようにデザインされた方法を用いて、臨床的な問いに関する研究を網羅的に見つけて批判的に吟味し、関連する研究を統合して要約することを意味する。

コクランの主要な特徴の一つは、最も信頼できるエビデンスはランダム化比較試験（randomized controlled trials：RCTs）に基づくエビデンスの系統的レビューであるという価値観に基づいて二次情報を生産している点にある。ランダム化比較試験とはいわゆる対照実験のことで、標的集団をランダムに介入集団と対照集団に割り付けることによって、集団間に系統的な差が生じることを避けて介入効果のみを評価することを目的とした実験的研究手法のことである。

コクランのアイディアは、Archiebald L. Cochrane が 1972 年に公刊した『効果と効率（*Effectiveness and Efficiency*）』という書籍から生まれた。そこで Cochrane が掲げたスローガンは、英国の国民医療保険サービス（National Health Services：NHS）は、「すべての有効な介入を無料で提供すべきである」¹⁹⁹というものであった。ポイントは、「すべての介入」ではなく「すべての有効な介入」である点だ。それを可能にするためには、どのような治療がなされたのかという歴史研究ではなく、ランダム化比較試験によって介入効果を科学的に測定し、害よりも益の方が大きい治療とケアを効率的かつ公平に提供することが必要である。ランダム化比較試験による介入効果の測定評価が重要な理由は、介入群と対照群をランダムに割り当てると被験者の個体差変数が偶然に配分されるため、介入効果が選択バイアスによって系統的に歪められることがなくなるためである。ただしラン

¹⁹⁸ コクランのウェブサイトを参照（Retrieved May 29, 2019, from <https://www.cochrane.org/>）。

¹⁹⁹ Cochrane (1972) p. 1 より引用。

ダム化比較試験に基づくエビデンスを生産しても、その系統的レビューが定期的に行われ、臨床医がすぐにアクセスできるようになっていなければ、最良のエビデンスに基づく有効な介入を効率的かつ公平に提供することはできない。したがって、もっとランダム化比較試験に基づくエビデンスをつくり、つたえ、つかわれるようにしなければならないと Cochrane は主張したのだった²⁰⁰。

コクランの二つ目の特徴は、再現性の高い良質なエビデンスをつくり、つたえるための方法論の開発と改善に力を注いでいる点にある。コクランがつくる系統的レビュー、すなわち「コクラン・レビュー (Cochrane Reviews)」の中心は、介入効果に関するレビューである。その生産と公表に責任をもつ組織を「コクラン・レビューグループ (Cochrane Review Group : CRGs)」という。2018 年 5 月時点で分野別に設定された 50 を超えるグループが、信頼できる最新のレビューをつくるために精力的に活動している²⁰¹。数多くあるレビューグループが良質で再現性の高い二次情報を生産できる大きな理由は、系統的レビューをつくるための優れた方法論が開発されているからである。コクランでは、「コクラン方法論グループ (Cochrane Methods Group)」が中心となって作成する『介入の系統的レビューのためのコクラン・ハンドブック (*Cochrane Handbook for Systematic Reviews of Interventions*)』を用いて系統的レビューが行われている。コクランの系統的レビューの方法論は、医療の世界で研究が遅れていた「つたえる」局面の質の研究を大きく進展させ、エビデンスに基づく医療における系統的レビューの方法論の開発と発展に大きく貢献していると評価されている²⁰²。

コクランの三つ目の特徴は、最良のエビデンスを臨床現場の医師が必要なときにすぐに使えるようにすることに大きなエネルギーを注いでいる点である。まず、エビデンスへの

²⁰⁰ ランダム化比較試験とその系統的レビューは、理想的な状況では選択バイアスを除去できるという意味で黄金律 (golden standard) であるものの、他のインパクト評価方法と同じように一長一短がある。そのため、後述する GRADE システムのように研究デザインの信頼性だけでなく、その他の要素も勘案してエビデンスの質を評価する方法が提案され、適用が進んでいる (相原 [2018] ; Rawlins [2018])。エビデンスに基づく医療で用いるべきエビデンスは臨床上の問いに答えるために最良の外部証拠であって、因果推論の手法によって先験的に決定できないことは、Sackett (1996) でも指摘されている。

²⁰¹ コクランのウェブサイトを参照 (Retrieved May 29, 2019, from <https://www.cochrane.org/>)。

²⁰² 津谷 (2000b) ; Chandler and Hopewell (2013)。

アクセスとコミュニケーションを容易にするために、ICT が大いに役立てられている²⁰³。例えばコクラン・レビューをつくり管理するためには、「Review Manager 5」というアプリケーションや「RevMan Web」というオンラインツールが用意されている。これらのツールを使えば、メタ・アナリシスのような定量分析や結果の図示化、所定の書式に基づくレポート作成等の作業を簡単に行うことができる。つくられた各種レビューはデータベース化されて「コクラン・ライブラリー」としてデジタル形式で提供されているため、利用者が自分の関心のある最新の医療情報にアクセスするのも容易である²⁰⁴。2018 年からは、「コクラン・クリニカルアンサー (Cochrane Clinical Answers : CCAs)」という、コクラン・レビューに基づく良質なエビデンスを簡潔にまとめてすぐ参照できるようにした文書も提供されている。

コクランの四つ目の特徴は、活動の透明性、開放性、独立性が高く、その活動が保健医療に関係する多様な分野でグローバルな規模で広がり、支持されていることである。コクランは 10 個ある活動原則 (The 10 principles of Cochrane) の一番目に「協働」を掲げ、多くの人を巻き込みながら開放的で透明性の高いコミュニケーションと意思決定を促進することに力を入れている。コクランの前進であるコクラン共同計画は 9 か国から 77 人の参加者を得た会合で 1993 年にスタートしたのだが、ボランティアな参加者は 2000 年の 64 か国 5,437 人から 2010 年 2 月の 112 か国 27,792 人へと 5 倍以上増えた。人数が増えただけでなく、多様性も向上した。コクラン・レビューグループのメンバーにおける低所得国または中所得からの参加者の割合は、2000 年の 5.6% から 2010 年の 19% へと高まったという²⁰⁵。コクランは非政府非営利団体なので、利益相反を心配する必要性も小さい。そのため、世界中の数多くの医療関係者や研究者、患者支援者などがコクランの活動に関わり、いわば公共財として良質なエビデンスを使うことのできる環境が形成されているの

²⁰³ コクランのウェブサイトの「Tools and software」でツールやソフトウェアを一覧できる (Retrieved May 29, 2019, from <https://community.cochrane.org/help/tools-and-software>)。

²⁰⁴ コクラン・ライブラリーとは、コクランが作成し Wiley が発行しているデータベース群のことである。その中心は、コクラン・レビューとそのプロトコルを格納した「系統的レビューのコクラン・データベース (Cochrane Database of Systematic Reviews : CDSR)」である。一部はコクラン・ライブラリーのウェブサイトから利用可能だが、完全版は有料でオンラインまたは電子媒体で提供されている。(コクラン・ライブラリーのホームページ, Retrieved at May 29, 2019, from <https://www.cochranelibrary.com>)。

²⁰⁵ Allen and Richmond (2011)。

である。2018 年 5 月時点では、130 か国を超える国から約 11,000 人のメンバーと約 78,000 人のサポーターがコクランの活動に関与している²⁰⁶。

コクランの取組みがグローバルな規模で広がっている背景には、上記に挙げた 4 つの特徴に加えて、もう一つ重要な特徴があると考えられる。それは、制度化から着手したのではなく、一気呵成に網羅的にやろうとしたわけでもなく、ひと握りの人たちが一部の分野で自発的に作り上げた実績をベースにしてスタートし、パッチワーク的に広がっていったという事実である。コクランは、Cochrane (1972) に衝撃を受けてその後の人生をエビデンスに基づく医療の確立に懸けることになった Iain Chalmers らが 10 年以上かけて周産期医療の分野で蓄積してきたランダム化比較試験の系統的レビューのデータベースを基礎にして開始された取組みであった²⁰⁷。いきなり保健医療の全分野を網羅しようとするのではなく、パッチワーク的に系統的レビューを行っていけばいずれ保健医療分野全体をカバーできるだろうという見通しのもとで、じょじょに扱う領域を広げ、ガイドラインやツール等を少しずつ開発し改善していったのだという²⁰⁸。つまりコクランは、「小さくはじめて、大きく育てる」という方針に基づいて、ボトムアップ型で広がっていったのであった。

1-3. エビデンスの確実性の評価と推奨度の評価の峻別

コクランは、図 3-1 に示したエビデンスに基づく医療の実践ステップのうち、「最良の証拠の収集 (Acquire)」及び「証拠の批判的評価 (Appraisal)」のステップを専門的に担っている。コクランのように系統的レビューをつくり、つたえる活動が盛んになったことで、診療ガイドラインの作成とその方法論の開発も急速に進展している。診療ガイドラインとは、エビデンスの系統的レビューと複数の治療選択肢に関する益と害の評価に基づいて患者ケアを最適化するための、推奨 (recommendation) を含む文書のことで、実態や原因がどうなっているかではなく、対策として何をすべきかという前景疑問 (foreground question) に答えるものである²⁰⁹。診療ガイドラインが充実することによって、医師が目の前にいる患者に対して、最良のエビデンスに基づいて科学的に妥当な手順に従って治療

²⁰⁶ コクランのウェブサイトを参照 (Retrieved May 29, 2019, from <https://www.cochrane.org/>)。

²⁰⁷ Smith and Rennie (2014) ; Cassels (2015) を参照。

²⁰⁸ 津谷 (2000a) を参照。

²⁰⁹ 相原 (2018) p. 2 及び p. 22 を参照。

やケアを行うことが容易にできる環境が整いつつある。

診療ガイドラインを代表とする保健医療のガイドラインを作成する際に国際的に用いられているシステムに、「GRADE (Grading of Recommendations, Assessment, Development and Evaluation)」がある²¹⁰。GRADE システムは保健医療のガイドライン作成の手引きであり、エビデンスの要約を作成し、推奨を決定するまでの手続きについて、透明性が高く科学的に妥当で構造化された手続きを提供している。GRADE システムにおけるエビデンスの確実性の評価方法や推奨の強さの判断方法は、地方自治体が住民や利害関係者等の合意に基づいてエビデンスに基づく政策を選ぶ際にも参考になると思うので、少し丁寧にみていく。

GRADE システムでは、エビデンスの確実性は4段階で、推奨の強さは2段階で表される。入手可能な最良のエビデンス総体 (body of evidence) の確実性は、まずアウトカムごとに評価される。最初に、研究デザインがランダム化比較試験なら最高ランクの「高い」、それ以外の観察研究なら上から3番目のランクの「低い」と評価される。かつては研究デザインに基づく信頼性の階層がより深かったが、現在の GRADE システムでは、研究デザインだけで安易にエビデンスの質を判断しないように、ランダム化比較試験以外の研究デザインを全て観察研究に含め、独立した対照群がないこと等によって多くの場合には確実性が「非常に低い」という最低ランクと位置づけられるような仕組みになっている。

研究デザインでエビデンスを2段階に格付けしたら、次にグレードを下げる5つの要因について評価する。すなわち、①研究の限界 (バイアスのリスク)、②結果の非一貫性・異質性、③エビデンスの非直接性、④データの不精確さ、⑤出版バイアスについて、「深刻」な場合にはグレードを1段階、「非常に深刻」の場合にはグレードを2段階下げる。ただし、同じ要因を機械的に重複カウントしてグレードを下げすぎないように注意する必要がある。

ダウングレード要因について評価したら、グレードを上げる3つの要因について評価する。アップグレード要因は原則として、上記の①から⑤のダウングレード要因のいずれにおいても深刻な問題が見られない観察研究に適用する。⑥効果の程度が大きい場合には、その程度によってグレードを1から2段階上げる。⑦用量反応勾配 (dose-response gradient)²¹¹がある場合には、グレードを1段階上げる。⑧考えうる全ての交絡因子の影

²¹⁰ GRADE システムの説明は相原 (2018) に基づく。GRADE システムについては、Dijkers (2013) 及び Schünemann et al. eds. (2013) も参考にした。

²¹¹ 「用量反応勾配 (dose-response gradient)」とは、曝露 (exposure) と相対リスクの間

響が同じ方向に働くことで見かけ上の効果が過大または過小に評価されていると考えられる場合には、グレードを1段階上げる。

以上のプロセスを経てアウトカムごとのエビデンス総体の確実性は「高い」、「中くらい」、「低い」、「非常に低い」の4段階に格付けされ、その結果はガイドライン作成ツールを用いて、所定の様式でエビデンスプロファイルや所見の要約として報告される。

推奨の強さと方向性の評価は、アウトカムごとにエビデンス総体の確実性を4段階で評価した後に行われる。まずアウトカムの相対的な重要性を再評価した後、①重大及び重要とされたアウトカム全般にわたる全体的なエビデンスの確実性、②望ましい効果（益）と望ましくない効果（害）のバランス、③価値観と意向²¹²、そして必要に応じて④必要資源量・コストを勘案して、推奨の強さを「強い」、「弱い/条件付き」の2段階で、推奨の向きを「推奨する」、「推奨しない」の2段階で判断する。

推奨の判断で興味深いのは、集合知を活かす工夫が埋め込まれている点である。推奨作成における各ステップで下される判断では、修正デルファイ法（RAND 法）等を用いて、全員または事前に設定した大多数が合意することが必要とされる。具体的には例えば、アウトカムの相対的な重要性を再評価するために、各パネリストには会議の前にメンバー間の意見交換なしに1～9点の評価尺度を使って個別評価を行うことが求められる（第1ラウンド）。1～9点の評点付けは、意思決定にとって重大（7～9点）、重要ではあるが重大ではない（4～6点）、重要ではない（1～3点）の3つに分類することが推奨されている。その結果を資料として、主に評価が食い違う分野を中心に会議で議論した後、再度個別に評価を行う（第2ラウンド）。斉一性のプレッシャーによる集団浅慮に陥ることを避けるため統一見解は決して強要されず²¹³、第2ラウンドの結果として得られた中央値や不一致率に基づいて推奨の程度が判定される。例えば参加者の70～75%が7～9点を付けたアウトカムを意思決定にとって重大とみなす、といった方法がとられる²¹⁴。

以上みてきたように GRADE システムは、エビデンスの確実性の評価と推奨の強さの判断を切り離し、その両方について科学的に妥当で透明性の高い方法で評価する点にその特

の関係性のことである。例えば1日当たり喫煙量と肺がんによる死亡リスクとの関係性を指す。

²¹² GRADE システムにおける「価値観と意向」とは、人々が保健医療のアウトカムにおく相対的な重要性のことである。

²¹³ 集団浅慮による判断ミスについては、本章第3節で述べる。

²¹⁴ 相原（2018）pp. 317-320 を参照。

徴がある。政策設計においては、科学的妥当性だけではなく、住民や市民にとっての必要性、標的集団の行政ニーズ、効果とコストのバランス等も重視する必要がある。それゆえ、エビデンスの科学的な確実性だけではなく患者の価値観や意向を重視した上で、介入または非介入をどのくらいの確信度で推奨できるかを透明性と再現性の高い方法で格付けするという GRADE システムの考え方と手法は、公共政策のインパクト評価においても大いに参考になるのではないかとと思われる。

2. エビデンスに基づくインパクト評価の実践方法

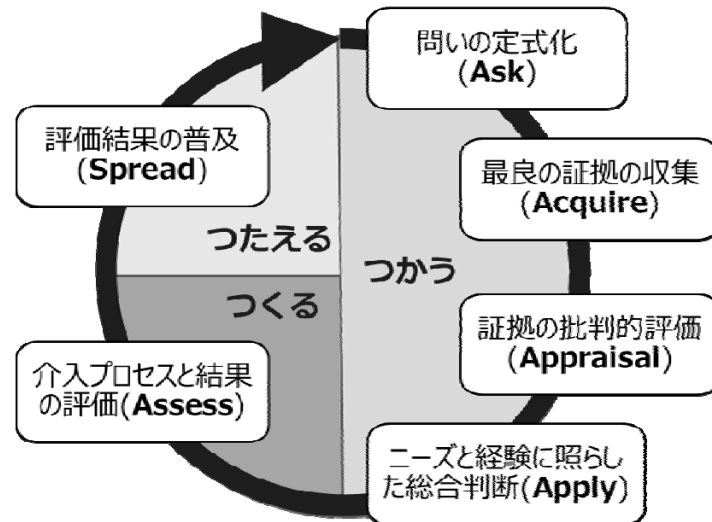
前節で整理した取組みは、エビデンスに基づく医療の取組み全体における、一部の取組みに過ぎない。他にも例えば、チェックリストの活用等により医療ミスを軽減した臨床現場での挑戦²¹⁵や、医療保険の費用対効果を大規模実験で明らかにしたランド医療保険実験²¹⁶等、興味深い取組みは存在する。それでもなお、エビデンスに基づく医療がグローバルな規模で急速に進展した大きな理由の一つは、科学的に信頼できる情報を生産し普及させようとするコクランや GRADE システムのような取組みによって、ある状態の患者に推奨される、最良のエビデンスに支えられた「標準治療」の情報に、臨床医が容易にアクセスできるようになったためだと言ってよいだろう。

では、目標達成に向けて正しい方向に進むために、地方自治体はエビデンスに基づく医療の経験から何を学ぶべきなのだろうか。図 3-2 に、エビデンスに基づくインパクト評価の実践ステップを示した。図 3-1 の「エビデンスに基づく医療の実践ステップ」に「つたえる」というプロセスを加え、公共政策における政策形成段階の政策分析から政策実施後の事後評価までのプロセスに適用できるように用語を微修正したものである。本節では図 3-2 に沿って、エビデンスを「つかう」、「つくる」、「つたえる」の順に地方自治体は何をすればよいのかを考察する。

²¹⁵ チェックリストの活用による医療ミス軽減の試みについては、本章第 3 節で後述する。

²¹⁶ Manning et al. (1987) ; Chernew (2008) を参照。

(図3-2) エビデンスに基づくインパクト評価の実践プロセス—5 A + 1 S ステップ



(出典) 図3-1を参考にして筆者作成²¹⁷

2-1. 効果が確かな政策を選ぶ—エビデンスを「つかう」

地方自治体がまず行うべきことは、外部の良質なエビデンスを「つかう」ことである。最初のステップは、問いの定式化である。GRADE システムでは、系統的レビューを行う前にまず問いを PICO フレームワーク形式で定式化してから²¹⁸、アウトカムを列挙して相対的な重要性を判定し、エビデンスの適格性基準等を決めることが求められる。調べる前に何を知りたいのかを明確にすることで、場当たりの調査によるバイアスを防ぐことができるからだ²¹⁹。第1章で述べたように、所与の制約の中で最大の成果を得るためには、証拠を集める前によく考えなければならないということである。

問いを明確に定義したら、関連研究を網羅的・系統的に収集して、詳しく分析する研究を取捨選択するというステップに移る。エビデンスに基づく医療では、エビデンスを探し始めるときには関連研究の系統的レビュー論文を探すのがよいとされる。保健医療分野以外でも、社会政策や開発援助等の分野では、特に英語で系統的レビューをつくり、つたえ

²¹⁷ 評価結果の「普及」には「disseminate」が用いられることが多いが、「spread」の方が一般的な日本人にとって馴染みがある英単語だと思われるので「spread」を用いた。

²¹⁸ PICO とは臨床上の問いを明確化するためのフレームワークで、「P」は患者 (patients) ないし母集団 (population)、「I」は介入 (intervention)、「C」は代替介入ないし対照介入 (comparison)、「O」は関心のあるアウトカム (outcomes) を意味する。

²¹⁹ Khan et al. (2003) を参照。

る活動が進んでおり、インターネット経由で豊富なエビデンスにすぐにアクセスできるような環境が整いつつある²²⁰。しかし政策分野が限られているし、海外のエビデンスが日本で妥当するとは想定しにくいケースも多いだろう。したがって資源が許す限りで外部の良質なエビデンスを系統的に収集し、その質を評価して適切に報告する方法を習得することが求められる。

地方自治体の職員が関連研究を網羅的・系統的に調べて適格性を評価し、科学的に信頼できる方法で統合するのは容易ではないかもしれない。しかし、政策の根拠に用いた資料をどのような基準でどのように選んだかという過程を記録してフローチャートを公表することは、技術的には難しいことではない。エビデンスの収集・選別プロセスの透明性と再現性を高めれば第三者による妥当性の検証が可能になるため、より有効でバイアスの小さい政策選択が促されるであろうし、権力はないが有効な政策を実施したいと願う志の高い人々にとっての強力な交渉ツールにもなりうる²²¹。

エビデンスに基づく医療における系統的レビューの報告方法については、「PRISMA-P 2015 (the Preferred Reporting Items for Systematic reviews and Meta-Analyses for Protocols 2015)」という広く用いられているガイドラインがある²²²。2009 年時点の PRISMA 声明では、系統的レビューのプロセスを図 3-3 に示したようなフローチャートを用いて公表することが推奨されている。フローチャートのような形式による、政策形成過程で参照したエビデンスの経時的記録と公開は、すぐにも取り入れるべき手続きである

²²⁰ 社会政策分野では、2000 年に設立された非営利組織である「キャンベル共同計画 (Campbell Collaboration)」がコクランの活動からスピノフしたデータベースを礎にして社会政策分野における系統的レビューとその方法論をつくり、つたえる活動を行っている(キャンベル共同計画のウェブサイトより。Retrieved at June 2, 2019, from <https://campbellcollaboration.org/>。津富 [2003] も参照)。開発援助分野では、前述した貧困アクションラボや世界銀行などがランダム化比較試験に基づくエビデンスをつくり、ひろめる活動を精力的に行っている(各ウェブサイトより。貧困アクションラボの活動は Retrieved at June 2, 2019, from <https://www.povertyactionlab.org/>、世界銀行の活動は Retrieved at June 2, 2019, from <http://www.worldbank.org/en/programs/knowledge-for-change>)。

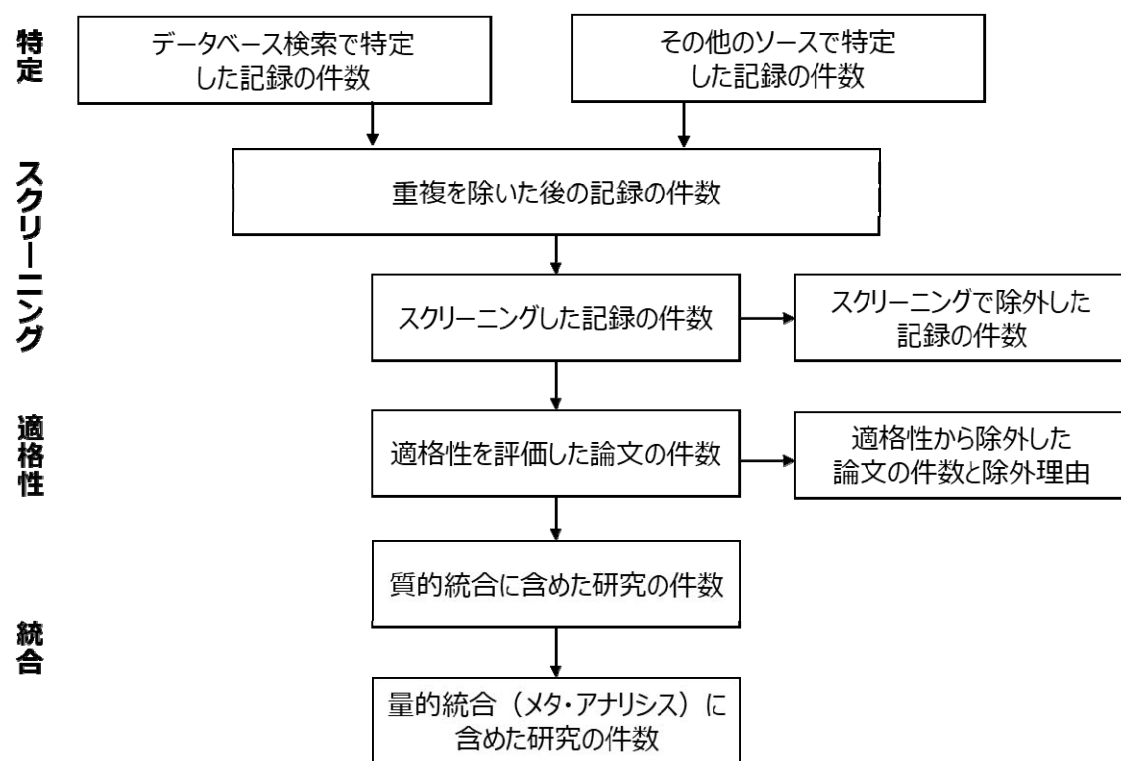
²²¹ 「エビデンスに基づく医療」という用語を初めて用いたとされる Gordon Guyatt は、エビデンスに基づく医療は、最も権威ある専門家や一番声の大きい人の言うことを聞き入れるのではなく、自分の患者にとって何が最良かを自分で判断するという驚くべき力を臨床医に与えてくれると述べている (Guyatt and Rennie eds. [2003] p.v)。

²²² PRISMA 声明については、Moher et al. (2009) ; 卓ほか (2011) ; Moher et al. (2015) を参照。

う²²³。

なお、行政職員が自分たちだけで資料を分析評価する必要はないし、第1章で述べたように文書だけを証拠に用いるのはむしろ避けるべきである。GRADE システムにおいて専門家の意見はエビデンスの分類に含まれておらず、含めるなら基本的に確実性が「非常に低い」と位置付けられるべきだとみなされているが、同時に専門家には信頼できるデータやエビデンスに関する有益なアドバイスを提供することが期待されている²²⁴。政策審議会等において専門家に対してデータやエビデンスに基づく意見を求めたり、データやエビデンスの質を明確な基準に基づいて専門家に評価してもらうことは、より多様で良質なエビデンスを効率的に入手するための簡便な手法として有用であろうし、専門家に対してよりバイアスの小さい意見の表明を促すことにもなるであろう。

(図3-3) 2009年版PRISMA声明で推奨されている系統的レビュープロセスの模式図



(出典) Moher et al. (2009) Figure 1 (doi:10.1371/journal.pmed.1000097.g001) を和訳。

²²³ 本稿第1章第4節図1-5に示した中核市調査のスクリーニング・プロセスは、PRISMA形式で表現した。

²²⁴ 相原(2018) p. 39を参照。

政策を決定するためにできるだけ良質なエビデンスを用いる理由は、行政が手を差し伸べなければならない人々に対して、できるだけ有効な政策を効率的かつ公平に講じるためである。資源の許す限りで利用可能な最良のエビデンスを収集して批判的に評価したら、経験やニーズ等に照らして政策を選び組み立てるステップに進む。エビデンスに基づく医療では、利用可能な最良のエビデンスを用いて、臨床的な専門知識と患者の状態や価値観を勘案して臨床的判断を下すことが望ましいとされるのだった。地方自治体による政策判断においても、エビデンスの確実性だけではなく、政治的・制度的制約のような行政に関する専門知識や行政現場での経験、そして住民と市民の価値観に基づいて、政策の費用対効果やリスク、誰のために何を犠牲にできるのか等を天秤にかけて価値判断することが求められる。そのような政策判断を公平で透明性・開放性の高いプロセスを通じて行う方法の一つは、GRADE システムで採用されている修正デルファイ法に基づく合意形成であるが、この方法がより良い意思決定をもたらすためには、参加主体の自律性だけではなく多様性も保証されている必要がある。集団としての推定精度が高まるのは判断の独立性が保たれているとき、つまり誤差相関がないときであり、参加主体が同じバイアスを持っているときには多数の判断を集めても実質的な標本サイズは大きくならないので、集団としての推定精度は高まらない²²⁵。したがって第1章第2節で「公聴会や地区住民懇談会では参加者が限られ政治的バイアスがかかりやすい点に注意する必要がある」と述べたように、政策判断を行う際には、意見に偏りがある参加者ばかりを集めて議論しても判断のバイアスは減じられない点に注意が必要だろう。

2-2. 介入プロセスと介入結果を評価する—エビデンスを「つくる」

政策を決定した後で地方自治体が行うべきことは、エビデンスを「つくる」こと、つまり、介入プロセスと介入結果を信頼できる方法で評価することである。エビデンスに基づく医療という考え方に初めて接すると、エビデンスがない介入をしてはいけないと考えてしまいがちであるが、標準的な手続きを実践しはじめると世の中にはエビデンスのないことの方が多いので、ではどうしようと考えようになるという²²⁶。公共政策のエビデンスは一般化がより困難なケースが多いと考えられるから、自治体評価の場合には尚更である

²²⁵ Kahneman (2011) Chapter 7 を参照。

²²⁶ 津谷 (2000a) p. 129 を参照。

う。したがって政策効果をしっかり評価し、そのときどきの社会情勢や標的集団の行政ニーズに合致した政策にアップデートし続けなければならない。

政策のインパクト評価の方法論については近年、日本語の良書が数多く出版されている²²⁷。中でも日本の中央政府におけるエビデンスに基づく政策形成への関心の高まりに大きな影響を与えたと思われる書物は、教育経済学を専門とする中室牧子が医療政策学・医療経済学を専門とする医師である津川友介と共にインパクト評価の方法論を解説した『「原因と結果」の経済学』である。同書によれば、政策効果を評価する際には、①原因と結果、つまり政策手段とアウトカムを明確に定義すること、②本当に政策がアウトカムに影響を与えたといえるのかを確認すること、③政策を講じなかった状態という反事実（反実仮想）を作り出して比較することが必要である²²⁸。①の説明は不要だろう。②を確かめるに当たっては、まったくの偶然ではないか、政策とアウトカムの両方に関係のある第三の要因²²⁹が介入効果に影響しているのではないか、逆の因果関係はないかをチェックするとよいとされる。③は比較可能なグループを作り出し、そのアウトカムをもっともらしい値で置き換えて比較してはじめて介入によるアウトカムの変化分を明らかにできるということである。

反事実をつくりだす方法として最も科学的に信頼できる研究デザインが、ランダム化比較試験である。よく設計されたランダム化比較試験に基づくエビデンスは、内的妥当性が高いだけでなく²³⁰、介入効果の算出方法が極めてシンプルなので、計量分析の専門知識がなくても介入効果を計算したり理解することが容易であるという特徴を持つ。エビデンスの質は研究デザインだけで決まるわけではないし、行政機関が追求すべきなのは学術研究のような方法論の精度ではないが、小規模でよいから厳格な方法で試行的に実験を実施して介入効果を確認してから本格的な政策実施に踏み出せば、費用対効果の高い政策をより確実に講じることが可能になる²³¹。実験以外の研究デザインを採用する場合でも、反事実をつくりだして介入効果を測定する方法を事前に決めておけば、一般に対照集団のないインパクト評価と比べてより確実性の高いエビデンスをつくることができる。政策効果をき

²²⁷ 例えば佐々木（2000）；山本（2015）；中室・津川（2017）；伊藤（2017）等。

²²⁸ 中室・津川（2017）。特に同書，pp. 180-182 を参照した。

²²⁹ 交絡因子のことである。

²³⁰ 内的妥当性とは、同じ研究を繰り返したときに同じ結果が再現される程度のことをさす。外的妥当性とは、ある研究で認められた研究結果を一般化しうる程度のことを指す。

²³¹ Halpern and Mason（2015）でも強調されている点である。

ちんと測定評価できなければ、成功や失敗から学習することはできない。まずは倫理的な問題が殆ど生じないような領域で、多くのコストがかからない小さな実験を試行してみる価値は大きいのではないだろうか。

2-3. データやエビデンスをつかってもらうーエビデンスを「つたえる」

地方自治体が政策形成の根拠に用いたデータやエビデンスの情報や、実際に講じた政策の効果に関する情報は、できるだけ再利用容易な形式で広く公開して市民や他自治体や専門家に使ってもらえるようにすべきである。そうすることによって次につかうことのできるエビデンスの量が増え、より公平で効果的な政策をより効率的に講じることができるようになるからだ。ところが行政の無謬性が前提とされる社会では、地方自治体がネガティブな政策効果を広く知らしめることは難しく、成功事例しか公表されなかったり、失敗事例も成功事例であるかのように偽装されてしまう恐れがある。

ネガティブな結果が出た研究はポジティブな結果が出た研究に比べて公表されにくいというバイアスを「出版バイアス」という。系統的レビューを行ってエビデンスを統合するとき出版バイアスに考慮しないと、介入の有効性を過大に、介入によるリスクを過小に評価することになる。介入の有効性や安全性の誤認は、誤った意思決定を誘発する避けねばならない事態である。したがって GRADE システムでは、出版バイアスがグレードダウン要因として考慮されている。公共政策に関する出版バイアスを防ぐためには、失敗から学習する文化を組織内外で醸成し、地方自治体がエビデンスを正直につたえることができるようにする必要がある。その対策については第4章で改めて検討する。

政策プロセスと政策効果の情報公開・データ開放に加えて不可欠なのは、コクランが協働を重視して開放的に活動しているように、エビデンスに基づくインパクト評価をつかい、つくり、ひろめるプロセスに市民や利害関係者や専門家を巻き込むことであろう。多様な人々と協働することで行政だけでは気づかないデータやエビデンスを入手しやすくなるであろうし、政策選択の失敗や政策運営のミス等を傍観者として無責任に批判することが難しくなり、ポジティブな変化が広がりやすくなることも期待される。中央政府にもまずは政策形成主体として、相対的に豊富な資源を駆使して、エビデンスをつかい、つくり、つたえる活動を、透明性・開放性の高いプロセスで進めることを期待したい。

3. 個人と組織の認知バイアスに対処する

3-1. 政策に基づく証拠形成と集団浅慮

評価学に科学的評価という概念を定着させるのに大きな影響を与えた Donald X. Campbell が、バイアスのあるエビデンスに基づいて有効な政策を選び出すのではなく、重要な政策問題を選び出し、ランダム化比較試験や回帰分断モデルのような科学的な方法で介入効果を検証することによって何が最適な政策か判断すべきだと主張したのは、1969年のことであった²³²。それから約 40 年後に米前大統領 Barack Obama 政権が挑んだ悪習の一つは、政治家がエビデンスを自分の正しいと思う政策を正当化するための武器として使うことであった²³³。そのようなエビデンスの使用法はしばしば、「政策に基づく証拠形成 (policy-based evidence making)」と揶揄される。

自分の信念を支持するためにエビデンスを使うという行為は、人々の判断や意思決定を歪めるために意図的に行われているとは限らない。政治家に限らず私たち人間は、仮説や自分の信念を検証するとき無意識にそれを肯定する証拠ばかりを探し、反証となる情報を無視したり集めようとしない傾向があるからだ。このような傾向は、「確認バイアス (confirmation bias)」と呼ばれる²³⁴。多様な分野で数多くの頑健なエビデンスによって支持されている認知バイアスの一つである。人間の認知にとって多くの場合「自分の見たものが全て」で、エビデンスの量や質よりも、自分の信念を肯定するエビデンスに基づいて首尾一貫したストーリーを紡ぐことの方が大切なのである²³⁵。

非系統的な臨床観察がランダム化比較試験の系統的レビューによって統合されたエビデンスに劣る理由は、標本サイズが小さくデータのバイアスが大きいだけでなく、人間の推論過程の欠点を免れえないからである²³⁶。日本の地方自治体が実施する政策についてはコクラン・レビューや診療ガイドラインのような文書がまだ存在しないので、地方自治体はエビデンスの確実性に注意を払うだけでなく、人間の推論過程に潜む認知バイア

²³² Campbell (1969) ; Rossi et al. (2004) p. 23 を参照。

²³³ Haskins and Margolis (2015) pp. 1-30 を参照。

²³⁴ 確認バイアスについては、Nickerson (1998) を参照。

²³⁵ 「自分が見たことが全て」を引き起こす認知バイアス（確認バイアスを含む）については、Kahneman (2011) Chapter 7 を参照。

²³⁶ Guyatt and Rennie eds. (2003) pp. 5-59 を参照。

スに対して自覚的に対処しなければならない。

確証バイアスによる判断のバイアスは、個人だけではなく集団でも生じうる。Irving L. Janis によれば、組織が「集団浅慮（group think：集団思考）」に陥ると、組織的意思決定が確証バイアスによって歪められてしまう²³⁷。集団浅慮とは、団結力の高い集団で無意識のうちに生まれる団結心によって、グループの仮説や信念に異論を唱えることよりも、集団の和を乱さないことが優先されてしまうという思考様式を指す。集団浅慮の先行条件とされる要素は、集団の凝集性（cohesion）、組織の構造的欠陥、そしてプレッシャーの強い状況である。組織の構造的欠陥とは、集団の孤立、公平なリーダーシップの伝統の欠如、方法論的な手順を要求する規範の欠如、及び、メンバーの社会的背景やイデオロギーの均質性等のことである。これらの先行条件があると、自信過剰、偏狭な党派的視野、そして斉一性（uniformity）への圧力という集団浅慮の兆候がみられるようになり、確証バイアスによって判断ミスを行ってしまうという。Janis（1982）によれば、ピッグス湾事件や真珠湾攻撃での意思決定の失敗は、集団浅慮に陥ったために代替案分析が不十分になったり、皆が賛成する選択肢のリスク分析を怠ってしまったり、手元にある情報だけで判断してしまうことによってもたらされたものだった。他方で、キューバミサイル危機やマーシャルプランでの政策判断の成功は、集団浅慮に陥らないように懐疑主義が貫かれ、儀礼的な議論や形式的な分析評価ではなく、忌憚ない議論と実質的な分析評価が行われたためにもたらされたのであった。

3-2. 確証バイアスによる判断ミスを組織的に防ぐ方法

では個人や組織は、認知バイアスに対してどのように対処すればよいのだろうか。Kahneman（2011）によれば、確証バイアスをはじめとする認知バイアスによる判断ミスを防ぐ方法は、原理的にはシンプルである。判断ミスが生じやすい状況を特定し、それを防ぐための方法を学習すればよい²³⁸。人は認知バイアスを完全に克服することはできないが、判断のバイアスに囚われやすい人とそうでない人がいることも分かっている。専門家の政

²³⁷ Janis（1982）。Janis の集団浅慮理論は組織的意思決定に関する古典的文献で、集団が優れた意思決定を行うために自覚的に避けなければならないものとして最近の研究でもしばしば取り上げられている（例えば Tetlock and Gardner [2015] Chapter 9； Sunstein and Hastie [2014]）。

²³⁸ Kahneman（2011）p.417（訳下巻, p. 272）を参照。

治判断の研究で有名な Philip E. Tetlock による一連の研究によれば、判断のバイアスが相対的に大きい人は、自らの思想信条を重視して物事を断定的に判断し、自分の失敗を認めない傾向がある。他方で判断のバイアスが相対的に小さい人は、自らの意見に対立する見解を含む多様な情報を収集して、複数の分析手法を用いてできるだけ正確で精緻な判断をしようと努力し、失敗したときはしっかり原因を検証して改善しようとする強い意志を持つ²³⁹。このような判断能力の差は、生来の資質よりも、自分の意見を一つの仮説と捉えて批判的に検証するか否かによって主に決まる²⁴⁰。その証拠に、難易度が適切な優れた問いに集中する、問題を細分化してから頭を使って解く、統計的基準率と事実に基づくエピソードの両方に注意を払う、自らと対立する見解を考える、できるだけ精緻な予想をする、失敗したら原因を特定する、といった予測能力を高めるための 10 の心得を示すだけで、被験者の予測能力の正確性は統計的に有意に高まったという²⁴¹。

注意しなければならないのは、第一に、自分の認知バイアスに気づくのは他人の認知バイアスに気づくよりもずっと難しいし²⁴²、仮に自分の認知バイアスを自覚できてもセルフコントロールは容易ではなく、バイアスが緩和されないこともあるという事実である²⁴³。他方で組織は、有用なチェックリストの活用や科学的に信頼できるエビデンスの参照などを強制できるため、個人よりも認知バイアスによる判断ミスを防ぐ潜在的能力が大きいとされる²⁴⁴。認知バイアスは自覚や修正が難しいことを前提にして、判断ミスを防ぐために組織的に対処しなければならない。

第二に、個人の認知バイアスとそれによる判断ミスをラボ実験や予測コンテストや選挙等を用いて測定する試みは様々な方法で行われてきたが、個人の確証バイアスや集団浅慮が政策の成功確率に影響することを再現性の高い方法で証明するのは、必ずしも容易ではない。例えば Janis (1982) の研究は定性的な事例研究に基づいているので、失敗事例の失敗は集団浅慮による判断ミスのせいで、成功事例の成功は集団浅慮を避ける工夫のためであることが再現性の高い方法で証明されているわけではない。Esserf (1998) は、Janis 『集団浅慮の犠牲者 (*Victims of Groupthink*)』の 1972 年初版公刊以後の 25 年間に行わ

²³⁹ Tetlock (2005) を参照。

²⁴⁰ Tetlock and Gardner (2015) Chapter 5 を参照。

²⁴¹ 前掲書, Chapter 1 を参照。

²⁴² Pronin et al. (2002) を参照。

²⁴³ 自我消耗については、Baumeister, et al. (1998) 及び Baumeister et al. (2007) を参照。

²⁴⁴ Kahneman (2011) pp. 19-30 及び pp. 408-418 を参照。

れた集団浅慮理論に関する歴史研究とラボ研究をレビューし、Janis の集団浅慮理論は組織的意思決定理論の発展にきわめて大きな貢献をしてきたが、彼が挙げた集団浅慮の先行条件や集団浅慮の兆候とされる要因については様々な疑義が呈されており、仮説の修正が必要かもしれないと結論づけている。例えば、組織の構造的欠陥の説明力は強いようだが、集団の凝集性はそれほど強い説明力を持っておらず、状況的要因は説明力があるとは言い難いようであるとする実証研究が多い²⁴⁵。しかし研究によって仮説の検証方法が多様であるし、定量的な実験研究が少ないため、頑健な結果を導出するのは困難な状況であるという。したがって先行研究に基づいて講じた手段が実際に効果を発揮するかどうかは、現場でテストしなければならない仮説だと心得ることが肝要である。

3-3. 予想と現実のギャップを可視化して、バイアスの自覚を促す

では、どうすればよいのだろうか。まずはバイアスの自覚を組織的に促すことが必要である²⁴⁶。個人の認知バイアスや組織の集団浅慮が政策の判断ミスに繋がったという因果関係を証明することは容易ではないかもしれないが、政策の失敗につながりうるような判断ミスの実態を可視化することは必ずしも難しい作業ではない。地方自治体が行政評価の文脈で公共サービスの重要性や満足度に関する住民意識調査を行って住民と行政との認識ギャップの分析を行っている例は多いし²⁴⁷、同様の関係性について市民と行政職員の双方に意識調査を行ってその結果を計量的に解析する学術研究もある²⁴⁸。ただし地方自治体によ

²⁴⁵ 「組織の構造的欠陥」とは、①集団の孤立、②公平なリーダーシップの伝統の欠如、③方法論的な手順を要求する規範の欠如、④メンバーの社会的背景やイデオロギーの均質性等のこと。「集団の凝集性」とは、集団の団結力の強さのこと。「状況的要因」とは、①集団への期待が低く、リーダーの提案以上の策は出てこないだろうと考える外部の脅威からの強いストレス、及び、②一時的に引き起こされる低い自尊心のこと。これらは集団浅慮の先行条件とされる。なお、Janis (1982) p. 244 に、それらの先行条件が集団浅慮をもたらし、政策判断の失敗に至るまでのフローチャートが掲載されている。

²⁴⁶ 後藤 (2017b) は、基礎自治体における福祉情報の広報実態に関する実証研究に基づいて、自治体職員の認知バイアスが無自覚のうちに福祉における情報の壁を作ってしまう可能性と、そのような事態への組織的対処の必要性について検討している。

²⁴⁷ 例えば宮城県震災復興・企画部震災復興政策課「宮城県の行政評価制度の概要」p.12 (Retrieved July 10, 2019 from <https://www.pref.miyagi.jp/uploaded/attachment/727516.pdf>)。

²⁴⁸ 例えば Melkers and Thomas (1998) ; Poister and Thomas (2007) ; 加藤 (2009) ; 山

る施策満足度調査は通常、一般住民に対して、地方自治体が行っている数多くの粒度の異なる政策や施策に関する評価を求めていることが多い。そのため、測定バイアスが大きすぎて結果の信頼性が低いし、調査項目の粒度が施策や施設という大きな単位であるため、政策や資源配分の改善に直接結びつけにくいことが多い²⁴⁹。学術研究でもやはり同様の問題をもつものも多く、またそもそも学術研究の場合には政策改善を直接の目的としていないことが多いので、地方自治体はその成果を政策改善に直接役立てられるケースは多くない。

そこで筆者は、調査項目を具体的な公共サービスに限定することで測定バイアスを小さくした上で、住民の評価やニーズに関する行政の仮説を質問票調査の設計に反映することによって、行政と住民の認識ギャップを自治体職員が直接把握できるようにすることを目的とした調査を試みた²⁵⁰。調査対象は、関東地方における人口規模4万人程度の基礎自治体（以下、Y自治体という）である。Y自治体広報担当課は、役場の広報活動に関する大規模な住民意識調査を2015年度に郵送配布・郵送回収の形式で実施し、その結果を2016年度以降の広報関連施策に反映することを計画していた。Y自治体職員の仮説を検証して行政と住民の認識ギャップを明らかにすることを目的の一つとして調査を実施し、その結果を学術研究のために用いたいという要望を示したところY自治体広報担当課の合意を得られたため、Y自治体の広報活動に関する住民意識調査を受託調査のかたちで実施することにした。学術目的による調査データの利用や結果の公表に関しては明文で同意を得たが、個人名や団体名を特定できないように表現等を工夫している。

まず2015年3月から7月にかけてY自治体の広報活動に関する実態調査を行い、住民ニーズや住民満足度等に関するY自治体広報担当課の仮説を調査票に反映した。具体的には、同年3月にY自治体の広報活動の実態及び課題等について書面にて自由記述形式で回

岡（2015）。

²⁴⁹ 例えば「安心できる地域医療の確保」や「だれもが住みよい地域社会の構築」のように操作化の度合いの低い政策について一般住民に満足度を尋ねても、回答者によってその言葉から想起される政策には大きな違いがあるため、同じ政策の満足度を調査していることにならない可能性が高い。また、「防災」、「公園」、「ゴミ処理」、「文化施設」、「高齢者福祉」、「小学校」のように粒度が異なる公共サービスを多数羅列してそれらの満足度を一般住民に尋ねるタイプの調査も多いが、そのような調査は調査項目の並び順をランダムに入れ替えるだけで満足度の平均値や相対的な順位が大きく変わるほどに測定バイアスが大きいと考えられる。

²⁵⁰ 本項の記述は、後藤（2016）を大幅に加筆修正したものである。

答してもらい、その回答をさらに深掘りするため、同月末に Y 自治体広報担当課職員に対して半構造化集団面接調査を実施した。その後、同課の仮説と筆者の理解に齟齬がないかを書面と対面で確認し、同年 6 月末に調査票を完成させた。調査票は住民基本台帳より性別年齢階層別に層化無作為抽出した Y 自治体に住む成人男女 3,000 人に対して郵送で発送し、郵送で回収した。調査期間は同年 7 月から 8 月で、調査結果を得る前のタイミングで調査票の全質問に対する Y 自治体広報担当課としての仮説を書面で最終確認し、調査結果に基づいて住民ニーズに関する行政の予想と住民の回答とのギャップについて検討した。有効回収率は 45.3%で、比率の差の有意検定の結果、回収標本の性別年齢階層 6 区分別構成比は統計的に有意に母集団と差があったため、母集団の性別年齢階層別構成比を再現するようにデータの重みづけを行って集計した。以下では公式ホームページの利用者満足度に関する調査結果を用いて、行政と住民の認識ギャップについて述べる。

ホームページの利用者満足度調査は、まず Y 自治体の広報担当課職員に対してホームページの利用者満足度に影響する品質に関する予想を尋ねることから始めた。Y 自治体が広報に関する住民調査を実施するのは初めてのことで、比較可能な過去の調査は存在しなかった。そのため、品質別利用者満足度の数値や相対順位を尋ねるよりも大まかな予想を尋ねる方が職員の予想を正しく把握できると考えられたことから、ホームページの品質のうち低い利用者満足度が予想される品質と高い利用者満足度が予想される品質について尋ねた。Y 自治体は本調査を実施する約 1 年前に公式ホームページをリニューアルしたばかりで、「情報の分類方法」、「検索機能の使いやすさ」、「トップページのデザイン」の三点を特に重視してリニューアルしたとのことだった。しかし、リニューアル後のホームページの評判は必ずしも良くなく、特に「表示速度」が遅いという不満を聞くことが多いため、「表示速度」の利用者満足度が最低だという予想だった。表示速度の次に満足度が低いのは、「情報の見つけやすさ」、「説明文の統一感」、「検索結果の分かりやすさ」で、ホームページリニューアルで重視した項目は比較的満足度が高いのではないかとのことだった。

Y 自治体では、住民の情報ニーズを把握し、広報内容と広報ニーズのギャップをチェックしてより良い情報を届けるのは広報担当課以外の各課（以下「原課」という）の役割であり、広報担当課が原課の広報内容に関与することは基本的になかった。広報担当課が広報活動において果たすべき責務は、原課の発信するコンテンツを適切な方法で届けることであって、コンテンツの質や量を高めることではなかった。このような役割分担の在り方は、Y 自治体に特有のものではない。2012 年に日本都市センターが全国の全市区の広報担

当課を対象に行った質問票調査によれば、広報担当部署以外の部署が個別に行っている広報を全庁的に把握する仕組みや取組みが「ある」と回答した自治体は15.1%しかなく、「ない」と回答した自治体が84.1%にのぼっていた²⁵¹。かかる役割分担の下では、広報担当課がコンテンツの実質的な内容ではなく情報媒体の機能やコンテンツの形式により関心を持つのは自然である。したがって、「検索機能の使いやすさ」のようなホームページの機能や「説明文の統一感」のようなコンテンツの形式に関わる項目が広報課題と認識されていたのだと考えられる。

さて住民と行政の認識ギャップを明らかにするためには、行政は自覚できていないが住民は重要だと考えるであろう項目についても調査する必要がある。そこで調査の約1年前に行われた公式ホームページのリニューアルで重視された3つの項目（「情報の分類方法」、「検索機能の使いやすさ」、「トップページのデザイン」）、住民満足度が低いとY自治体広報担当課職員が予想した4つの項目（「表示速度」、「情報の見つけやすさ」、「説明文の統一感」、「検索結果の分かりやすさ」）に、ホームページの情報内容に関わる2項目（「コンテンツの充実度」、「文章の分かりやすさ」）と「総合的満足度」を加えた合計10項目について、ホームページの利用者満足度を調べた²⁵²。ホームページの利用者満足度は、過去1年以内にY自治体の公式ホームページにアクセスしたことがある379人に対して満足度を「満足」、「やや満足」、「どちらともいえない」、「やや不満」、「不満」という5件法で尋ねた。

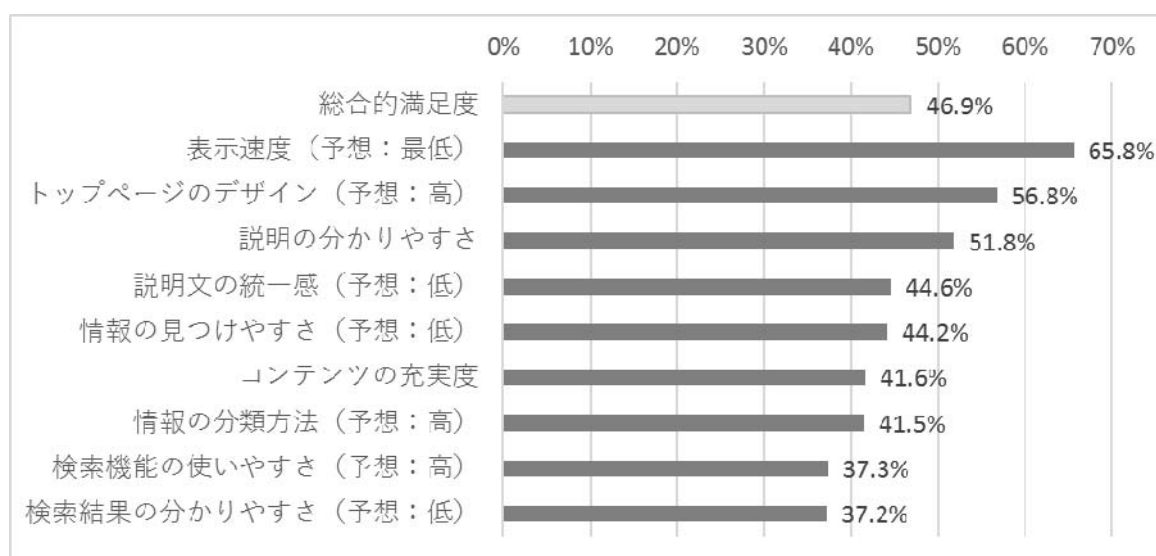
その結果を図3-4に示した。公式ホームページのリニューアルで重視された3つの項目は、高い利用者満足度が予想されていたので項目の後ろに「(予想：高)」と記した。Y自治体広報担当課職員が低い利用者満足度を予想した4つの項目には後ろに「(予想：低)」または「(予想：最低)」と記した。ホームページの総合的満足度は50%弱で、事前に尋ねた広報担当課としての予想値は45%だったので、ほぼ行政の予想通りであった。しかし第一に、ホームページのリニューアル時に重視して改善した3つの項目のうち、「トップページのデザイン」の利用者満足度は比較的高かったものの、「情報の分類方法」及び「検索機

²⁵¹ 日本都市センター編(2013)。同調査の有効回収標本サイズは478、有効回収率59.0%。

²⁵² ホームページの機能に関する項目数と、ホームページのコンテンツに関する項目数を揃えた。前者は「検索機能の使いやすさ」、「検索結果の分かりやすさ」、「表示速度」、「トップページのデザイン」の4項目、後者は「情報の分類方法」、「説明文の統一感」、「コンテンツの充実度」、「文章の分かりやすさ」の4項目である。「情報の見つけやすさ」は「総合的満足度」と同じく、複数の要因の結果となる変数とみなした。

能の使いやすさ」の利用者満足度は行政の予想に反して低い方だった。第二に、Y自治体広報担当課職員によって利用者満足度が低いと予想された4つの項目のうち、「検索結果の分かりやすさ」の利用者満足度は行政の予想どおり低かったが、利用者満足度が最低だと行政が予想した「表示速度」の利用者満足度は最も高く、行政が利用者満足度は低いと予想していた「説明文の統一感」や「情報の見つけやすさ」の利用者満足度は中程度であった。

(図3-4) Y自治体ホームページの品質別利用者満足度 (n=379)



(注) 数値は「満足」、「やや満足」と回答した人が有効回答に占める割合。各項目の () 内は Y自治体広報担当課の予想利用者満足度水準を表す。

調査結果から、どのような示唆を得ることができるだろうか。第一に、職員による住民満足度予想の的中率が高くなかったという事実は、勘や思い込みに基づくサービス改善の危険性を示唆している。今回、Y自治体の広報担当課職員は「表示速度」の利用者満足度が最も低いと予想していたが、その理由はたまたま耳にすることの多かったクレームを代表的な住民評価だと思い込んでしまったからであった。これは、利用可能性ヒューリスティックに起因する確証バイアスによって利用者満足度を読み間違えてしまった典型例ではないかと推察される²⁵³。広報担当課職員が各課の所掌するコンテンツの改善ではなく自ら

²⁵³ 利用可能性ヒューリスティックとは、容易く思い浮かぶかどうかによってある事象の確率を見積もってしまうという判断方法のこと。利用可能性ヒューリスティックについて

の担当課の所掌する広報媒体の機能改善を重視していたという事実も、職員が「自分の見たものが全て」という認知バイアスに陥っていた可能性を示唆する。確証バイアスのような認知バイアスは避け難いものなので、勘や経験だけに頼らずに、標的集団からの多様なフィードバックを得て判断することが重要だと言えるだろう。

第二に、ホームページのリニューアル時に重点的に改善したはずの機能に対する利用者満足度が必ずしも高くなかったという事実は、サービスイン前の設計開発段階で利用者を巻き込み、協働してサービスを改善する必要性を示唆していると考えられる。政策は仮説である。仮に理想的なニーズ評価と信頼性の高いエビデンスに基づいて公共サービスを開発したとしても、その公共サービスが本当に標的集団のニーズに合致しているかは、実際に試してみなければ分からない。一般にホームページの大幅リニューアルを頻繁に行うことは難しいから、サービスイン前に利用者を巻き込んでテストを行うことが肝要であろう

²⁵⁴。

本調査にはいくつかの限界がある。例えば本調査の調査票設計では、筆者の仮説が行政の仮説や住民のニーズを歪めている可能性がある。全質問に対する行政の仮説を明文化して広報担当課職員に書面と対面で確認してもらったので、行政の仮説の反映という点では大きな問題は生じていないと考えられるが、住民ニーズの反映という点では筆者のバイアスのリスクが残る。また、本研究の手法に汎用的な有効性があるのか、今回の調査が住民目線の行政サービス改革に実際に役立てられるかを具体的に確認するには至っていない。調査分析手法を磨き上げて調査の費用対効果を高めながら、バイアスの自覚が判断ミスの是正につながるのかを実証することが今後の大きな研究課題である。だが一方で、職員の仮説を調査票に反映して仮説検証型で住民調査を行えば、行政と住民の認識ギャップを単純集計結果から簡単に把握できるので、統計のトレーニングを受けていない行政職員でも調査結果を自分で分析評価できるというメリットを示すことはできたのではないかと思う。利用者へのフォーカスグループ・インタビューや後述する Wiki 調査のようなオープンエンド型調査を用いれば、評価者のバイアスをより小さくすることができる。そのような工夫を講じた上で定量調査を実施すれば、より歪みのないフィードバックに基づいてサービ

は、Tversky and Kahneman (1973) ; Tversky and Kahneman (1974) ; Kahneman (2011) Chapter 12 and 13 を参照。

²⁵⁴ Plan と Do の間に学習や実験のプロセスを設定して素早く学習する仕組みについては、第4章第3節で述べる。英国政府が顧客を巻き込みながら政府ポータルの開発を進めた事例については、後藤 (2012) を参照されたい。

スを改善することができるだろう。なお、後付けバイアスを防ぐために、実査の前に仮説の記録を残す等の工夫が必要なことを忘れてはならない。

本項ではこれまで、行政による住民ニーズの予想と現実の住民ニーズとの認識ギャップについて筆者が実施した調査の結果を述べてきた。行政の予想と現実とのギャップを測る方法は、他にもある。例えば 2005 年に米国の行政管理予算局（Office of Management and Budget：OMB）が 47 の規制影響評価について事前評価と事後評価とのギャップを調べたところ、米連邦政府の各省庁は効果も費用も過大評価する傾向があるが、費用よりも効果を過大評価する傾向があること、規制の費用対効果比について省庁の予想が正しかったのは 23%、過大だったのは 47%、過小だったのは 30%で、予想が正しい場合の方が少なく、一定の方向に誤るわけではないこと等が明らかにされている²⁵⁵。

行政による住民ニーズの予想と現実の住民ニーズとの認識ギャップも、規制の費用対効果に関する事前評価と事後評価のギャップも、測定評価して可視化しなければ、ギャップの有無や方向性を自覚して改善することはできない。自分の認知バイアスに気づくのは他人の認知バイアスに気づくよりも難しいし、第 4 章で述べるように人間にはミスを認めたがらないという性質があるから、予想と現実とのギャップの分析評価は政策担当課以外の第三者が行う必要がある。第三者評価を可能にするような組織的仕組みを早急に整えて予想と現実のギャップを可視化し、フィードバックを踏まえた改善を図ることが望まれる。

3-4. 認知バイアスに組織的に対処する

バイアスの自覚が行動変容を引き起こすとは限らない。だが少なくとも、バイアスによる判断ミスを防ぐための対策が必要だという気づきを促すことにはなるだろう。では、バイアスを自覚した個人や組織は、認知バイアスにどのように対処すればよいのだろうか。

バイアスの自覚が意識や行動を変える効果は持続しないという問題には、チェックリストの活用が有効だと考えられる。エビデンスに基づく医療では判断ミスを防ぐためにチェックリストが多用されており²⁵⁶、エビデンスに基づく医療のテキストにも別添でハンディ

²⁵⁵ Sunstein (2013) Chapter 8 を参照。米行政管理予算局によって過大でも過小でもない正確な事前評価がなされたとみなされる範囲は、事後評価の指標値との差が±25%以内に収まっている場合である。

²⁵⁶ 医療ミスを防ぐためのチェックリストの有用性に関する代表的な研究は、Peter Pronovost らによる、中心静脈カテーテル関連の感染症を防止するためのチェックリスト

なチェックリストが付いている²⁵⁷。公共政策についても、エビデンスに基づくインパクト評価の実践に役立つチェックリストを開発し、その有効性を検証して改善しながら適用できる政策プロセスや政策領域を広げていくことが望まれる。

Gawande (2009)によれば、予想外の問題が生じたとき、病院では伝統的に専門家に権限を与えて彼らの判断に任せるという方法がとられてきた。だが、専門家の判断に任せるという方法では、現代の複雑な高層建築物を建てることはできない。そのため建設業界は棟梁の判断に任せるというかつての伝統的方法を放棄し、建設工事の工程表に事前に話し合いの予定を入れ、専門家たちに確実にコミュニケーションを取らせることで不測の事態に対応するという方法を採用している。単純ミスを防ぐためのチェックリストとコミュニケーションを促すためのチェックリストを併用するという方法を用いれば、マニュアル化された単純業務のミスを減らせるだけではなく、予想できない複雑で難しい問題にも対処できることが建設業界では証明されてきた。

緊急手術を行う外科医と同じようにミスが人命に直結し、時に一分一秒を争う判断が求められるパイロットの業務では、1 ページに大きな見やすい文字でシンプルかつ明確に書かれた数行のチェックリストがいくつも用意され、平時と非常時に活用されているという²⁵⁸。どのチェックリストも使いやすさを重視して入念に作られており、いつチェックを行うかという一時停止点 (pose point) が明確で、一つのチェックには基本的に1分ないし1分半足らずの時間しかかからないようにすることで手順が省略されないように工夫されている。それでもチェックリストが実用に耐えられるかは、実際に試してみなければ分からない。そのため、パイロットがフライトシュミレーターでチェックリストをどのように使うのかを何時間も観測して、無駄を徹底的に削ぎ落とした上で実用に供されている。

航空安全対策の優れた点は、各航空会社が慎重な模擬実験を重ねて、現場で素早く使えるチェックリストを作成していることだけではない。航空事故から得られる教訓を航空業

を用いた一連の研究である。その研究の経緯と広がりについては、Pronovost (2010)に詳しい。

²⁵⁷ 例えば Sackett et al. (2000) には、予後や診断に関する研究結果の妥当性に関するチェックリストや系統的レビューの結果の信頼性に関するチェックリスト等が縦 10 cm、横 16.5 cm 四方のラミネート加工のカラフルなカードにまとめられ、巻末に添付されている。Guyatt and Rennie eds. (2003) には、治療に関する論文をレビューするときのチェックリストや、診断検査結果を解釈するためのノモグラムの読み方解説などが縦 12.5 cm、横 7.5 cm 四方のコンパクトなカードに要約されている。

²⁵⁸ Gawande (2009) Chapter 6 を参照。

界全体で学び、新しい手順に反映して迅速に国内外に普及させる仕組みが確立されていることも特筆に値する。Syed (2015)によれば、航空事故が起こるとすぐに航空会社とは独立した調査機関や監督行政機関が事故機の残骸やその他様々な証拠をくまなく調査して報告書を一般公開することとされており、航空会社にはその報告書に記載されている勧告を履行する責任が発生するのだという。パイロットに対するエラーマネジメント教育も徹底しており、人間の記憶力と判断は不確かで、その事実を認識して定められた手順に従わないと死者が出ることをパイロットは徹底的に学ぶそうだ²⁵⁹。

医師でありライターでもある Atul Gawande らが主導した国際保健機構 (World Health Organization : WHO) の国際プロジェクトでは、それらの取り組みからヒントを得て作成したチェックリストが手術時の医療ミスの削減に与える影響を確かめるべく、実証実験が行われた。手術のリスクを引き下げするために、手術の一時停止点ごとに単純な手順が飛ばされることを防ぐためのチェック項目とチームワークの向上を図るためのコミュニケーションのチェック項目を入れたチェックリストが作成され、世界各国の 8 つの病院で実験が行われた。3 か月間で 4000 人近い患者のデータを集めて分析した結果、深刻な合併症の発症率は 36% 下がり、死亡率は 47% 下がり、感染症は半分になり、再手術の確率は 75% になるという大きな成果が得られたという²⁶⁰。

集団浅慮のような確証バイアスによる判断ミスを避けるためには、集団のもつ多様性という利点を大いに生かし、集合知を引き出すためのツールを用いるという対処方法もありうる。例えば「All Our Ideas」というウェブサイトを使えば、政策問題、目標（または期待されるアウトカム）、あるいは政策手段について、群衆（クラウド）の力を利用して新しいアイデアを取り入れながら優先順位を付けることができる²⁶¹。具体的には、「All Our Ideas」の Wiki 調査 (Wiki surveys) に回答者として参加すると、ある質問に対して 2 つの

²⁵⁹ Syed (2015) Chapter 1 を参照。

²⁶⁰ Haynes et al. (2009) ; Gawande (2009) Chapter 7 を参照。

²⁶¹ 「All Our Ideas」はプリンストン大学に拠点を置く、社会データを収集するための新しい方法を構築することを目的とした研究プロジェクトである。「All Our Ideas」のサイトでは、簡単に新しい Wiki 調査をつくったり集計結果をビジュアル化することができる。コードはオープンソースとして公開されている (Retrieved July 26, 2019, from <https://www.allourideas.org/>)。Wiki 調査の設計思想は、Salganik and Levy (2015) に詳しい。国連の持続可能な開発部門も、持続可能な開発目標 (Sustainable Development Goals: SDGs) に関する人々の優先順位を把握するために、「All Our Ideas」の Wiki 調査を用いているという (Sunstein and Hastie [2014] Chapter 7 を参照)。

具体的な選択肢を与えられ、それらの選択肢を選ぶか、「決められない」という選択肢を選ぶか、あるいは自分のオリジナルな選択肢を追加することができる。次の回答では、基本的に人々に選ばれる確率が高い順に選択肢が表示されていく。選択肢の数と選択肢の優先順位が参加者からのフィードバックに基づいて時間とともに進化するため、伝統的な質問票調査よりも効率的に多くの人々の声を反映して、より民主的に政策の優先順位を決めることができる。

指標の優先順位づけを市民に委ね、選択式と自由回答形式の質問項目の併用によって調査主体のバイアスを排除するというアイディアは、古くから提案されてきた²⁶²。そのアイディアに ICT を活用して、調査客体による指標提案と指標の優先順位づけをより低コストでより多くの人々に開かれた方法で実現したのが、Wiki 調査である。Wiki 調査のようなオープンエンド型調査を用いれば、大量のデータを収集してそれをより簡単に分析できるようになるため、ビッグデータの特性を生かした事後層化による予測精度の向上も容易になると期待される²⁶³。前述したように保健医療分野ではエビデンスをつくり、つたえ、つかうために ICT が積極的に有効活用されている。例えば GRADE システムでは、アウトカムの相対的重要性や推奨の質について修正デルファイ法に準じた方法で合意形成する際にも ICT ツールを用いることができる。既に世界には、憲法改正案の作成さえもクラウドソーシングの手法を取り入れて透明性と開放性の高い方法で行っている国がある²⁶⁴。公共政策においても、データやエビデンスの収集、政策選択、政策評価などの場面でもっと ICT を有効活用し、政策判断のバイアスを小さくすると共に、より開放的で民主的な政策運営を行うべきではないだろうか。

²⁶² Schuman and Presser (1979) ; Salganik (2018) pp. 111-115 を参照。

²⁶³ Salganik (2018) Chapter 3 を参照。

²⁶⁴ スタックラー (2014) 第 4 章 ; レイブルック (2019) 第 4 章。集団浅慮を避けながらより民主的な意思決定を促す方法には、討論型世論調査 (deliberative poll) のような熟議民主主義 (deliberative democracy) の実践手法を取り入れるという方法もある (フィッシュキン [2011])。「討論型世論調査」とは、代表性を持つことが期待される無作為抽出によって選ばれた市民が、対立する意見を吟味した上で集団浅慮を避けてより良い意思決定ができるように、議論に参加しない専門家集団の支援やモニタリングのもとで小グループに分けて熟議し、熟議前後の無記名アンケートによって意見分布を把握する方法のことである。「熟議民主主義」とは、政治的に同等な立場にある人々の間でのある種の話し合いによって、様々な問題について真剣に考えた市民全体の意見を意思決定に反映しようとする考え方のことである。

4. 政策プロセスの透明性と開放性の大切さ

バイアスに系統的に対処するために組織的な工夫を凝らすほど、注意しなければならないことがある。それは、判断や意思決定を改善するために講じたはずの手段がいつの間にか目的化してしまうという事態である。例えば開発経済の分野では、世界銀行や国際連合のような国際機関がランダム化比較試験に基づくエビデンスを大量に生産していることもあり、ランダム化比較試験に基づくエビデンスは短期間でグローバルな規模で飛躍的に増えているが、エビデンスの利用は必ずしも進んでいないという²⁶⁵。もしランダム化比較試験によるインパクト評価を重視する学界や国際援助機関の潮流が、効果の高い政策を実践することよりも途上国で社会実験を行って新しいエビデンスをつくることを目標にさせてしまっているとすれば、憂うべき事態である。

目的の倒錯という現象は、まさに日本の地方自治体による政策評価でみられる。田中（2014）によれば、日本の自治体評価の全般的特徴の一つは、事務事業を主たる対象とした目標管理型の実績評価を制度化して、毎年一回程度の頻度で、専ら内部評価によって全庁的、継続的、画一的に制度を運用していることである。全ての自治体評価が形骸化しているわけではないであろうが、少なくない自治体における政策評価は有効活用されているとはいえない状況にあるようだ²⁶⁶。政策の効果効率を高め、市民への説明責任を果たすために導入されたはずの制度が、一年に一度政策評価書を作成するという恒例行事となってしまうのであれば、評価制度を早急に見直さなければならない²⁶⁷。

自治体評価が形骸化している理由が無知（ignorance）のせいであれば、本章でこれまで述べてきたことに留意してインパクト評価を行うことで改善できるであろう。また、惰性（inertia）のせいであれば、第4章で考察する仕組みを取り入れることで状況は改善されるだろう。では自浄作用の不足には、どのように対処すればよいだろうか。前節では意図

²⁶⁵ Cameron et al. (2015)；浅岡(2016)を参照。

²⁶⁶ 山谷（2008）。ただし序章で述べたように、「援助は役に立っているか」や「自治体経営の効果効率は高まっているか」という抽象的で大きな問いに科学的に答えることが難しいのと同じように、「自治体評価は役に立っているか」という問いに科学的な方法で答えることは難しいと考えられる。

²⁶⁷ 評価制度の形骸化は、地方自治体だけではなく中央政府でもみられる。中央政府の政策評価の多くが政策の正当化のために行われているという指摘は、大橋ほか（2011）を参照。事業の目標と実績とのズレを示す指標は無益な事業の際限ない拡大を正当化しかねないという指摘については、西尾（1990）第7章、特に pp. 285-286 を参照。

せざる結果として都合の良いエビデンスだけがつまみ食いされる事態にどのように対処するかを論じたが、イデオロギー (ideology) によって最良のエビデンスが意図的に無視されることも十分にありうる²⁶⁸。

自浄作用の不足という問題に対しては、政策プロセスの透明性と開放性を高めて、外部評価をもっと強化するしかないのではないかと思われる。その取組みを現場発でボトムアップに進めるのは不可能だと思うかもしれない。しかし一国のレベルで見ればトップダウンの事例ではあるものの、外発的ではなく内発的に、外部評価を有効活用して、政治的利害によるエビデンスの無視という事態を免れた世界的に注目されている事例がある。それは、Ernesto Zedillo Ponce de León 元メキシコ大統領の下で開始された「プログレッサ計画 (Program for Health and Education and Nutrition : PROGRESA)」である²⁶⁹。

プログレッサ計画とは、子どもの健康や教育について責任ある親になることを条件に、農村部の貧困世帯の母親にかなりの額の現金を支給するというプログラムである²⁷⁰。この計画は、貧困対策として現物給付ではなく条件付き現金給付を行った点に加えて、政策プロセスに科学的手法を大胆に取り入れ、エビデンスに基づく政策をランダム化比較試験によって小規模で試行して効果を確認した上で全国展開するという手法を採用したという点で、画期的な政策だった。計画を開始した 1997 年当時は小規模だが統計的に有意な結果を得ることのできる規模で介入群と対照群がランダムに割り当てられ、ベースライン調査が行われた。その後 2000 年までに、メキシコの全世帯の約十分の一、地方の貧困世帯の

²⁶⁸ Banerjee and Duflo (2011) は、「イデオロギー (ideology)」、「無知 (ignorance)」、「惰性 (inertia)」という「3つの I」を克服して、人々の意思決定方法や政策の細部を見過ごさず、実験を恐れずに、データとエビデンスに基づいて効果が確かな開発援助政策を講じるべきだと主張している。この段落では、彼らのフレームワークを参考にして、形骸化した自治体評価が維持されてしまう理由を整理した。

²⁶⁹ プログレッサ計画はその後「Oportunidades」、さらには「Prospera」と名称変更されて継続している。

²⁷⁰ プログレッサ計画で現金給付を受け取るためには、居住地や所得水準等のプログラムの適格性基準を満たしていることに加えて、全世帯員が健康診断と予防医療を受ける、乳幼児と授乳中の母親と妊娠中の女性はクリニックに通って健康診断を受け、栄養補助食品をもらい、栄養と公衆衛生に関する教育を受ける、といった条件を守らなければならなかった。条件をクリアした世帯の母親には、世帯収入の約 2 ～ 3 割に当たる現金が原則として三年間支給された。父親ではなく母親に支給されたのは、父親よりも母親の方が子どもに対する責任感が強く、子どものためにお金を使うと期待されたからである。小学校 3 年生から中学 3 年生までの子どもには奨学金も給付された (Gertler [2004] を参照)。

約4割をカバーする規模まで拡大し、公募で決められた外部の第三者によって、標的集団のターゲティングが正確で濫給や漏給はないか、プログラムは就学率や病気の発生率や身長等のアウトカムに対して統計的に有意な効果をもたらしたか等が評価された。フォローアップ調査も次々に行われ、メキシコ政府は調査結果を踏まえてすぐに軌道修正しながら、結果を積極的に公開していった²⁷¹。

Zedillo 元メキシコ大統領が目指したのは、確かな効果のある貧困対策を実施して、学校に通える健康な子供たちを増やして将来その子供たちが自分で働いて稼げるようにすることだけではなく、大統領が交代する度に前政権の政策を否定し、新しい政策を講じるというメキシコの悪しき慣行を断ち、次の大統領に引き継がれる貧困対策を講じることも、プログレッサ計画の重要な目標だった。この政治的な意味におけるプログレッサ計画の成功は、政策プロセスの透明性を高め、政策効果を外部の第三者が肯定的に評価したことでもたらされた。プログレッサ計画は、議会がその結果を無視できなくなったために大統領が交代しても規模を拡大しながら引き継がれていっただけではなく、貧困対策として国外にも広がり、ラテンアメリカ地域における標準的な政策メニューになった。ことほどさように、エビデンスを用いた政策形成を実践し、その政策プロセスの透明性・開放性を高めることは、エビデンスを無視したいという誘惑をもつ政治家や権力者に打ち克つ力を人々に与えるのである。

大きな変化は、小さな変化の積み重ねによって生じる。コ克蘭やプログレッサ計画のグローバルな広がり、個人の感情や党派的利害を超えて公益の増進を図ろうとする人々が、当事者目線の目標を掲げ、多くの人々を巻き込みながら透明性の高い方法でエビデンスに基づく政策を実践することによって、社会をより善い方向へと導くようなグローバルな変化がボトムアップ型で生じうることを鮮やかに示しているのではないだろうか。

5. 小括

本章では、地方自治体が目標に向かって正しい方法でやり抜くためにどうすればよいか

²⁷¹ PROGRESA 計画ないし Oportunidades 計画全般については Levy (2006) を、ターゲティングについては Skoufias et al. (2001) を、就学率については Schultz (2004) を、健康や栄養状態のアウトカムについては Behrman (2010) を、長期的な能力に与える影響については Fernald et al. (2009) を、外部評価については浜口・高橋 (2008) を参照した。

という問いについて、エビデンスのバイアスと認知のバイアスに注目して考察してきた。エビデンスに基づく医療を推進するために、コクランは良質なエビデンスをつくり、つたえる活動をグローバルに行っている。その取組みは、世界中の数多くの医療関係者等と協働しながら、透明性と再現性の高いプロセスに基づいて行われていた。また診療ガイドラインの作成等に用いられる GRADE システムでは、エビデンスの確実性の評価と推奨の強さの判断が切り離され、透明かつ再現性の高い方法に基づいて確実性の評価を行い、関係者間の合意形成の中立性を重視したプロセスによって推奨の程度を決定することが要求されていた。エビデンスに基づく医療は、良質で統合的な研究エビデンスやそれを活かした実践の方法論をつくり、つたえる活動に支えられて急速に発展し、勘や非系統的な臨床経験、病態生理学的な推論、そして権威ではなく、再現性の高い方法で批判的に吟味したエビデンスこそが重要だという方向へと、医療のパラダイムシフトをもたらしている²⁷²。

公共政策においては、良質なエビデンスにすぐアクセスできる環境が十分には整っていない。そのためエビデンスを「つかう」段階では、特に政策の根拠に用いたエビデンスのスクリーニング・プロセスと評価結果を公表し、第三者による妥当性検証を容易にすることが重要ではないかと考えられる。エビデンスを「つくる」段階では、反事実と比較すること、そのために、倫理的・経済的な問題の小さい分野で小さな対照実験を試してみることが推奨される。エビデンスを「つたえる」段階では、政策プロセスの透明性と開放性を高めて多様な人々と協働することによって、良質なエビデンスが増え、政策の効果効率や公平性が高まるのではないかと期待される。

非系統的な臨床観察がランダム化比較試験の系統的レビューによって統合されたエビデンスに劣る理由は、データのバイアスだけではなく、認知バイアスによる判断ミスの恐れが大きいためであった。集団浅慮のような認知バイアスによる判断ミスを防ぐためには、まず、バイアスの自覚を組織的に促す必要がある。人は自らのバイアスを自覚しにくいものなので、第三者評価を可能にする仕組みを整えて行政の予想と現実のギャップを可視化する等の方法を用いて、バイアスの自覚を組織的に促す必要があるだろう。チェックリストの活用やクラウドの力を用いた意思決定の促進という方策も、認知バイアスによる判断ミスを防ぐために有効であろう。

自浄作用の不足という問題に対しては、政策プロセスの透明性と開放性を高めて、外部評価を強化することが求められよう。それが現場発のボトムアップ型で実現すると期待す

²⁷² Guyatt and Rennie eds. [2003] p. 5 を参照。

るのは楽観的過ぎるかもしれないが、メキシコのプログレッサ計画やコ克兰の活動のグローバルな広がり、公益に資する取組みが自発的・内発的に生じる可能性を示唆している。確かな効果が期待される仕組みを積極的に取り入れて小さな実践を積み重ねることによって、政策に関係する人々の相互理解が広がり、公共サービスのイノベーションが次々と生まれるような大きなムーブメントに繋がることが望まれる。

第4章 目標を達成するまでやり抜く組織の構築方法²⁷³

システム工学の専門家で「学習する組織 (learning organization)」という概念を提唱した Peter Senge によれば、何らかの偉業を成し遂げた組織で、目標や価値観や使命が組織全体に浸透していない例を挙げるのは難しい²⁷⁴。だが一方で、良い目標や価値観や使命があれば成功が保証されるわけではない。明確で説得力のある目標を掲げ、業界で卓越した地位を築いて尊敬を集めてきた企業でも、崇高な目的を達成した結果として生じたものであったはずの利益の拡大をいつしか追求するようになり、衰退してしまう例はいくつも挙げることができる²⁷⁵。

しかも、チャレンジには抵抗や失敗がつきものである。確かなエビデンスに基づいて政策を設計しても、それが政策現場で期待通りに機能するとは限らない。どれほど未然に防ごうとしても、失敗は避けられるものではない。よって組織が良い目標を達成するまでやり抜くためには、失敗情報を組織で共有し、失敗から学ぶことが不可欠である。しかし行政の無謬性という言葉があるように、行政は過ちを犯さないことを前提にしてしまうと、失敗を認めて失敗から学ぶことができなくなってしまう。そうすると、改革よりも前例が優先されたり、失敗がひた隠しにされて同じ過ちが何度も繰り返されてしまうことになるだろう。ではどうすれば、目標達成までの過程で生じるであろう様々な困難を乗り越え、目標を達成するまでやり遂げる組織を作ることができるだろうか。

本章では、主にポジティブ心理学²⁷⁶及び組織科学分野の文献研究に基づいて、失敗から系統的に学び、目標を達成するまでやり抜く組織を作るためにはどうすればよいかを検討する。Senge の言うように、学習する個人がいなければ、学習する組織はありえない²⁷⁷。よってまず、失敗しても目標を達成するまでやり抜く人の特徴を、ポジティブ心理学の研究成果に依拠しながら明らかにする。次に、失敗から学ぶ「能力」を組織的に高める方法と、失敗から学ぼうとする「意欲」を組織的に高める方法について考察する。最後に本稿

²⁷³ 本章は後藤 (2019c) を加筆修正したものである。

²⁷⁴ センゲ (1995) pp. 17-18 を参照。

²⁷⁵ Collins (1994) ; 同 (2009) を参照。

²⁷⁶ 精神疾患を治すためではなく、潜在能力を高め、人生を豊かにすることに焦点をおく心理学研究は「ポジティブ心理学」と呼ばれ、特に 21 世紀に入って盛んに研究されるようになった。その経緯については、Seligman and Csikszentmihalyi (2000) を参照。

²⁷⁷ センゲ (1995) p. 165 を参照。

の考察結果を要約する。

1. 成功する人の特徴

成功する人の特徴を明らかにするためには、先に「人生の成功とは何か」を定義しなければならない。スタンフォード大学で心理学を教える Carol S. Dweck によれば、まさに人生の成功をどのように定義するかが人生の成功確率を決める。Dweck (2006) は、成功と失敗に関する 2 つの相反する考え方を「成長思考 (growth mindset)」と「固定思考 (fixed mindset)」と称して比較している²⁷⁸。成長思考は、「人は変われる」という信念に支えられている。成長思考をもつ人々にとって、成功も失敗も一時的な出来事であり、自分の人生を運命づけるものではない。彼らにとって正解することではなく学習し成長することこそが成功であり、壁にぶつかっても試練に立ち向かい努力を積み重ねれば壁を乗り越えられる可能性が高まると信じている。だから、うまくできる自信がなくても失敗を恐れず、楽な気持ちで挑戦し続けることができる。

それに対して固定思考をもつ人は、人の能力は固定的で変わらないと思い込んでいる。固定思考を持つ人たちにとって、成功とはまさに正解して勝負に勝つことであり、失敗は自らの無能さを曝け出す恐怖である。だから失敗を避けるために難しいことにはチャレンジせず、簡単に成功できる楽な道を選ぶ。それでも失敗してしまったときには、教訓を得て次に生かすことではなく、自分のせいではないと言い訳したり、自分より劣る人を参照して自尊心を充たすことを好む。その結果、学習して成長するチャンスを失い、潜在能力を十分に発揮できなくなってしまう。このようにして、成長思考を持つか固定思考を持つかによって、人生の成功確率は大きく変わっていく。失敗を学習のチャンスと捉えることができる人は、失敗を敗北や挫折と捉える人よりも、自分の潜在能力を発展させて、結果的により高い成果を達成することができると Dweck (2006) は指摘する。

²⁷⁸ 成長思考と固定思考に関する Dweck の初期の研究は、Diener and Dweck (1978) 及び (1980) を参照。才能ではなくマインドセットこそが人生の成功のカギを握っていることに Dweck が気づいたのは、学童期の子どもたちに難しいパズルを解かせて、人は失敗したときにどうするかを研究していた頃であったという。そこには、パズルが解けないという躓きを失敗とみなすのではなく学習するチャンスと捉えて楽しむ子どもたちがいた。マインドセットがしなやかな子どもたちにとって、成功とはパズルが解けることではなく、自分のベストを尽くして学習し成長することだったのである。

人生の成功を決めるのは才能というよりマインドセットであるという事実は、以前から他の研究でも支持されてきた。1960年代から「学習性無力感 (learned helplessness)」の研究に取り組んできた Martin E. Seligman は、楽観的であるか悲観的であるかというものの考え方がメンタルヘルスや寿命、学業や仕事、スポーツなどの業績を決定することを、様々な実験で示している。Seligman (1990) のいう楽観主義者とは、悪いことがあってもそれは一時的なことで、自分のせいではなく、環境や運や他者のせいだと考える人のことである。逆に悲観主義者とは、悪いことはいつまでも続き、その原因はすべて自分にあると考える人のことである。楽観主義者が悲観主義者よりも成功するのは、壁にぶつかっても、たとえ失敗しても、簡単には諦めず粘り強く挑戦し続ける傾向があるからだ。例えばメットライフ (Met Life) という米国最大の生命保険会社の保険外交員 1 万 5,000 人を 2 年間追跡した調査によれば、楽観主義者は悲観主義者と比べて売上が多く、その差は 1 年後よりも 2 年後の方が顕著だった。どれほど優秀な保険外交員でも、一件の新規顧客を獲得するために、その何倍もの件数だけ断られる経験をする。そのため時が経てば経つほど、楽観主義か悲観主義かという考え方の違いが成功を説明する要因として重要になってきたのではないかと Seligman は推察している²⁷⁹。

では、どうすれば楽観主義になることができるのだろうか。Angela L. Duckworth によれば、まずは才能や知性ではなく成長思考が人生で成功できるかどうかを決めるという風に考え方を改め、次に物事を楽観的に捉えられるようにトレーニングすれば、逆境でも粘り強くやり抜く力が備わる。そうして得られる「粘り強さ (perseverance)」は、分野を問わず大きな成功を収めた人たちに共通する「グリット (やりぬく力: grit)」を構成する二大要素の 1 つである。グリットを構成するもう 1 つの要素は、「情熱 (passion)」である。それは、何かに夢中になって熱心に取り組むことではなく、自分が何を求めているのかをよく理解して方向性を定め、ひとつのことに専念して長い間取り組み続けることを指す。情熱と粘り強さは重なりを持つが異なる概念で、何かを諦めずに粘り強く取り組み続けることよりも、専心すべきたった一つの目標を見つけて、その達成を追い求め続けることの方が難しいようだ²⁸⁰。

ものの考え方や努力が人生の成功の全てを説明するわけではないが、IQ (知能指数) で測られる知能と人生の成功との関係性はそれほど強くないとする同一集団の長期追跡調査

²⁷⁹ Seligman (1990) Chapter 6 を参照。

²⁸⁰ Duckworth et al. (2007) ; Duckworth (2016) を参照。

研究結果や²⁸¹、生まれつきの資質が業績に影響するのは初期段階だけで、一定のスキルレベルに達したトッププレイヤーの成功と遺伝的能力との関係性を示すエビデンスは殆どないという心理学のレビュー論文²⁸²が示すのは、努力の大切さである。生まれつきの才能はスキル形成の初期段階では業績に一定の影響を与えるかもしれないが、どれほど才能に恵まれていても努力しなければスキルを高めることはできないし、スキルがなければ業績を上げることはできない。しかも努力は、スキルを加法的に高めるのではなく、乗法的に高める。様々な時代の多様な分野で、一握りの卓越した人たちがその他大勢を遥かに凌駕する業績を量産する現象が観察されるのは、努力が乗数効果をもつからである²⁸³。長期的な成功に対して強い説明力をもつのは、生まれつきの才能よりも努力なのである。

では、一つのことに打ち込んで努力し続ければ、いつか成功できるのだろうか。卓越した成果を出す人々について長年研究してきた Anders K. Ericsson は、単に努力するのではなく、改善に向けた意識的な努力を積んではじめて傑出したプレイヤーになれると指摘する。確かに、やればできると信じて長期間努力し続けることは大切である。例えば世界トップクラスのバイオリニストを育てるベルリン芸術大学でバイオリンを学ぶ学生のうち、世界的なソリストになる可能性がある「S ランク」集団の学生は、他の集団の学生よりも、入学までに一人での練習に費やしてきた時間が圧倒的に長かった。具体的には、「S ランク」集団が 18 歳までに一人での練習に費やした 1 人当たり総練習時間は平均 7,410 時間であるのに対して、優秀だがスーパースターほどではない「A ランク」集団の 1 人当たり総練習時間は 5,301 時間、ソリストコースに受からず教員コースに入学した「B ランク」集団の 1 人当たり総練習時間は 3,420 時間と、極めて大きな差があった。傑出したバイオリニストになるには膨大な練習時間が必要であり、才能のある音楽家の間でさえも、平均する

²⁸¹ Terman and Oden (1959) には、高 IQ の男女約 1,500 人を約 10 歳のときにおける知能テストの結果に基づいて選出、その後の学歴、健康状態、職業、生活満足度等を 35 年もの長期にわたって追跡した調査の結果が示されている。同書によれば、約 10 歳のときの知能で測られた才能はその後の健康状態や職業や人間関係の安定性等に影響するようだが、その影響力はそれほど大きなものではなかった。彼らの研究には調査対象の属性の偏り等への批判があるものの、同一集団を低い脱落率で長期追跡した貴重な調査研究であるため、多数の追跡調査や関連調査が行われている。例えば Subotnik et al. (1989) ; Friedman and Martin (2011)。

²⁸² Ericsson (2014) を参照。

²⁸³ Seligman (2011) Chapter 6 を参照。

と練習時間の多い者ほどより大きな成功を収めていた²⁸⁴。

ただし大事なのは単なる投入時間ではなく、正しい方法で努力することである。一般には、何か一つのことに打ち込んで練習すればするほどスキルは向上すると考えられているかもしれない。しかし事実は異なり、医療でも料理でもテニスでも教育でも、同じことを同じやり方でやり続けていてもスキルは上達せず、むしろ緩やかに劣化することが様々な研究で示されている²⁸⁵。では、どのような方法で努力すればよいのだろうか。信頼性の高い科学的エビデンスが存在するのは、音楽やチェスのように長い伝統があり、トップになるための競争が激しい分野である。そのような分野では、傑出した業績とは何かが明確で、そのためにどのような練習を積みばよいかが標準化されており、自分の能力をわずかに上回る課題に挑戦し続けるための練習メニューを組み立てられる教師やコーチが存在する。そのような分野における卓越したプレイヤーにみられる共通点は、自分が快適にパフォーマンスできるコンフォートゾーンから少し飛び出すような「限界的練習 (deliberate practice)」を長期にわたって行っていることである。スキルの伸びが頭打ちになったとき壁を乗り越えるために大切なのは、才能でも意欲でも今までの方法を踏襲したまま努力量を増やすことでもなく、フィードバックを踏まえて練習方法を変えることなのである。

人生の成功に関するポジティブ心理学のエビデンスが示しているのは、生まれつきの才能よりも、ものの考え方、努力の量、そして努力の質が、人生の成功を説明する力が強いということであった。もし私たちが人生の大きな目標の達成に繋がるような一つのことに専念し、やればできると信じて正しい方法で課題に取り組む、粘り強く諦めずに努力し続けることができれば、自らの潜在能力を大いに発展させて成長し、結果的により意欲的な成果を達成できる可能性は高まると考えられる。

では、成功する人に共通する特徴をもつ人々を組織的に育てることはできるのだろうか。失敗しても成功するまでやり抜くことを組織の文化として定着させることは、可能なのだろうか。多くのポジティブ心理学研究は、成功に繋がる思考法や習慣は個人がトレーニングによって後天的に習得できるだけでなく、組織的にも強化できると説いている。では、どうすれば、壁にぶつかっても失敗しても、大きな目標を達成するまでやり抜く組織文化

²⁸⁴ Ericsson et al. (1993) を参照。同論文では行われていなかった二重盲検化等を行って同論文の頑健性を追試した研究に、Macnamara and Maitra (2019) がある。

²⁸⁵ Ericsson and Pool (2016) Chapter 5 に様々な分野における研究成果が紹介されている。医師の経験年数と医療の質については、Choudhry et al. (2005) を参照。

を醸成することができるのだろうか。

2. 組織が失敗から系統的に学習するためにはどうすればよいか

1990年代後半から「失敗学」を提唱してきた畑村洋太郎によれば、失敗を生かすシステムを組織的に確立することはそれほど難しいことではない。例えば失敗を起こした当事者が失敗をレポートにまとめて必要と思われる部署に情報発信することを義務づければ、失敗情報を適時に共有することができる²⁸⁶。後述する OKR (Objectives and Key Results) という業績管理システムも、成功事例や失敗事例を情報共有する仕組みとして機能させうる。

失敗から組織的に学習する仕組みには、失敗データベース以外にも様々な実践例がある。中でもよく知られているのは、病院や消防のように小さなミスが生死に影響するような組織での実践例であろう。医療看護の現場では、医療事故やその手前のインシデント（いわゆるヒヤリ・ハット）が生じた場合、職員がインシデント・アクシデントレポートを提出し、後日その解析が行われて、重要事例については原因究明と再発防止を目的としたケース会議でピアレビューが行われる²⁸⁷。救急救命活動でも、一般に全搬送事例について活動の協調性、迅速性、合規性などが事後検証されてデータベースに登録され、その後に二次検証にかけられる²⁸⁸。

失敗が人命に直結する分野やインシデントの事後検証が制度的に義務付けられている分野では、失敗情報が共有されるのは当然だと思われるかもしれない。しかし第一に、同じ制度のもとでも制度の運用方法は組織によって多様でありうる。例えば児童虐待分野の場合、同じ児童虐待関連法令に基づいていても、2018年8月時点で児童虐待事案の情報を自治体と警察との間で全件共有していたのは8府県に留まっていた²⁸⁹。児童虐待事案の情報共有さえあまり進んでいないのだから、児童虐待事案における失敗情報が組織の壁を超えて共有されているとは期待し難いだろう。第二に、同じ組織のもとでも人によって対応は変わりうる。例えば同じ病院内でも、部門長の開放性（openness）によって職員が投薬ミスを

²⁸⁶ 畑村（2000）を参照。

²⁸⁷ 遠山ほか（2007）を参照。

²⁸⁸ 2003年3月26日付「消防救第73号」（総務省消防庁救急救助課長）・「医政指発第0326002号」（厚生労働省医政局指導課長）、各地のメディカルコントロール協議会の事後検証実施指針や各自治体の救急活動に係る事後検証実施要綱等を参照。

²⁸⁹ 2018年8月31日付産経新聞朝刊。

進んで報告しようと思うかどうか、実際の投薬ミスがどのくらい起こるかが変わるという²⁹⁰。失敗から学ぶプロセスで特に難しいのは、失敗情報を共有するための仕組みを作ることではなく、失敗情報を進んで共有し、失敗から積極的に学ぶ方向へと組織のマインドセットを変え、人々の行動を変えることなのである。

児童虐待や投薬ミスのようなケースでも失敗情報の共有が難しいのなら、失敗が人命に直結するリスクが小さい地方自治体の一般行政部門で失敗を正直に認め、失敗から積極的に学ぶ組織風土を醸成することは可能なのだろうか。Heath and Heath (2010)は、なかなか変わらないように見えるものも、①象使いに進むべき方向を教え、②象にやる気を起こさせ、③目的地までの経路を示せば変えられることを、豊富な事例をもとに論じている。「象使い」と「象」という用語は、道徳心理学者である Jonathan Haidt の著作で示された用語である²⁹¹。「象使い」は心理学で「システム 2」といわれる熟慮に使われる遅い思考を、「象」は心理学で「システム 1」といわれる自動的に働く直感的な速い思考を指す²⁹²。

この①から③が具体的に何を意味するのかを、開発援助の現場での事例を用いて示してみよう²⁹³。1990 年にタフツ大学の 2 人の研究者がセーブ・ザ・チルドレンという NGO(Non Governmental Organization: 非政府組織)の仕事で幼児に蔓延する栄養失調を改善するためにベトナムの首都ハノイに降り立った時、まだベトナム国民にはベトナム戦争(アメリカ戦争)の記憶が色濃く残っていた。人々は米国の巨大 NGO を信用しておらず、現地の外務大臣からは「半年間で成果を出してほしい」と言われた。通常は現地オフィスを整えて専門家を雇い、何が問題かを特定して食糧を購入し農業技術を導入するために 2 年を要することを半年でやらなければならなくなった 2 人は、かつて文献で読んだ「優れた逸脱者(positive deviance)」に注目することにした。ここでいう優れた逸脱者とは、同じように貧しい貧困家庭の児童の中に一定程度いる、栄養状態のよい児童のことである²⁹⁴。現地のボランティアに協力してもらいながら治験対象の村に住む全幼児の体重を測り、各家庭の経済状態を格付けし、貧しいのに幼児の栄養状態が良い家庭を探し、一軒一軒聞き取り

²⁹⁰ Edmondson (2004) を参照。

²⁹¹ ハイト (2006) を参照。

²⁹² 判断や意思決定に関する脳の働きを「システム 1」と「システム 2」に分ける考え方は二重過程理論(dual-process theory)と呼ばれる。同理論については Kahneman (2011) を参照。

²⁹³ Pascale et al. (2010) Chapter 2. Heath and Heath (2010) Chapter 2 も併せて参照。

²⁹⁴ Zeitlin et al. (1990) を参照。

調査をして、優れた逸脱者のいる家庭の実践が通常の貧困家庭とはどのように異なっているのかを特定した。その結果、食事を2回ではなく数回に分けて与えていること、毎回の食事に少量のカニやエビや葉っぱを加えていること、食事中に汚いものに触ったら何回でも手を洗わせていること等が分かった。これらは極貧層でも可能な習慣であった。

象使い、つまり理性がどうすればよいか分かっても、象、つまり感情が動かなければ行動変化は持続しない。そこで象にやる気を与えるために Richard Pascale らが取り入れたのは、実地研修と成果の即時共有であった。優れた逸脱者の習慣はすぐに実践可能なものだったが、それまでの食習慣や衛生習慣とは大きく異なっていた。大人が不在の昼間に何度も食事を与えるのは手間がかかるし、小さな幼児にカニやエビや葉っぱを与えるのはまだ早いと考えられていた。だから、座学で知識を授けても、行動変容につながらない可能性があった。そこで、栄養失調の幼児をもつ母親と世話人が子どもたちを連れて近隣家庭に出かけ、栄養価の高い食材をもらって、優れた逸脱者の方法を実践するための調理法や衛生法などを自らの手を動かしながら現場で学ぶという実地研修を一定期間毎日行った。その後、全員を集めて幼児の体重を測定し、成果情報をその場で共有した。彼らの関わった22のベトナムの村で、栄養失調の幼児は2年間で65%から85%も激減し、優れた逸脱者アプローチは国内外に広がっていった。

この事例は、Heath and Heath (2010) では「①象使いに方向を教える」事例とされているが、実際には「②象にやる気を与える」と「③目的地までの経路を示す」の要素も入っている。まず、①「食事は2回ではなく数回」、「食事にはカニかエビか葉っぱ」、「汚いものに触ったら何度でも手を洗う」といった、シンプルで行動志向の具体的な活動目標が示されていた。次に、②理性ではなく感情に働きかける仕掛けがあった。すなわち、外から持ち込まれた解決策ではなく自分たちで発見した解決策だからやる気になるし、皆で実地研修するのは楽しいし、研修後に子どもの体重が増えたら嬉しい。つまり、頭ではなく心が動く仕掛けがあった。さらに、③やり続けたいくなるようなスキル習得と動機づけの仕組みが埋め込まれていた。実地研修を何度も繰り返せば新しい調理法と衛生法は自然と身につくし、コミュニティ全体で取組み、皆で業績測定をしているので、大きな目標を皆で協力して達成しようという気になる。よって行動変化は持続し、他のコミュニティにも広がったのである。

①の目標設定に関わる問題は、第2章で既に論じてきた。したがって以下では、②と③に関連する問題について、能力の問題、つまり「できない」を「やればできる」にする仕

組みと、意欲の問題、つまり「やりたくない」を「やりたい」にする仕組みに分けて、それぞれ検討する。

3. 失敗から学習する能力を組織的に高める方法

3-1. 体験型の研修と採用基準の変更

失敗から学習する能力を組織的に高めるための伝統的な手段は、研修である。優れた逸脱者の事例でも示されているように、研修を実際の行動変容に繋げるためには、知識習得を目的とした座学型の研修ではなく、能力向上を目的とした体験型の研修が有用だと言われる²⁹⁵。個人に失敗から学ぶという目的意識を持たせるためには、失敗を疑似体験させて、失敗から学ぶことの大切さを頭ではなく体に染み込ませることが大切だ、ということである。エキスパートの直感的判断に関する研究を重ねてきた Klein (2007) は、事業の実施前に事業が失敗に終わったと想定して「死亡前死因分析」(premortem) をすることで、成功確率を高められるのではないかと提案している。

失敗してもやり抜く能力を組織的に高める方法として研修以外に確かなエビデンスがある方法の一つは、採用段階でスクリーニングして楽観的な人を採用するというものである。Seligman らの研究によれば、生命保険会社は保険外交員の採用段階でスクリーニングを行って悲観主義者ではなく楽観主義者を採用すれば、保険契約高を増やしながら採用コストを節約することができる。楽観主義者は成長思考をもっており、何度も断られても簡単に離職せず、粘り強く営業を続けるからである²⁹⁶。

人材の採用基準を変えるのは、成長思考の人の多い組織をつくるために有力な方法かもしれない。しかし日本の公務員採用プロセスにおいて、楽観主義か悲観主義か、あるいは成長思考か固定思考かを測る心理テストの結果を重視して合否を決めるというアイディアは、容易には受け入れられない可能性が高いだろう。仮に受け入れられる可能性があるとしても、それが善いことなのかという疑問も残る。

²⁹⁵ 畑村 (2000) 第4章から第6章; Bohnet (2016) pp. 82-99 を参照。

²⁹⁶ Seligman (1990) pp. 95-115 を参照。ただし楽観主義の人は自信過剰に陥って判断ミスをするリスクも大きいので、論理や統計に基づく判断が必要な契約交渉や会計等の部門ではマイルドな悲観主義の人たちが必要だとされる (同書, Chapter 14)。

しかも、そもそも職場で失敗することは楽しいものではなく、人間には自分の失敗を歪曲したり隠したりする傾向が備わっている。人に生来備わっている性質に反して失敗から組織的に学べるようになるためには、働いている人が安心して挑戦し、失敗情報を安心して共有できるような環境をつくるのが肝要である²⁹⁷。そのために欠かせないと考えられるのは、マインドセットをリフレーミングした上で、業務システムを失敗前提のシステムへと変更することである。

3-2. マインドセットのリフレーミング

フレーミングとは、心理学用語で「ものごとの解釈の枠組み」を指す。フレーミング効果とは、同じ選択問題でも問題の提示の仕方（フレーム）が異なるだけで人々の選好が変わってしまうことを意味する²⁹⁸。組織が失敗から素早く学んで最善を尽くせるようになるためには、組織構成員が失敗や疑問について率直に話して支援を求めることができるように、組織のあらゆるレベルのリーダーはあらゆる機会を捉えて「失敗は恥や罪ではなく学習の好機である」とリフレーミングし続けることが重要だとされる²⁹⁹。

成長思考で失敗を捉えて最善を尽くそうとする組織文化は、日頃からの無数の積み重ねによって初めて醸成されるものである。2012年に倒産寸前の Nokia 社で最高経営責任者（Chief Executive Officer：CEO）に就任した Risto Siilasmaa は、同社復活の原動力となった経営哲学の核心を「疑い深い楽観主義（paranoid optimism）」と称している³⁰⁰。彼の言う「疑い深い楽観主義」とは、どれほど困難な状況でも必ず打開策があると信じた上で、最悪の事態を想定して行動することでリスクを最小化し、最善の手を打とうとするマインドセットのことである。新しいマインドセットを組織に埋め込み、新しい行動様式を組織文化にするために、彼はまず、組織のメンバーが安心して何でも率直に議論し、考えうる全てのシナリオを想定した上で日々改善を重ねて前進できるようにするために取締役がとるべき行動原則を定めた。従来は解決策を示さずに悪い知らせだけを上司に報告するのは批判すべきこととされていたが、新しいマインドセットでは悪いニュースは良いニュース

²⁹⁷ Edmondson（2012）を参照。

²⁹⁸ Tversky and Kahneman（1981）を参照。

²⁹⁹ Kotter and Cohen（2002）；Edmondson（2012）を参照。

³⁰⁰ Siilasmaa（2018）を参照。

とされ、解決策を見つけるために上司と部下が一緒に話し合うことが奨励された。2012年1月に倒産寸前の危機を迎えていた同社は、主力事業を高品質通信インフラ設備の製造・開発に移し、2018年には同分野で世界第二位の企業になるまで復活を遂げたのだった³⁰¹。

3-3. 失敗からの学習を促す業務システムへの変更

リーダーが失敗を恥や罪ではなく学習のチャンスだとリフレーミングし、失敗を罰するのではなく失敗情報の共有をむしろ称賛することで、人は失敗を率直に話しても大丈夫だと安心できるようになるかもしれない。それでもなお、自分をよく見せたいとか、失敗するのは怖くて嫌だという人間の思考様式を根本から変えることは容易ではないだろう。ただし、人の価値観が変わらなくても人の行動を変えることはできる。そのために有効だとされている方法は、業務システムを失敗からの学習を促す仕組みに変えるという方法である。その代表的な方法には、1) 業務プロセスに組織学習メカニズムを埋め込む、2) 計画(Plan)と実行(Do)の間に学習や実験のフェーズをあらかじめ設定する、3) 期限付きの別ユニットを立ち上げて大胆に試行錯誤できるようにする、4) 失敗を前提とした目標管理システムに変える、というパターンがある。

1) 業務プロセスへの組織学習メカニズムの埋め込み

Madson and Desai (2010)によれば、組織は成功体験よりも失敗体験からより多くを学ぶ。1957年の世界初人工衛星スプートニク1号の打ち上げから2004年までに世界で行われた30機関による4,646の打ち上げ事例を用いてロケットを軌道に乗せる成功確率の決定要因を確かめたところ、成功でも小さな失敗でもなく大きな失敗をした経験が次の打ち

³⁰¹ Nokia社はフィンランドを代表する企業で、2011年まで世界最大の携帯電話端末メーカーであった。しかし米アップル社のiPhoneと米グーグル社のAndroid対応端末の登場によって激変した携帯端末市場の新しい競争に乗り遅れ、2012年1月には米コダック社のように経営破綻するのではないかと取り沙汰されていた。その後、携帯電話端末事業をマイクロソフト社に売却して、主力事業を無線技術を中心とする高品質通信インフラ設備の製造・開発に移し、その業界で世界第二位の企業として復活を遂げた(Siilasmaa [2018]を参照)。なおiPhoneは米アップル社製のスマートフォン端末のことであり、Androidは米グーグル社製の携帯端末向けオペレーティングシステムのことである。

上げの成功確率を高めていたという。ただし、このエビデンスはより多くを学ぶためには大きな失敗を犯す必要があるという教訓を示しているのではなく、小さな失敗から学ぶことは難しいという事実を示していると考えの方が適切であろう。小さな失敗から組織的に学ぶためには、組織構成員が小さな失敗を無視したり隠したりせず、タイミングを逃さず活用することが必要だからである。

組織学習の成功要因を長年研究してきた Amy Edmondson は、失敗にはコンテキストに応じたアプローチが必要であって万能薬はないと警鐘を鳴らしつつ、一般に失敗から学ぶためには失敗に気づき、失敗を分析し、失敗を試みることが重要だと主張する³⁰²。そのメカニズムを業務プロセスに埋め込む方法としてよく知られているのは、日本のトヨタ自動車社における総合的品質管理（Total Quality Management：TQM）に基づくカイゼン活動である。藤本（1997）によれば、理念型としてのトヨタ的な開発・生産方式の特徴の一つは、生産性向上、品質改善、その他の製造問題の解決を継続的かつ全社的に行う組織学習メカニズムが機能している点にある。その核心は、無駄を発見し、それを体系的に削ぎ落とす仕掛けを作ることにある。問題解決サイクルの流れに沿ってその生産システムの改善能力の特徴を挙げると、問題発見を強制する仕掛け、現場への問題解決権限の委譲、問題解決ツールの標準化、改善案の迅速な実験・実施、標準の累積的改訂となる。問題発見を強制する仕掛けとは、問題を共有したくなるような動機づけの仕組みではなく、不良品問題のリアルタイムでの視覚化や、整理・整頓・清掃の徹底による異常の早期検知のような、現場の人々の問題発見能力を高める仕組みのことである。カイゼンのための問題解決手法や手順そのものが標準化・ルーティン化されており、誰でも使いこなせるように教育が徹底されている。しかもカイゼンは提案される前に実験され、既知の改善事例が事後的に正式提案されてくるために、膨大な数のカイゼン提案を迅速に処理して、その成果をすぐに標準化・マニュアル化することができる³⁰³。つまりトヨタ式のカイゼンとは、ごく小さな問題や失敗からも絶えず学び、組織の生産能力を体系的に高める仕組みであり、問題や失敗を放置したり隠したりすることを不可能にする仕掛けなのである。

³⁰² Edmondson（2012）Chapter 5 を参照。

³⁰³ 藤本（1997）第 2 章を参照。

2) 学習や実験のフェーズの事前設定

どのような業務にも適した万能薬はない。不確実性が大きく成功するためには試行して失敗することが避けられない業務では、通常のプロセスからの逸脱を事前に失敗と定義して対処することはできない。不確実性が大きく新たな知識を開発するために試行が不可欠なイノベーションの業務での生産方法として注目されているのが、「リーンスタートアップ」である³⁰⁴。リーン生産システムと同じようにリーンスタートアップにも多様な要素とバリエーションがあるが、その核となる要素は、Plan（計画）と Do（実行）の間に実験ないし学習というフェーズを設定し、はじめから完璧を目指さずに安価な試作品を作って計画段階の仮説が正しいかどうかを素早く実験して検証し、イノベーションを系統的に促すというプロセスにある³⁰⁵。実験による検証後の規模拡大というアイデア自体は、新しいものではない。リーンスタートアップで新しいのは、試行錯誤による失敗は当然生じるものだとして想定されており、その代わりに顧客の行動変化が起きたか否かを検証するためのフィードバックを実験で得て、科学的に評価することが要求される点であろう。

実験による検証後の規模拡大というアイデアは、制度イノベーションにも応用可能である。例えば近年、生産性向上特別措置法（平成 30 年法律第 25 号）に基づく新技術等実証制度が導入された。同制度は、AI（人工知能）、IoT（Internet of Things）、ビッグデータ、ブロックチェーン等の革新的な技術やビジネスモデルを用いたアイデアについて、効果が認められてから規制を改革するのではなく、参加者や期間を限定し、参加者の同意を得ること等を条件に「まずやってみる」ことを許容し、実証による情報や資料をもとに規制改革・政策形成を行うために導入された制度である³⁰⁶。同制度では、実証が上手くいかなかった場合におけるデータも貴重な資産として用いて、「実証による政策形成」を推進することが予定されている。つまり、実証の失敗を前提として、失敗情報を活かすことがあらかじめ企図されている。実証による政策形成の有効性自体を科学的に検証する仕組み

³⁰⁴ Ries（2011）を参照。

³⁰⁵ このアイデアは、トヨタ式のカイゼン活動の一部を切り取ったものでもある。リーンスタートアップでは、本番リリースないしサービスイン前のプロトタイプینگが公式なプロトコルとして要求されている。

³⁰⁶ 日本経済再生本部革新的事業活動評価委員会（2018）「規制のサンドボックス制度に関する基本資料集」、同委員会第 1 回配布資料、2018 年 8 月 31 日付, Retrieved May 22, 2019, from <http://www.kantei.go.jp/jp/singi/keizaisaisei/committee/dai1/sankou.pdf>.

も併せて制度化し、制度イノベーションの有効性を継続的・系統的に高めることが期待される。

3) 期限付きの専門組織の立ち上げ

不確実性が大きく新しい試みが組織の既存のプロセスや価値基準にあまり適合しない場合には、親組織とは別にイノベーション・ユニットないしチームを立ち上げ、スタートアップ企業のように大胆に試行錯誤できるイノベーション・サンドボックスを用意するという方法も有効でありうる。イノベーション・サンドボックスとは、製品・サービスまたは顧客の一部のみを実験対象にして、期限付きで仮説をテストし、標的顧客の行動変容につながったかを複数の同じ評価指標を用いて繰り返し評価することで成否を判断する仕組みのことである。Eric Ries によれば、困難に直面したり失敗しても諦めずに成功を追求する組織文化を長期にわたって維持するためには、イノベーションを成し遂げようとするチームが組織構成員の目に見えている前で力を発揮できるように、イノベーション・サンドボックスを用意することが肝要だ³⁰⁷。

イノベーションを促すために新組織を立ち上げ、時限付きで仮説をテストするという方法は、政府組織でも近年盛んに実践されている。その代表的な成功例の一つは、英国政府が内閣府内に立ち上げた行動インサイトチーム（The Behavioural Insights Team：BIT）、通称「ナッジ・ユニット」であろう³⁰⁸。ナッジとは、行動科学のエビデンスを活かして、人々の選択の自由を残したまま人々の行動をより良い方向へ導こうとするために用いられる情報ないし介入のことである。その概念が Thaler and Sunstein (2008) で示された 2008 年当時にも、行動科学の知見を政策に活かす取り組みはいくつも存在していた。だが、その成果を政府自らが様々な公共政策の効果を高めるために系統的に役立てようとする試みはなかった。それを世界で初めて実践したのが、ナッジ・ユニットであった。

ナッジ・ユニットは、①少なくとも 2 つの主要な政策分野で変化を起こす、②英中央政府内で行動科学の手法に対する理解を広める、③費用対効果を少なくとも 10 倍にする、

³⁰⁷ Ries (2011) Chapter 12 を参照。組織のプロセス及び価値基準との適合度によってイノベーションを生み出す組織戦略を変えるべきという指摘については、Christensen and Overdorf (2000) を参照。

³⁰⁸ ナッジ・ユニットの立ち上げ期からの歴史や成果については、Halpern (2015) 及び Thaler (2015) Chapter 16 を主に参照。

という3つの達成目標を掲げて、たった7名の職員で2010年に英内閣府内に立ち上げられた。ナッジ・ユニットには、2年間で目に見える成果を出せなければ組織を解散するという条件も課されていた。そのため初期の段階では、実施コストが低く、高い費用対効果を期待できる分野に焦点を当てた対照実験が進められ、確実に成果を上げていった³⁰⁹。設立から2年後、行動科学のエビデンスに基づく介入の効果は英中央政府内で認められ、職員数は2013年には15人、2015年には50人にまで増え、同ユニットは英国以外にもオフィスを構える社会目的企業（social purpose company：SPC）へと変貌を遂げた³¹⁰。この試みは、より科学的に信頼できるエビデンスに基づいて政策効果を評価し、より良い意思決定をしていこうとする英国政府全体の動きへと繋がると共に³¹¹、英国外の政府における行動科学の受容にも大きな影響を与えている³¹²。

なぜナッジ・ユニットは成功したのだろうか。立ち上げ期から2019年4月現在に至るまでナッジ・ユニットを率いている心理学者 David Halpern によれば、影響力の強いナッジ・ユニットを立ち上げるためには、「APPLES」という頭字語で表される6つの条件を備えることが重要である。第一に、政府組織内で権限をもつ人の支持を得ること（Administrative support）。第二に、政治家の支持を得ること（Political support）。第三に、各分野の専門家と政府で豊富な経験をもつ人材の両方をうまく登用すること（People）。第四に、職場が議会や政府組織に20分以内で着く距離にあること（Location）。第五に、ランダム化比較試験のような科学的方法に基づいて実証すること（Experiment）。最後に、分野ごとに行動科学における最新の学術研究を参照すること（Scholarship）である³¹³。なおナッジは、政治的・党派的な利害に資することを目的にして悪用されることがありうる

³⁰⁹ 例えば税滞納者に送る督促状の文面について行動科学の知見を取り入れて工夫を凝らした文章を何パターンか用意し、ランダム化比較試験でその効果を測るという取り組みがなされた。原課の責任者が行動科学への理解が深い人物だったことに加えて、例年発行している督促状の文面を変えただけなので追加的なコストは極めて低く、確かなエビデンスに基づいて文章を作成したので効果は極めて高かったという。

³¹⁰ 社名は Behavioural Insights Limited。2018年の1年間に31か国でランダム化比較実験等のサービスを提供したという（Retrieved May 23, 2019, from <https://www.bi.team/>）。

³¹¹ 例えば2013年には英国中央政府内に「what works network」が設置され、エビデンスに基づく政策形成を英政府全体で目指す動きへと発展していった。

³¹² OECD（2017）等を参照。

³¹³ Halpern（2015）p. 58を参照。「APPLES」には挙げられていないが、ナッジ・ユニットがサンセット条件付きの明確な目標を掲げた組織であったことも、確かな成果を上げるために必要な条件であったのではないかと思われる。

³¹⁴。本稿で繰り返し必要性を述べてきた政策プロセスの透明化・開放化は、ナッジの悪用を防ぐためにも不可欠であることを強調しておきたい。

4) 失敗を前提とした目標管理システムへの変更

業務プロセスを失敗前提の仕組みに変える4つ目の方法は、失敗を前提とした目標管理システムで業績評価や進捗管理を行うというものである。その一つの例は、従来型の目標管理制度（Management by Objectives：MBO）に代わる新しい目標管理制度として注目されている OKR（Objectives and Key Results）である³¹⁵。OKR の特徴の一つは、挑戦を奨励するために、簡単には達成できない意欲的な目標の設定を求める点にある。意欲的な目標の成功確率は高くないし、定量指標の目標達成度だけでは成果を捉えにくいこともある。よって、例えば7割の達成度でも成功とみなしたり、機械的な評価だけではなく定性評価も加えて総合評価するといった工夫が講じられる。なお多くの場合、評価は四半期に一度、モニタリングはより頻繁に行われ、例えば3割の達成度だと失敗だとみなされて原因を説明する責任が課される。すべての業績が可能な限りリアルタイムかつ二次利用容易な形式でデータベース化されて社内でオープンに管理されるのも、OKR の主な特徴の一つである。そのため OKR の業績管理システムは、成功や失敗を共有するためのデータベースとしても機能させうる。

OKR には、組織全体の目標が部門ごとの目標に連動し、部門の目標が個人の目標に連動するツリー型構造を有するという特徴もあるのだが、その際に、目標達成度を人事考課と直接連動させないという原則がある。目標達成度が人事考課と直結していると、意欲的な目標設定や失敗情報の共有が避けられてしまう可能性が高いからである。それぞれの目標は、上位目標に紐づけられている下位目標を達成すれば上位目標がほぼ自動的に達成されるように設計される。そのため目的と手段の倒錯が生じにくく、何より優先すべきタスクが明確になるので不要な仕事をやらなくて済むというメリットがある。第2章で述べた達成目標の充たすべき要件に留意しながら OKR のような仕組みを取り入れることによって、実用的でありながら意欲性や重要性という要件を充たした目標管理制度へと自治体評価を

³¹⁴ ナッジの提唱者の一人である Richard H. Thaler は、公益に反するナッジを「スラッジ」と称し警鐘を鳴らしている。ナッジは nudge for good を、スラッジは nudge for evil を指す用語である（Thaler [2018] を参照）。

³¹⁵ Doerr (2018) を参照。

刷新することが望まれる。

第3章第3節で述べたチェックリストの活用も、失敗を前提とした目標管理システムとみなすことができる。政策形成プロセスにいくつかの一時停止点を設けて、系統的な資料収集や代替案分析のような重要な手順が飛ばされるというミスを防ぐためのチェック項目と、チームワークの向上を図って認知バイアスによる判断ミスを防ぐためのコミュニケーションのチェック項目から成るチェックリストを作成して運用すれば、個人や集団が認知バイアスを自覚できなくても行動が変わり、組織はより失敗の少ない意思決定ができるようになると思われる。

4. 失敗から進んで学習する意欲を組織的に高める方法

前節では、失敗から学習するための能力を組織的に高める方法について述べてきた。しかし、挑戦による失敗に寛容な雰囲気を作り出し、失敗から組織的に学習できる仕組みを整えても、人々がやる気にならなければ組織は動かない。では、どうすれば組織構成員の意欲を系統的に高めることができるのだろうか。

主流派経済学は従来、個人が自発的に行動を起こすときの内的な動因（drive）よりも、個人に行動を起こしたいと思わせる外的な誘因（incentive）を重んじてきた。そして前節で紹介したナッジの概念を生んだ行動経済学が挑んだのは、経済主体が誘因に対して合理的に反応するという仮定であった。経済学における主流派経済学と行動経済学との関係と対をなすように、行動主義的心理学は活動それ自体がその活動の目的であり報酬になるような内発的動機づけ（intrinsic motivation）よりも、ある活動によって得られる報酬または避けられる罰という外発的動機づけ（extrinsic motivation）を重視してきた。そして外発的動機づけが人々のやる気を高めるという仮説に挑んだのが、内発的に動機づけられた自己充足的な活動に関する研究を行ってきた研究者たちであった。

1970年代前後から盛んになった内発的動機づけ研究によって示されたのは、ある活動への外的報酬は行動の理由を外部に向けさせ、内発的動機づけを抑制することによって、その活動へのやる気を高めるところかむしろ低下させうるという事実である³¹⁶。この事実は、人をやる気にさせるために外的報酬に頼るのをやめて、内的報酬だけに頼るべきだという

³¹⁶ Deci (1971) ; Deci (1975) ; Csikszentmihalyi (1975) ; Amabile et al. (1976) ; Deci and Flaste (1995) 等を参照。

ことを意味するわけではない。私たちの行動は多くの場合、外的報酬と内的報酬の両方に動機づけられている。フィードバックを含む全ての外的報酬は、統制的側面と、自律性や有能感に関する情動的側面の両面を有する。そのため、外部環境は内的報酬に対して促進的にも抑制的にも働きうる³¹⁷。もし制御的側面が相対的に強く顕れれば外的報酬は内発的動機づけを低下させてしまうであろうが³¹⁸、もし自律性や有能感に関する情動的側面が相対的に強く顕れれば、外的報酬は内的報酬を増大させ、人は活動それ自体を目的として楽しみ始めることになるだろう。しかも、最初は意思に反して他律的に動機づけられていた活動が内発的な報酬をもたらすようになることは、よくある。同じ人に同じ外的環境を与えても、それが内発的動機づけを抑制する外的報酬として作用するか、内発的動機づけを増大させる機制として作用するかは、それを受け取る自己の状態によって変化する³¹⁹。したがって問うべきことは、外発的動機づけと内発的動機づけのどちらを強化すべきかという二者択一ではなく、外発的動機づけの副作用が抑制され、内的報酬の増大が促されるのは人がどのような状態にあるときで、内的報酬の低下が予期されるのはどのような場合かということである。では、どのようなときに人は活動に対して内発的に動機づけられ、どのようなときに外発的動機づけによる内的報酬の低下を避けることができるのだろうか。

4-1. 理性ではなく感情に働きかける

人をやる気にさせるために大切なことの一つは、理性ではなく感情に働きかけることである。John P. Kotter は、組織を変革する上で最も重要なことは、実際に見て、触れて、心を動かすことによって人の行動を変えることだと指摘する³²⁰。例えば調達を一元化していないことでどれだけの無駄が生じているかを数字で示すだけでは、組織は動かないかもしれない。しかし各工場が別々に調達した値段の異なる 424 種もの手袋を山積みにして、実際の調達価格の値札をつけて取締役会や各工場の入り口で見せれば、調達改革の必要性に異論を唱える人は殆どいなくなるというわけだ³²¹。

退屈なデータより山積みされた手袋や生き生きとしたエピソードの方が想起されやすく、

³¹⁷ Amabile (1993)；鹿毛 (1994) を参照。

³¹⁸ Ariely et al. (2009) を参照。

³¹⁹ Csikszentmihalyi (1990) を参照。

³²⁰ Kotter (1995)；Kotter and Cohen (2002) を参照。

³²¹ このエピソードは Kotter (2002) に掲載されている実例である。

人の行動を変える力を持つのは、人間の認知能力や許容量には限界があるためだと考えられる。人間の判断と選択に系統的なエラーを生じさせるバイアスは多種多様だが、その要因になっているのは、人は見たものだけで直感的に判断しがちであるというという習性だ³²²。例えば人は、遠くの国にいる多数の児童の危機よりも、近くにいるたった一人の児童の危機を重視してしまいがちである。また、地震のような自然災害よりもマラリアや下痢のような日常的な病気の方が致死リスクが高いと分かっているにもかかわらず、後者よりも前者を重視してしまいがちである³²³。人の行動様式を変え、組織を変えたければ、人は論理的思考を要するような統計的事実を軽視し、記憶から容易に呼び出せる事柄や感情に従って直感的に判断しがちであるという習性を考慮に入れなければならないだろう。

4 - 2. 自律性の感覚を重んじる

人がある活動をやる気になったとき、その活動に内発的に動機づけられるのは、特に自律性 (autonomy) と有能感 (competence, self-efficacy, or self-confidence) を感じられるときだと Ryan and Desi (2000) は指摘する³²⁴。自律性とは、自らの行動を自分自身で決定しているという感覚のことである。目標の押しつけ、監視、評価、競争、報酬などはすべて、特定の行為をさせるような統制的側面が強まると内発的動機づけを低下させ、活動それ自体への熱意や興味を失わせ、業績を低下させる傾向がみられることが知られている。

ただし、自律性の感覚を保つために目標の設定や評価を止める必要はない。計画策定や目標設定のプロセスから主体的に参加した方が自己決定の感覚が強まり、目標達成に向けた内発的動機づけが高まると一般には信じられているかもしれないが、必ずしもそうともいえない³²⁵。なぜなら第一に、追求すべき目標が所与だとしても、それをどのようにやる

³²² Tversky and Kahneman (1974) ; Kahneman (2011) Chapter 7 を参照。

³²³ 近接性バイアスについては他に、Huber et al. (2011) ; Rosling et al. (2018) Chapter 4 and 5。感情ヒューリスティックによってリスク判断が歪んでしまうという現象については、Slovic et al. (2007) に詳しい。

³²⁴ 自律性と有能感ほどには活動を内発的に動機づける力が大きくないが、関係性 (relatedness) も内的報酬として感じられることがあるとされている。関係性とは、特定の人、集団、あるいは社会に属しているという感覚のことで、いわば人との絆を感じられるという感覚のことである。

³²⁵ 目標設定プロセスへの参加と業績の関係については、Locke and Latham (2002) pp. 707-708 を参照。

かにある程度選択の余地があり、自分で環境をコントロールしているという意識を持つことができれば自己決定感は強まりうる。例えば同じ高齢者介護施設で殆ど同じような待遇を受けた入居者でも、どのような人生にするかは自分次第だと世話係に伝えられ、自分が選んだ鉢植えを自分で管理する自由と、毎週映画を見るのを木曜にするか金曜にするかを選べる権利を与えられた入居者は、あなたの健康は職員がしっかり管理すると伝えられ、職員が選んだ鉢植えを職員が管理し、映画を木曜に上映するか金曜に上映するかを毎週職員が選んだ入居者と比べて、統計的に有意に身体的に健康で死亡率も低かったという³²⁶。

第二に、自己決定感は活動する時の外的環境だけではなく、その人の家庭環境や文化によっても左右される。例えば、個人主義ではなく集団主義を重んじる家族や文化で育った人は、自分が選んだ目標ではなく、むしろ自分に大きな影響を与えている人や自分の属する集団が選んだ目標の方が内発的動機づけを高める力が大きい場合があるという³²⁷。

第三に、選択肢が多ければ多いほど人の意思決定が最適化されたり、企業の利益が増えたりするわけでもない。人の認知能力や処理量には限界があるため、多すぎる選択肢はむしろ人の意思決定を阻害する傾向がみられる³²⁸。

最後に、内発的動機づけに関するメタ・アナリシスによれば、外的報酬と活動が事前に直接的に関係づけられていなければ、外的報酬が内的報酬を低下させる効果は生じにくい。また、関心の弱いタスクに対する外的報酬や関心の強いタスクに対する言語的強化のような非物的報酬は、内発的動機づけを強化することもある³²⁹。

外的報酬の制御的側面を弱め、自律性を強化することでイノベティブな活動への内的報酬を高めている好事例の一つは、米 Google 社が採用してきた「20%ルール」であろう。Google 社のエンジニアは、業務時間の 20%、つまり週に 1 日を普段の業務とは異なる業務に当てることが義務づけられており、Gmail や Google News のような革新的サービスが 20%ルール適用事業の下で生まれたものであることが知られている³³⁰。20%ルールがなくても同様の革新的サービスを生み出すことができたかもしれないが、金銭的報酬を増やすよりも時間の自由度を高める方が創造的な活動への内的報酬を高める力が強いことを支持

³²⁶ 1972 年に米国の高齢者介護施設で行われた実験については、Langer and Robin (1976) を参照。当該研究を含む自己決定感と健康との関係については、Iyengar (2010) を参照。

³²⁷ Iyengar and Lepper (1999) ; Iyengar (2010) を参照。

³²⁸ Iyengar and Lepper (2000) ; Heiss et al. (2013) を参照。

³²⁹ Cameron et al. (2001) を参照。

³³⁰ Doerr (2018) を参照。

するエビデンスは他にもある³³¹。

現場の自律性を高めることは、手段の目的化を避け、変化の大きい環境に適応しながら最善を尽くすためにも役立つ。「部下に目標を伝えつつ、その達成方法を指示しない」というシンプルなルールは、「委任型指揮（mission command）」と呼ばれる。これは予測の難しい事態に対処するための有効な方法で、戦争における軍隊の指揮方法や企業における組織共通の目標達成方法として有効だとされる³³²。前述したように日本では第一次分権改革によって機関委任事務が廃止されて基礎自治体の権限と責任が増え、基礎自治体が様々な政策分野の計画を自ら策定して実行することが求められるようになったが、中央政府が補助金受給要件等を細部にわたって規定していることが多い³³³。それでは中央政府が示した雛型に基づく計画策定とその進捗管理のための文書作成業務に多くの資源が注がれてしまうだけではなく、自律性の感覚が養われないために目標達成の意欲が喚起されず、目的と手段の倒錯という弊害が生じてしまう恐れがある。

委任型指揮とは反対の細かすぎる規則が政府のパフォーマンスを低下させていた象徴的な事例がある。Steve Kelman が 1993 年に Bill Clinton 元米大統領のもとで連邦調達政策局（Office of Federal Procurement Policy : OFPP）の局長になったとき、米国防総省（Department of Defense : DoD）は兵士用のチョコチップクッキーを調達するために 20 ページにわたる詳細な仕様書を用意し、市販のクッキーと比べて法外な価格で購買していたという。政府の不正を防ぐために定められた細かな調達規則が、政府調達のパフォーマンスを著しく妨げていたのである。過剰な制約が公益を損なう結果を招かないようにするために、現場の自律性を高め、規則ではなく使命重視で実績を適切な方法でモニタリングすることが求められよう³³⁴。

³³¹ 例えば Amabile（1976）を参照。

³³² Tetlock and Gardner（2015）Chapter10 を参照。

³³³ 例えば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することを目的とした地方創生事業において、業績測定型の行政評価が中央政府による統制のために用いられていることについて、山谷（2016）を参照。男女平等参画計画における事例については第 1 章第 3 節を参照。

³³⁴ Osborne and Gaebler（1992）Chapter 4, 5, and 9；上山（2002b）を参照。

4-3. 成果を素早く実感できるようにする

人が自律性と同じくらい内発的に動機づけられる活動は、有能感を感じられる活動である。例えば数学の問題を解いたり見知らぬ場所を発見することは、Mihaly Csikszentmihalyi のいうフロー経験をもたらす有能感を感じられる活動になりうる³³⁵。業績あるいは成果を上げることは、それ自体が内的報酬になり活動の目標になりうるということだ。したがって、前進を実感できるように変化を細分化したり、ゴールが近いと感じさせることは、有能感を養い、内発的動機づけを高める可能性が高いだろう³³⁶。前述した英国のナッジ・ユニットの「小さくはじめて、大きく育てる」というアプローチは、不確実性の高い事業で失敗リスクを抑えるためだけではなく、その活動に関わる人々の有能感を高めて、活動の内的報酬を高めるためにも有効な方法だと考えられる。

4-4. 外発的動機づけを用いる際に注意すべきこと

最後に、外発的動機づけを用いる際の注意事項について整理しておく。これまで述べてきたように、外発的動機づけを用いる際にまず注意しなければならないことは、内的報酬が大きい活動に対して外的報酬を与えることで内発的動機づけを低下させないようにすることである。逆に言うと、失敗情報の共有や失敗からの学習のように内発的に動機づけられにくい活動については、それを奨励するために外的報酬を用いることを積極的に検討すべきであろう。

自発的には行われにくい組織や社会としては奨励したい活動にインセンティブを付与したり、逆に民間部門に委ねていると過剰になる活動にサンクションを課すというのは伝統的な経済政策の一手法だが、「罰しない」というインセンティブの付与方法もありうる。第3章第3節で述べたように、航空業界では業界全体で航空事故の失敗から学習する仕組みが機能しているが、その背景には失敗した者を罰しないことによって失敗情報の共有を促す仕組みがある。具体的には、規制当局等による航空事故の調査結果を民事訴訟で証拠として採用することが法的に禁じられていたり、パイロットが小さなミスを起こしたとき

³³⁵ Csikszentmihalyi (1975) ; 同 (1990) を参照。フローとは、生活に意味づけと楽しさを与える深い没入経験を表す概念である。

³³⁶ Bandura and Schunk (1981) ; Kotter (1995) を参照。

には 10 日以内に報告書を提出すれば処罰されないという決まりがあるのだという³³⁷。

外的報酬を用いるときには、公平性や透明性にも注意を払う必要がある。人は公平性に関する感覚に敏感な動物で、自分や他人が不公平に扱われることを嫌う性向がある³³⁸。一方で、人は公平であると感じられるときに、新しいやり方を学ぶことに積極的になることが知られている³³⁹。意思決定や評価の透明性を高め、人々が納得して仕事に取り組めるようにすることは、公平性の問題だけではなく、イノベーションに関わる問題でもあるのだ。

5. 小括

本章では、主にポジティブ心理学及び組織科学分野の文献研究に基づいて、失敗から系統的に学び、目標を達成するまでやり抜く組織の構築方法を検討した。その結果、失敗から系統的に学ぶ能力を組織的に高めるためには、失敗を好機だと捉えるようにマインドセットをリフレーミングした上で、失敗からの学習を促す業務システムに変更することが有効であることが分かった。その代表的な方法は、業務プロセスに組織学習メカニズムを埋め込む、計画と実行の間に学習や実験のフェーズを事前に設定する、期限付きの組織を立ち上げて大胆に試行錯誤できるようにする、失敗を前提とした目標管理システムに変更する、という方法であった。他方で失敗から系統的に学ぼうとする意欲を組織的に高めるためには、感情に働きかけ、自律性と有能感を感じられるように情報提供やインセンティブ設計を行うことが求められることが分かった。

本稿で提案した方策が日本の自治体現場で実際に機能するかどうかは、今後の実証すべき研究課題として残されている。どれほど未然に防ごうとしても、失敗は避けられるものではない。航空会社が慎重な模擬試験を重ねて入念にチェックリストを作成するように、本稿で提示した仮説が、行政の無謬性という組織規範をもつとされる日本の地方自治体で

³³⁷ Syed (2015) Chapter 1。日本の航空事故や重大インシデントに関する調査機関は、運輸安全委員会である。その調査の目的は刑事責任の追及ではなく事故の原因究明及び再発防止・被害軽減であることから、事故調査は犯罪捜査のために認められたものと解釈してはならないことや、求めに応じて事故状況等を報告したことによって解雇その他の不利益な取扱いを受けないことが法律で明記されている（運輸安全委員会設置法第 8 条第 5 項及び第 30 条）。

³³⁸ Bolton and Ockenfels (2000) ; Dawes et al. (2007) ; Tricomi et al. (2010) を参照。

³³⁹ Frey et al. (2004) を参照。

どの程度有効なのかを慎重に確かめ、現場で機能する仕組みを細部に拘って丁寧に設計し実装することによって、失敗から系統的に学ぶことのできる風通しの良い組織を構築することが求められる。

終章 結論と今後の研究課題

1. 結論

本稿はこれまで、より良い自治体経営の実現を促す政策評価システムについて考察してきた。とりわけ、地方自治体の職員がボトムアップで変えることができるソフトな制度に注目して、政策プロセスの各段階で地方自治体は何をすべきかを考察してきた。本稿を終えるに当たって、各章から得られた主要な結果を本稿全体の考察を踏まえて改めてまとめておこう。

第1章「自治体ニーズ評価の理想と現実」では、政策科学の文献に基づいてニーズ評価の理想的な手続きを特定した上で、地方自治体によるニーズ評価の実態と課題を筆者が実施した2種類の社会調査結果を用いて明らかにした。その結果、地方自治体によるニーズ調査またはニーズ評価のための市民向け質問票調査の多くは明確な仮説がないまま行われていること、多くの地方自治体が政策の標的集団を代表しているとはいえないデータに基づいてニーズ評価を行っていること等が分かった。しかし現在はニーズ調査やニーズ評価の結果が再利用容易な形態で公表されていない場合が多いため、外部の第三者がその妥当性を検証するのは容易ではない。かかる状況を改善するために、第1章では理想的なニーズ評価を行おうとする意欲に着目してニーズ評価の社会的インパクトを確認したが、本稿の考察全体を踏まえれば、理想的なニーズ評価を行う能力を高める方法も浮かび上がってくる。誤った政策問題を解いている場合、どれほど政策を改善しても解くべき問題は解決されない。第1章で明らかにした理想的なニーズ評価に求められる手続きを現場で機能する具体的なルールに落とし込み、その効果を透明性と開放性の高い方法で検証することによって、政策問題の分布や標的集団のニーズをよりの確に把握できるようになるだろう。

第2章「達成目標が充たすべき要件」では、良い達成目標が充たすべき要件を考察し、先行研究に基づいて、「現実的だが意欲的かつ具体的で、実効性を伴う簡潔な目標が業績を高める。それに加えて、重要性が大きく、政策によって直接的に達成できる測定可能な目標であれば、目標設定や評価のバイアスを小さくすることができる。」という仮説を立てた。そしてその仮説の妥当性を、潜在能力の拡大に貢献したと考えられる成功事例で検証したのだった。すなわち、一般指標デザインに基づく差の差法を用いて、岡山県総社市「障がい者千人雇用事業」が地域の障害者就業率及び一般就労者割合を高めたことを定量的に明

らかにし、同事業の目標が本稿の考える達成目標の充たすべき7要件を全て充たしていたことを確認した。本稿の仮説検証方法は定性的な方法であり、今後より再現性の高い方法や異なる対象を用いて追試することが求められる。それでもなお、基礎自治体の障害者就労支援策が地域の障害者就業率及び一般就労者割合を高めることをおそらく初めて定量的に明らかにすると共に、当事者目線のアウトカム目標を定めてその指標値を定期的にモニタリングし、そのデータを公表し続けていれば第三者によるインパクト評価が可能であると示せたことは、日本の障害者就労支援策の在り方を問い直す上でも、当事者目線の意欲的な目標を定めることの重要性を示す上でも、一定の学術的・社会的な意義が認められるのではないかと思う。管見によれば、日本の障害者就労支援策では当事者目線のアウトカム目標が設定されている例は少なく、殆どの地方自治体は行政区域内の障害者就労者数を把握していない。そのような状況が大きく変わり、より良い目標が設定され追求されることが期待される。

第3章「バイアスとエビデンス」では、地方自治体が目標に向かって正しい方法でやり抜くための方策について、エビデンスのバイアスと認知バイアスによる判断ミスに注目して考察した。保健医療の分野では1990年代からエビデンスに基づく医療の取り組みが進み、利用可能な最良のエビデンスをつかい、つくり、ひろめる活動が急速に広まっている。認知バイアスに関する研究も進み、予想と現実のギャップの可視化によるバイアスの自覚、チェックリストの活用、集団浅慮に陥らないように集合知を活かす調査手法等、認知バイアスへの対処法について既に豊富な知見が蓄積されつつある。しかし、かつて David L. Sackett がエビデンスに基づく医療について述べていたように、エビデンスに基づく政策形成は行政職員の経験が無用にし、住民と市民の選択を無視するものではないし、料理本のようなレシピでもない³⁴⁰。エビデンスに基づくインパクト評価の方法論も、認知バイアスによる判断ミスを防ぐための仕組みも、制度さえ整えれば日本の地方自治体の政策現場で期待通りに機能するとは限らない。政策プロセスの透明性と開放性を高めた上で、確かな

³⁴⁰ 元の表現は、「エビデンスに基づく医療は、“料理本”のような医療ではない。現場の臨床医が最良の外部証拠を個々の臨床経験及び患者の選択と統合するというボトムアップのアプローチが必要なので、レシピ通りに料理するように、隷従的に個々の患者のケアをすることにはならない。(Evidence based medicine is not "cookbook" medicine. Because it requires a bottom up approach that integrates the best external evidence with individual clinical expertise and patients' choice, it cannot result in slavish, cookbook approaches to individual patient care.)」(Sackett et al. [1996] p. 72 より引用)である。

効果が期待される政策を積極的に試みてデータを取り、政策効果を確かめてアップデートし続ける継続的な努力が求められる。

第4章「目標を達成するまでやり抜く組織の構築方法」では、失敗から系統的に学び、目標を達成するまでやり抜く組織を作るためにはどうすればよいかを検討した。その結果、失敗から系統的に学ぶ能力を組織的に築くためには、失敗を好機だと捉えるようにマインドセットをリフレーミングした上で、失敗からの学習を促す業務システムに変更することが有効であることが分かった。また、失敗から学ぼうとする意欲を組織的に高めるためには、感情に働きかけ、自律性と有能感を感じられるように情報提供やインセンティブ設計を行うことが必要であることが分かった。どれほど未然に防ごうとしても、失敗は避けられるものではない。よって地方自治体が良い目標を達成するまでやり抜くためには、失敗情報を組織で共有し、失敗から組織的に学ぶことが不可欠である。本稿で提示した失敗から組織学習するための方策の効果を第3章で説明した方法で実証的に確かめて現場で機能する仕組みに改良することによって、失敗から系統的に学ぶことのできる風通しの良い組織を構築することが必要であろう。

2. 今後の研究課題

本稿では、より良い自治体経営を促す政策評価システムの構想を目指して、政策問題のニーズ評価の方法、政策の達成目標の設定方法、バイアスを克服して効果が確かな政策を講じる方法、そして目標達成までやり抜く組織の構築方法について検討し、前節で述べたような研究成果を得ることができた。しかし、本稿の検討課題から捨象された研究課題も多い。ここでは特に重要だと考えられる研究課題について、いくつかの方向性を提示したい。

第一に、本稿で提示した仮説が期待通りの効果をもたらすのかを日本の地方自治体のデータを用いて検証し、日本の自治体評価のあるべき姿とそこへの移行方法をより詳細に検討する必要がある。特に第4章で提案した諸制度について、それらが行政の無謬性という組織規範をもつとされる日本の地方自治体の組織文化を変え、失敗からの組織学習を本当に促すのかを実証的に確かめることが求められよう。運用方法等の微妙な違いによって、制度や政策の帰結は大きく変わりうる。地方自治体のコンテクストに最も適した政策評価システムとそのマネジメント方法をできるだけ具体的に定式化し、形骸化が懸念されてい

る自治体評価の刷新を促す知見を提示することが求められる。第二に、本稿は既存の社会経済構造を変えずにガバナンスや政策を改善する方法を中心に検討してきた。今後は、基礎自治体が現場発でボトムアップに着手できる改革だけではなく、中央政府や広域政府が関与すべき課題、住民や事業者や市民が果たすべき役割についても考察し、政策評価に関する公共ガバナンスの在り方を構想し実証することが大きな研究課題として残されている。より信頼できるデータやエビデンスに基づいて、より良い自治体経営が持続的に推進されるために、さらなる研究を進めていきたい。最後に、本稿はより良い自治体経営を実現するための評価システムに焦点を当てて考察してきたが、序章で述べたように自治体経営の全般的状態を把握するための社会指標の開発も重要な研究課題である。特にガバナンス要件とパフォーマンス要件については社会で広く合意された社会指標はないと思われるため、より一層の研究が求められる。

3. 謝辞

本稿は、多くの方々のご指導やご助言がなければ完成させることはできなかった。東京大学大学院情報学環の須藤修教授には、大学院生時代に指導教員として研究指導をしていただいただけでなく、現在に至るまで数え切れないほどの示唆に富むご指導やご助言、そして数多くの研究機会を賜った。筆者が政策評価を専門的に研究することになった一つの契機は、須藤教授が座長を務めていた旧・IT戦略本部（現在のIT総合戦略本部）「電子政府評価委員会」において電子行政に関する定量分析結果を発表する機会を2007年にいただいたことであった。また本稿を完成させることができたのは、須藤教授の長年にわたるご指導と叱咤激励があったからである。

本稿審査に当たっては、東京大学大学院情報学環の石崎雅人先生、同高木聡一郎先生、東京工科大学の榎俊吾先生、内閣府計量分析室の村舘靖之先生から、有益なご助言を頂いた。深謝の意を表したい。

須藤教授が座長を務めたNTTオープンラボ「『経済取引』探求グループ研究会」（1997年度～1998年度）は、東京工業大学の出口弘教授及び東京工科大学の榎俊吾教授と共同研究する機会を与えてくれた。出口教授は早い段階からプラットフォーム財のロックインがイノベーションに与える影響について鋭い考察をしており、自由な発想と確かな方法論が独創的な研究を生み出すことを間近で見せて下さった。筆者の出身大学院の研究室の先輩

でもある榑教授は、様々な業務を遂行しながら体系的な研究業績を積み重ねて博士論文を完成させ、博士号を取得していた。その記憶は本稿執筆中に繰り返し想起され、本稿完成まで筆者を後押ししてくれた。

文部科学省科学研究費補助金特定領域研究「ITの深化の基盤を拓く情報学研究」柱A06計画研究「グローバルな規模で進展する情報経済と新たな社会制度デザインに関する研究」（2001年度～2005年度）及び同特定領域研究「情報爆発時代に向けた新しいIT基盤技術の研究」柱B01計画研究「知識社会経済システムの共創的発展とそのガバナンスに関する研究」（2005年度～2010年度）では、東京大学大学院情報学環の田中秀幸教授と共同研究する貴重な機会を得た。多彩な研究課題を並行して進めながら丁寧な実証研究を積み重ねる田中教授の研究手腕に多くの刺激をいただいた。

本稿は、その他にも様々な共同研究プロジェクトや研究助成の支援に支えられている。まず本研究は、筆者が研究代表として遂行してきた文部科学省科学研究費補助金若手研究（A）「公共IT投資の費用対効果に関する調査研究」（2011年度～2012年度及び2014年度～2017年度、研究課題番号：23680025）及び同基盤研究（B）「地方自治体による雇用政策の効果と雇用政策の分権化の有効性に関する調査研究」（2018年度～2023年度予定、研究課題番号：18H00860）の助成を受けている。それらの助成がなければ、第2章で述べた岡山県総社市による障害者就労支援策のインパクト評価を行うことは決してできなかった。

同じく筆者が研究代表として受託した「情報媒体等利用状況調査」（2015年度）及び「水戸市男女平等参画に関する市民意識調査及び水戸市男女平等参画推進状況事業所調査の分析に関する研究」（2018年度）は、基礎自治体の日常業務で見え隠れする認知バイアスや、各種計画策定のために行われる社会調査におけるデータのバイアスに注目する好機を提供してくれた。しかしそれ以上に貴重であったのは、基礎自治体の職員と非公開の場で何度もコミュニケーションを交わしたり、フィールドに出て地域住民や利害関係者の声に耳を傾ける機会を得たことにより、補完性の原理の重要性と、実践で役立つ系統的な評価システムの必要性を実感できたことにある³⁴¹。本稿の実証研究の基礎となった社会調査に協力

³⁴¹ 「補完性の原理」とは効率的に管理できる最小単位で物事を扱おうとする考え方の中で、そのために政府間の事務分担に際して基礎自治体を優先し、次に広域自治体を優先し、中央政府は広域自治体でも担うことができない事務事業のみを担うとする原則のことである（タレブ [2017] p. 159；大森 [2008] p. 5を参照）。補完性の原理に基づく労働政策の分権化の可能性については、後藤（2004）で考察している。

してくれた基礎自治体職員の方々には、本研究に関する貴重なご助言や示唆をいただいた。
記して感謝申し上げます。

地方自治体や中央政府の各種委員会等に参加する機会を得て、資源制約がタイトになる中で多方面からの様々な要求に対処しなければならない厳しさや、フィールドから遠い場所でインセンティブ装置や技術的基盤を設計することの難しさ等を窺い知ることができたことも、本研究を構想する上で大きな刺激になった。貴重な機会を与えてくれた関係諸氏に厚く御礼申し上げます。最後に、複数のネットワークへの重複参加を許容する柔軟な職場環境と、日々研究の刺激を与えてくれる茨城大学の同僚と学生、そしていつも支えてくれる家族に、心からの感謝の意を表したい。

参考文献

- 阿部彩 (2008)『子どもの貧困—日本の不平等を考える』岩波書店.
- Acemoglu, Daron and James Robinson (2012) *Why Nations Fail: The Origins of Power, Prosperity and Poverty*, Currency. (鬼澤忍訳 [2013]『国家はなぜ衰退するのか—権力・繁栄・貧困の起源』, 上下巻, 早川書房.)
- 相原守夫 (2018)『診療ガイドラインのための GRADE システム』, 第 3 版, 中外医学社.
- Allen, Claire and Kiley Richmond (2011) The Cochrane Collaboration: International activity within Cochrane Review Groups in the first decade of the twenty-first century, *Journal of Evidence-Based Medicine*, 4(1), pp. 2-7.
- Amabile, Teresa M. (1993) Toward new conceptualizations of intrinsic and extrinsic motivation in the workplace, *Human Resource Management Review*, 3(3), pp. 185-201.
- Amabile, Teresa M., William DeJong, and Mark R. Lepper (1976) Effects of externally imposed deadlines on subsequent intrinsic motivation, *Journal of Personality and Social Psychology*, 34, pp. 92-98.
- Ariely, Dan, Uri Gneezy, George Loewenstein, and Nina Mazar (2009) Large stakes and big mistakes, *Review of Economic Studies*, 76, pp. 451-469.
- 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), pp. 19-32.
- Atkinson, Tony, Bea Cantillon, Eric Marlier, and Brian Nolan (2002) *Social Indicators: The EU and Social Inclusion*, Oxford University Press.
- Bandura, Albert and Dale H. Schunk (1981) Cultivating competence, self-efficacy, and intrinsic interest through proximal self-motivation, *Journal of Personality and Social Psychology*, 41, pp. 586-598.
- Banerjee, Abhijit V. and Esther Duflo (2011) *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*, PublicAffairs. (山形浩生訳 [2012]『貧乏人の経済学—もういちど貧困問題を根っこから考える』みすず書房.)
- Bardach, Eugene and Eric M. Patashnik (2016) *A Practical Guide for Policy Analysis: The Eightfold Path to More Effective Problem Solving*, 5th edition, Sage. (白石賢司・鍋島

- 学・南津和広訳 [2012]『政策立案の技法—問題解決を「成果」に結び付ける 8 つのステップ』東洋経済新報社.)
- Baumeister, Roy E., Ellen Bratslavsky, Mark Muraven, and Dianne M. Tice (1998) Ego depletion: Is the active self a limited resource?, *Journal of Personality and Social Psychology*, 74(5), pp. 1252-1265.
- Baumeister, Roy E., Kathleen Vohs, and Dianne Tice (2007) The strength model of self-control, *Current Directions in Psychological Science*, 16, pp. 351-355.
- 米国行政学会・行政経営センター, 上山信一監訳監修 (2001)『行政評価の世界標準モデル—戦略計画と業績測定』東京法令出版.
- Behrman, Jere R. (2010) The International Food Policy Research Institute (IFPRI) and the Mexican PROGRESA anti-poverty and human resource investment conditional cash, *World Development*, 38(10), pp. 1473-1485.
- Bentham, Jeremy (1891) *A Fragment on Government*, edited with an Introduction by F.C. Montague, The Clarendon Press.
- Beveridge, William (1942) *Social Insurance and Allied Services*, HMSO.
- Boardman, Anthony E., David H. Greenberg, Aidan R. Vining, and David L. Weimer (2006) *Cost-Benefit Analysis: Concepts and Practice*, 3rd edition, Prentice Hall.
- Bohnet, Iris (2016) *What Works: Gender Equality by Design*, Belknap Press. (池村千秋訳 [2018]『WORK DESIGN(ワークデザイン)—行動経済学でジェンダー格差を克服する』エヌティティ出版.)
- Bok, Derek (2010) *The Politics of Happiness: What Government Can Learn from the New Research on Well-Being*, Princeton University Press. (土屋直樹・茶野努・宮川修子訳 [2011]『幸福の研究—ハーバード元学長が教える幸福な社会』東洋経済新報社.)
- Bolton, Gary E. and Axel Ockenfels (2000) ERC: A theory of equity, reciprocity, and competition, *American Economic Review*, 90(1), pp. 166-193.
- Booth, Charles (1892) *Life and Labour of the People in London*, Volume 1, Macmillan.
- Cameron, Drew B., Anjini Mishra, and Annette N. Brown (2015) The growth of impact evaluation for international development: How much have we learned?, *Journal of Development Effectiveness*, doi: 10.1080/19439342.2015.1034156.
- Cameron, Judy, Katherine M. Banko, and W. David Pierce (2001) Pervasive negative

- effects of rewards on intrinsic motivation: The myth continues, *The Behavior Analyst*, 24(1), pp. 1-44.
- Campbell, Donald (1969) Reforms as experiments, *American Psychologist*, 24, pp. 409-429.
- Cassels, Alan (2015) *The Cochrane Collaboration: Medicine's Best-Kept Secret*, Agio Publishing House.
- Chandler, Jackie and Sally Hopewell(2013)Cochrane methods: Twenty years experience in developing systematic review methods, *Systematic Reviews*, 2, p. 76.
- Chernew, Michael E. and Joseph P. Newhouse (2008) What Does the RAND Health Insurance Experiment Tell Us About the Impact of Patient Cost Sharing on Health Outcomes?, *American Journal of Managed Care*, 412-414.
- Choudhry, Niteesh K. Choudhry, Robert H. Fletcher, and Stephen B. Soumerai (2005) Systematic review: The relationship between clinical experience and quality of health care, *Annals of Internal Medicine*, 142(4), pp. 260-273.
- Christensen, Clayton M. and Michael Overdorf (2000) Meeting the challenge of disruptive change, *Harvard Business Review*, 78(2), pp. 66-76.
- Cochrane, Archiebald L. (1972) *Effectiveness and Efficiency: Random Reflections on Health Services*, Nuffield Provincial Hospitals Trust.
- Collier, Paul (2009) *Wars, Guns, and Votes: Democracy in Dangerous Places*, Harper Collins. (甘糟智子訳 [2010]『民主主義がアフリカ経済を殺すー最底辺の10億人の国で起きている真実』日経BP社.)
- Collins, Jim (1994) *Built to Last: Successful Habits of Visionary Companies*, HarperBusiness. (山岡洋一訳 [1995]『ビジョナリー・カンパニーー時代を超える生存の原則』日経BP社.)
- Collins, Jim (2009) *How the Mighty Fall: And Why Some Companies Never Give In*, JimCollins. (山岡洋一訳 [2010]『ビジョナリー・カンパニー3ー衰退の五段階』日経BP社.)
- Cronbach, Lee J. (1980) *Toward Reform of Program Evaluation*, Jossey-Bass.
- Csikszentmihalyi, Mihaly(1975) *Beyond Boredom and Anxiety: Experiencing Flow in Work and Play*, Jossey-Bass. (今村浩明訳 [1991]『楽しむということ』思索社.)

- Csikszentmihalyi, Mihaly (1990) *Flow: The Psychology of Optimal Experience*, Harper Perennial. (今村浩明訳 [1996]『フロー体験 喜びの現象学』世界思想社.)
- Dawes, Christopher T., James H. Fowler, Tim Johnson, Richard McElreath, and Oleg Smirnov (2007) Egalitarian motives in humans, *Nature*, 446, pp. 794-796.
- Deci, Edward L. (1971) Effects of externally mediated rewards on intrinsic motivation, *Journal of Personality and Social Psychology*, 18, pp. 105-115.
- Deci, Edward L. (1975) *Intrinsic Motivation*, Springer. (安藤延男・石田梅男訳 [1980]『内発的動機づけ—実験社会心理学的アプローチ』誠信書.)
- Deci, Edward L. and Richard Flaste (1995) *Why We Do What We Do: Understanding Self-Motivation*, Putnam Adult. (桜井茂男訳 [1999]『人を伸ばす力—内発と自律のすすめ』新曜社.)
- Diener, Carol I. and Carol S. Dweck (1978) An analysis of learned helplessness: Continuous changes in performance, strategy, and achievement cognitions following failure, *Journal of Personality and Social Psychology*, 36, pp. 451-462.
- Diener, Carol I. and Carol S. Dweck (1980) An analysis of learned helplessness: II. The processing of success, *Journal of Personality and Social Psychology*, 39(5), pp. 940-952.
- Dijkers, Marcel (2013) Introducing GRADE: A systematic approach to rating evidence in systematic reviews and to guideline development, *KT Update*, 1(5), Retrieved July 19, 2019, https://ktdrr.org/products/update/v1n5/dijkers_grade_ktupdatev1n5.pdf.
- Doerr, John (2018) *Measure What Matters: How Google, Bono, and the Gates Foundation Rock the World with OKRs*, Portfolio.
- Duckworth, Angela L. (2016) *Grit: The Power of Passion and Perseverance*, Scribner. (神崎朗子訳 [2016]『やり抜く力 GRIT(グリット)—人生のあらゆる成功を決める「究極の能力」を身につける』ダイヤモンド社.)
- Duckworth, Angela L., Christopher Peterson, Michael D. Matthews, and Dennis R. Kelly (2007) Grit: Perseverance and passion for long-term goals, *Journal of Personality and Social Psychology*, 92(6), pp. 1087-1101.
- Dweck, Carol S. (2006) *Mindset: The New Psychology of Success*, updated edition, Random House. (今西康子訳 [2016]『マインドセット—「やればできる!」の研究』

草思社.)

Edmondson, Amy C. (2004) Learning from mistakes is easier said than done: Group and organizational influences on the detection and correction of human error, *The Journal of Applied Behavioral Science*, 40(1), pp. 66-90.

Edmondson, Amy C. (2012) *Teaming: How Organizations Learn, Innovate, and Compete in the Knowledge Economy*, John Wiley & Sons. (野津智子訳 [2014]『チームが機能するとはどういうことか』英治出版.)

Ericsson, K. Anders (2014) Why expert performance is special and cannot be extrapolated from studies of performance in the general population: A response to criticism, *Intelligence*, 45, pp. 81-103.

Ericsson, K. Anders, Ralf Th. Krampe, and Clemens Tesch-Romer (1993) The role of deliberate practice in the acquisition of expert performance, *Psychological Review*, 100(3), pp. 363-406.

Ericsson, K. Anders and Robert Pool (2016) *Peak: How to Master Almost Anything*, Viking. (土方奈美訳 [2016]『超一流になるのは才能か努力か?』文藝春秋.)

Esserf, James K. (1998) Alive and well after 25 years: A review of groupthink research, *Organizational Behavior and Human Decision Processes*, 73(2-3), pp. 116-141.

Fernald, Lia C. H., Paul J. Gertler, and Lynnette M. Neufeld (2009) 10-year effect of Oportunidades, Mexico's conditional cash transfer programme, on child growth, cognition, language, and behaviour: A longitudinal follow-up study, *Lancet*, 374, pp. 1997-2005.

Fischhoff, Baruch (1975) I knew it would happen: Remembered probabilities of once-future things, *Organizational Behavior and Human Performance*, 13(1), pp. 1-16.

Fischhoff, Baruch (1977) Perceived informativeness of facts, *Journal of Experimental Psychology Human Perception & Performance*, 3(2), pp. 349-358.

フィッシュキン, ジェイムズ・S (曾根泰教監修, 岩木貴子訳) (2011)『人々の声が響き合うとき—熟議空間と民主主義』早川書房.

Frey, Bruno S. (2008) *Happiness: A Revolution in Economics*, The MIT Press. (白石小百合訳 [2012]『幸福度をはかる経済学』NTT出版.)

- Frey, Bruno S. and Alois Stutzer (2002) *Happiness and Economics: How the Economy and Institutions Affect Human Well-Being*, Princeton University Press. (佐和隆光監訳・沢崎冬日訳 [2005]『幸福の政治経済学—人々の幸せを促進するものは何か』ダイヤモンド社.)
- Frey, Bruno S., Matthias Benz, and Alois Stutzer (2004) Introducing procedural utility: Not only what, but also how matters, *Journal of Institutional and Theoretical Economics*, 160(3), pp. 377-401.
- Friedman, Howard S. and Leslie Martin (2011) *The Longevity Project: Surprising Discoveries for Health and Long Life from the Landmark Eight-Decade Study*, Hudson Street Press.
- 藤本隆宏 (1997)『生産システムの進化論—トヨタ自動車にみる組織能力と双発プロセス』有斐閣.
- Galbraith, John Kenneth (1958) *The Affluent Society*, Houghton Mifflin. (鈴木哲太郎訳 [1960]『ゆたかな社会』岩波書店.)
- Gawande, Atul (2009) *The Checklist Manifesto: How to get Things Right*, Metropolitan Books. (吉田竜訳 [2011]『あなたはなぜチェックリストを使わないのか?—重大な局面で“正しい決断”をする方法』晋遊舎.)
- 玄田有史・三谷直紀・川口章 (2000)「<学界展望>労働経済学研究の現在--1997~99年の業績を通じて」,『日本労働研究雑誌』, 42(2・3), pp. 2-34.
- Gilbert, Daniel T. (2006) *Stumbling on Happiness*, Knopf. (熊谷淳子訳 [2007]『幸せはいつもちょっと先にある—期待と妄想の心理学』早川書房.
- Gollwitzer, Peter M. (1999) Implementation intentions: Strong effects and simple plans, *The American Psychologist*, 54(7), pp. 493-503.
- Gollwitzer, Peter M. and Paschal Sheeran (2006) Implementation intentions and goal achievement: A meta-analysis of effects and processes, *Advances in Experimental Social Psychology*, 38, pp. 69-119.
- 後藤玲子 (2004)「雇用政策の成功要因に関する経済学的考察—OECD 加盟の欧米諸国の経験とわが国への含意」,『社会・経済システム』, 25, pp. 154-167.
- 後藤玲子 (2007)「パブリック・イノベーションと電子行政」,『茨城大学社会科学論集』, 44, pp. 57-77.

- 後藤玲子 (2009) 「自治体 IT ガバナンスが電子自治体の業績に与える影響」, 『茨城大学社会科学論集』, 48, pp. 39-52.
- Gotoh, Reiko (2009) Critical factors increasing user satisfaction with e-government services, *Electronic Government*, 6(3), pp.252-264.
- 後藤玲子(2012) 「利用者ニーズの先をいくー海外にみるプロアクティブ型行政サービス」, 『月刊 LASDEC』, 42(3), pp. 13-19.
- 後藤玲子 (2016) 「市民と行政の認識ギャップー某自治体の社会調査から得られる示唆」, 『第 22 回社会情報システム学シンポジウム学術講演論文集』, A4 版 6 頁.
- 後藤玲子(2017a) 「地方自治体の政策運営におけるデータ活用の実態と課題ー水戸市による市民向け質問票調査の実態調査結果」, 『社会科学論集』, 1, pp. 45-56.
- 後藤玲子 (2017b) 「福祉における情報の壁ー自治体の事例調査に基づく考察」, 『社会政策』, 9(2), pp. 135-146.
- 後藤玲子 (2019a) 「地方自治体による障害者就労支援策の効果測定方法に関する一考察」, 『社会情報学会関東支部研究発表会要旨集』, A4 判 2 頁, 2019 年 2 月.
- 後藤玲子(2019b) 「地方都市居住者の管理職への到達度を規定する要因ー男女差に焦点を当てて」, 『社会科学論集』, 4, pp. 29-44.
- 後藤玲子(2019c) 「どうすれば失敗してもやり抜く組織を作ることができるか」, 『社会科学論集』, 5, pp. 1-20.
- Grove, Andrew S. (1983) *High Output Management*, Random House. (小林薫訳 [2017] 『HIGH OUTPUT MANAGEMENTー人を育て、成果を最大にするマネジメント』日経 BP 社.)
- Guyatt, Gordon and Drummond Rennie eds. (古川壽亮・山崎力監訳) (2003) 『臨床のための EBM 入門ー決定版 JAMA ユーザーズガイド』医学書院.
- 南風原朝和 (2014) 『続・心理統計学の基礎ー統合的理解を広げ深める』有斐閣.
- ハイト, ジョナサン (藤澤隆史・藤澤玲子訳) (2006) 『しあわせ仮説ー古代の知恵と現代科学の知恵』新曜社.
- Halpern, David (2015) *Inside the Nudge Unit: How Small Changes Can Make a Big Difference*, Ebury Press.
- Halpern, David and Danielle Mason (2015) Radical incrementalism, *Evaluation*, 21(2), pp. 143-149.

- 浜口信明・高橋百合子（2008）「条件付き現金給付による貧困対策の政治経済学的考察—ラテンアメリカの事例から」, 『国際経済雑誌』, 197(3), pp. 49-64.
- Harington, Micheal (1962) *The Other America: Poverty in the United States*, Macmillan Publishing. (内田満・青山保訳 [1965] 『もう一つのアメリカ』 日本評論社.)
- Haskins, Ron and Greg Margolis (2015) *Show Me the Evidence: Obama's Fight for Rigor and Results in Social Policy*, Brookings Institution Press.
- 畑村洋太郎（2000）『失敗学のすすめ』 講談社.
- ハトリー, ハリー・P（上野宏・上野真城子訳）（2004）『政策評価入門—結果重視の業績測定』 東洋経済新報社.
- Haynes, Alex B., Thomas G. Weiser, William R. Berry, Stuart R. Lipsitz, Abdel-Hadi S. Breizat, Patchen Dellinger, Teodoro Herbosa, Sudhir Joseph, Pascience L. Kibatala, Marie Carmela M. Lapitan, Alan F. Merry, and Krishna Moorthy, et al. (2009) A surgical safety checklist to reduce morbidity and mortality in a global population, *New England Journal of Medicine*, 360, pp. 491-499.
- Heath, Chio and Dan Heath (2010) *Switch: How to Change Things When Change Is Hard*, Currency. (千葉敏生訳 [2013] 『スイッチ! —「変わらない」を変える方法』 草思社.)
- Heiss, Florian, Adam Leive, Daniel McFaddenc, and Joachim Winter (2013) Plan selection in Medicare Part D: Evidence from administrative data, *Journal of Health Economics*, 32, pp. 1325-1344.
- 本多則恵・本川明（2005）「インターネット調査は社会調査に利用できるか—実験調査による検証結果」, 『労働政策研究報告書』 17, Retrieved May 10, 2019, from <http://www.jil.go.jp/institute/reports/2005/documents/017.pdf>.
- 星野崇宏（2009）『調査観察データの統計科学—因果推論・選択バイアス・データ融合』 岩波書店.
- Huber, Michaela, Leaf Van Boven, A. Peter McGraw, and Laura Johnson-Graham (2011) Whom to help? Immediacy bias in judgments and decisions about humanitarian Aid, *Organizational Behavior and Human Decision Processes*, 115, pp. 283-293.
- Ichimura, Hidehiko and Christopher Taber (2001) Propensity-Score Matching with Instrumental Variables, *American Economic Review*, 91(2), pp. 119-124.
- Imbens, Guido W. and Donald B. Rubin (2015) *Causal Inference for Statistics, Social, and*

- Biomedical Sciences: An Introduction*, Cambridge University Press.
- 猪木武徳・大竹文雄編 (2001)『雇用政策の経済分析』東京大学出版会.
- 伊藤公一朗 (2017)『データ分析の力ー因果関係に迫る思考法』光文社.
- 岩田正美 (2007)『現代の貧困ーワーキングプア／ホームレス／生活保護』筑摩書房.
- Iyengar, Sheena S. (2010) *The Art of Choosing*, Grand Central Publishing. (櫻井祐子訳 [2010]『選択の科学』文藝春秋.)
- Iyengar, Sheena S. and Mark R. Lepper (1999) Rethinking the value of choice: A cultural perspective on intrinsic motivation, *Journal of Personality and Social Psychology*, 76(3), pp. 349-366.
- Iyengar, Sheena S. and Mark R. Lepper (2000) When choice is demotivating: Can one desire too much, *Journal of Personality and Social Psychology*, 9(6), pp. 995-1006.
- Janis, Irving L.(1982) *Groupthink: Psychological Studies of Policy Decisions and Fiascoes*, 2nd edition of *Victims of Group Think*, Houghton Mifflin.
- Johnson, Claire (2008) Evidence-based practice in 5 simple steps, *Journal of Manipulative and Physiological Therapeutics*, 31(3), pp. 169-170.
- 鹿毛雅治 (1994)「内発的動機づけ研究の展望」『Japanese Journal of Educational Psychology』, 42, pp. 345-359.
- Kahneman, Daniel (2011) *Thinking, Fast and Slow*, Farrar, Straus and Giroux. (村井章子訳 [2014]『ファスト&スローーあなたの意思はどのように決まるか?』上下巻,早川書房.)
- 金本良嗣・蓮池勝人・藤原徹 (2006)『政策評価ミクロモデル』, RIETI 経済政策分析シリーズ 15, 東洋経済新報社.
- 加藤哲男(2009)「市民と行政職員の都市基盤に関する意識の比較」,『土木計画学研究・論文集』, 26(2), pp. 297-306.
- 勝又幸子 (2006a)「第 1 回障害者実態生活調査の概要」, 勝又幸子編 (2006)『厚生労働科学研究費補助金障害保健福祉総合研究事業 障害者の所得保障と自立支援施策に関する調査研究 (H17-障害-003) 平成 17 年度 総括研究報告書』所収, pp. 15-46.
- 勝又幸子 (2006b)「障害者福祉・雇用問題の実態と課題ー政府統計調査から見えてくるもの」『Int'lecowk』, 965, pp. 15-24.
- Kelman, Steve (2005) *Unleashing Change: A Study of Organizational Renewal in*

- Government*, Brookings Institution Press.
- Khan, Khalid S., Regina Kunz, Jos Kleijnen, and Gerd Antes (2003) Five steps to conducting a systematic review, *Journal of Royal Society of Medicine*, 96(3), pp. 118-121.
- Klein, Gary (2007) Performing a Project Premortem, *Harvard Business Review*, 85: 18-19.
- 近藤 絢子 (2014)「雇用確保措置の義務化によって高齢者の雇用は増えたのか—高年齢者雇用安定法改正の政策評価」『日本労働研究雑誌』, 56(1), pp. 13-22.
- Kotter, John P. (1995) Leading change: Why transformation efforts fail, *Harvard Business Review*, 73(2), pp. 59-67.
- Kotter, John P. and Dan S. Cohen (2002) *The Heart of Change: Real-Life Stories of How People Change Their Organizations*, Harvard Business School Press.
- Kuroda, Sachiko and Isamu Yamamoto (2012) Impact of overtime regulations on wages and work hours, *Journal of the Japanese and International Economies*, 26(2), pp. 249-262.
- Langer, Ellen and Judith Rodin (1976) The effects of choice and enhanced personal responsibility for the aged: A field experiment in an institutional setting, *Journal of Personality and Social Psychology*, 34(2), pp. 191-198.
- Levy, Santiago (2006) *Progress against Poverty: Sustaining Mexico's PROGRESA-Oportunidades Program*, Brookings.
- ロビンズ, ライオネル (小峯敦・大槻忠史訳) (2016)『経済学の本質と意義』京都大学学術出版会.
- Locke, Edwin A. and Gary P. Latham (1990) *A Theory of Goal Setting & Task Performance*, Prentice-Hall.
- Locke, Edwin A. and Gary P. Latham (2002) Building a practically useful theory of goal setting and task motivation: A 35-year odyssey, *American Psychologist*, 57(9), pp. 705-717.
- Locke, Edwin A. and Gary P. Latham (2006) New directions in goal-setting theory, *Current Directions in Psychological Science*, 15(5), pp. 265-268.
- Lundin, Martin, Perola Öberg, Cecilia Josefsson (2015) Learning from success: Are successful governments role models?, *Public Administration*, 93(3), pp. 733-752.
- Macnamara, Brooke N. and Megha Maitra (2019) The role of deliberate practice in expert

- performance: Revisiting Ericsson, Krampe & Tesch-Römer (1993), *Royal Society Open Science*, 6: 190327, Retrieved August 27, 2019, from <https://doi.org/10.1098/rsos.190327>.
- Madson, Peter M. and Vinit Desai (2010) Failing to learn? The effects of failure and success on organizational learning in the global orbital launch vehicle industry, *Academy of Management Journal*, 53(3), pp. 451-476.
- Manning, Willard G, Joseph P. Newhouse, Naihua Duanm, Emmett B. Keeler, and Arleen Leibowitz (1987) Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment, *American Economic Review*, 77(3), pp. 251-277.
- 松尾貴巳 (2009) 『自治体の業績管理システム』 中央経済社.
- McKillip, Jack (1987) *Need Analysis: Tools for Human Services and Education*, Sage.
- McLaughlin, John A. and Gretchen B. Jordan (2004) Using logic models, In Wholey, Joseph S., Harry P. Hatry, and Kathryn E. Newcomer eds. (2004) *Handbook of Practical Program Evaluation*, 2nd edition, John Wiley & Sons, pp. 7-32.
- Melkers, Julia and John Clayton Thomas (1998) What do administrators think citizens think?: Administrator predictions as an adjunct to citizen surveys, *Public Administration Review*, 58(4), pp. 327-334.
- 宮川公男 (1994) 『政策科学の基礎』 東洋経済新報社.
- 宮本幸平 (2004) 『自治体の財務報告と行政評価』 中央経済社.
- Moher, David, Alessandro Liberati, Jennifer Tetzlaff, Douglas G Altman, and PRISMA Group (2009) Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the PRISMA Statement, *PLOS Medicine*, 6(7): e1000097, Retrieved from July 19, 2019, from <https://doi.org/10.1371/journal.pmed.1000097>.
- Moher, David, Larissa Shamseer, Mike Clarke, Davina Ghera, Alessandro Liberati, Mark Petticrew, Paul Shekelle, Lesley A. Stewart, and PRISMA Group (2015) Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015 Statement, *Systematic Reviews*, 4(1), Retrieved from July 19, 2019, from <https://doi.org/10.1186/2046-4053-4-1>.
- 村松岐夫 (1999) 『行政学教科書—現代行政の政治分析』 第 2 版, 有斐閣.
- 内閣府編 (2018a) 『平成 30 年版男女共同参画白書』 勝美印刷.

- 内閣府編（2018b）『平成 30 年版障害者白書』勝美印刷.
- 内藤和美・山谷清志編（2015）『男女共同参画政策—行政評価と施設評価』晃洋書房.
- 中室牧子・津川友介（2017）『原因と結果の経済学』ダイヤモンド社.
- Nickerson, Raymond S. (1998) Confirmation bias: A ubiquitous phenomenon in many guises, *Review of General Psychology*, 2(2), pp. 175-220.
- 日本都市センター編（2013）『都市自治体の広報分野における課題と専門性—478 市区のアンケート調査結果を通じて—』（日本都市センターブックレット No. 32），日本都市センター.
- 西尾勝（1990）『行政学の基礎概念』東京大学出版会.
- Norton, David P. and Robert S. Kaplan (1992) The balanced scorecard—Measures that drive performance, *Harvard Business Review*, 70(1), pp. 71-79.
- OECD (2010) *Sickness, Disability and Work: Breaking the Barriers*, OECD. (岡部忠信・田中香織訳 [2011]『図表でみるメンタルヘルスと仕事—疾病、障害、仕事の障壁を打ち破る』明石書店.)
- OECD (2017) *Behavioural Insights and Public Policy: Lessons from Around the World*, OECD.
- 大橋弘・金本良嗣・岸本充生・澁谷和久・田辺国昭・八田達夫（2011）「わが国における政策評価：この 10 年を振り返って—パネル討論—I」，阿部顕三・大垣昌夫・小川一夫・田淵隆俊編（2011）『現代経済学の潮流 2011』，東洋経済新報社，第 6 章所収，pp. 157-196.
- 大森彌（2008）『変化に挑戦する自治体—希望の自治体行政学』第一法規.
- 太田聰一（2017）「地域経済が抱える課題と労働市場」，川口大司編『日本の労働市場—経済学者の視点』第 4 章所収，有斐閣，pp. 104-126.
- 大谷信介編著（2002）『これでいいのか市民意識調査—大阪府 44 市町村の実態が語る課題と展望』ミネルヴァ書房.
- 岡檀（2013）『生き心地の良い町—この自殺率の低さには理由（わけ）がある』講談社.
- 奥野正寛編（2008）『ミクロ経済学』東京大学出版会.
- 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』，8(1), pp. 19-38.
- 小野吉昭・倉橋節也（2013）「市町村の情報システムの費用分析」，『経営情報学会』，22(3),

pp. 183-203.

Osborne, David and Ted Gaebler (1992) *Reinventing Government: How the Entrepreneurial Spirit Is Transforming the Public Sector*, Basic Books

Pascale, Richard, Jerry Sternin, and Monique Sternin (2010) *The Power of Positive Deviance: How Unlikely Innovators Solve the World's Toughest Problems*, Harvard Business Review Press.

Patton, Michael Quinn (2012) *Essentials of Utilization-focused Evaluation*, Sage.

Poister, Theodore H. and John Clayton Thomas (2007) The wisdom of crowds: Learning from administrators' predictions of citizen perceptions, *Public Administration Review*, 67(2), pp. 279-289.

Pronin, Emily, Daniel Y. Lin, and Lee Ross (2002) The bias blind spot: Perceptions of bias in self versus others, *Personality and Social Psychology Bulletin*, 28(3), pp. 369-381.

Pronovost, Peter and Eric Vohr (2010) *Safe Patients, Smart Hospitals: How One Doctor's Checklist Can Help Us Change Health Care from the Inside Out*, Hudson Street Press.

Ravallion, Martin and Shaohua Chen (2007) China's (uneven) Progress against poverty, *Journal of Development Economics*, 82, pp. 1-42.

ラヴァリオン, マーティン (柳原透監訳) (2018) 『貧困の経済学』上巻, 日本評論社.

Rawlins, Michael D. (2018) *Evidence-based Medicine*, *Medicine*, 46(7), pp. 388-392.

Redelmeier, Donald A. and Eldar Shafir (1995) Medical decision making in situations that offer multiple alternatives, *Journal of the American Medical Association*, 273, pp. 302-305.

レイブルック, ダーヴィッド・ヴァン (岡崎晴輝, デイミトリ・ヴァン・オーヴェルベーク訳) (2019) 『選挙制を疑う』法政大学出版局.

Ries, Eric (2011) *The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses*, Currency. (井口耕二訳 [2012] 『リーン・スタートアップ・ムダのない起業プロセスでイノベーションを生み出す』日経 BP 社.)

Robbins, Lionel (1935) *An Essay on the Nature and significance of Economic Science*, 2nd ed., Macmillan. (小峯敦・大槻忠史訳 [2016] 『経済学の本質と意義』京都大学学術出版会.)

- Roemer, John E. (1996) *Theories of Distributive Justice*, Harvard University Press. (木谷 忍・川本隆史訳 [2001]『分配的正義の理論—経済学と倫理学の対話』木鐸社.)
- Roese, Neal J. and Kathleen D. Vohs (2012) Hindsight bias, *Perspectives on Psychological Science*, 7(5), pp. 411-426.
- Rosling, Hans, Ola Rosling, Anna Rosling Rönnlund (2018) *Factfulness: Ten Reasons We're Wrong about the World and Why Things Are Better Than You Think*, Sceptre. (上杉周作・関美和訳 [2019]『FACTFULNESS (ファクトフルネス) —10の思い込みを乗り越え、データを基に世界を正しく見る習慣』日経 BP 社.)
- Rossi, Peter H., Mark W. Lipsey, and Howard E. Freeman (2004) *Evaluation: A Systematic Approach*, 7th edition, Sage. (大島巖・平岡公一・森俊夫・元永拓郎監訳 [2005]『プログラム評価の理論と方法—システムティックな対人サービス・政策評価の実践ガイド』日本評論社.)
- Rowntree, Benjamin S. (1901) *Poverty: A Study of Town Life*, McMillan.
- Ryan, Richard M. and Edward L. Deci (2000) Intrinsic and extrinsic motivations: Classic definitions and new directions, *Contemporary Educational Psychology*, 25, pp. 54-67.
- 龍慶昭・佐々木亮 (2000)『「政策評価」の理論と技法』, 増補改訂版, 多賀出版.
- Sackett, David L. (1996) Evidence Based Medicine: What it is and what it isn't, *BMJ Clinical Research*, 312, pp. 71-72.
- Sackett, David L. (1997) Evidence-based medicine, *Seminars in Perinatology*, 21(1), pp. 3-5.
- Sackett, David L., Sharon E. Straus, W. Scott Richardson, William Rosenberg, and R. Brian Haynes (2000) *Evidence-Based Medicine: How to Practice and Teach EBM*, 2nd edition, Churchill Livingstone.
- 坂井豊貴 (2013)『社会的選択理論への招待—投票と多数決の科学』日本評論社.
- 坂井豊貴 (2015)『多数決を疑う—社会的選択理論とは何か』岩波書店.
- 坂之上洋子(2014)「障がい者が自立すれば、みんなが元気になれる一国に NO! を突き付けた岡山県総社市長の戦い(上)」,『東洋経済 ONLINE』, Retrieved July 14, 2019, from <https://toyokeizai.net/articles/-/34321>.
- Salganik, Matthew J. and Karen E. C. Levy (2015) Wiki surveys: Open and qualifiable social data Collection, *PLoS ONE*, 10(5): e0123483, doi: 10.1371/ journal. pone.

023483.

Salganik, Matthew J. (2018) *Bit by Bit: Social Research in the Digital Age*, Princeton University Press.

佐々木亮 (2006) 「ODA 分野における「エビデンスに基づく評価」の試み: 「貧困アクションラボ」の動向」, 『日本評価研究』, 6(1), pp. 43-54.

佐々木亮 (2010) 『評価論理—評価学の基礎』多賀出版.

佐藤郁哉 (2015) 『社会調査の考え方』上巻, 東京大学出版会.

Schultz, T. Paul (2004) School subsidies for the poor: Evaluating the Mexican Progresa poverty program, *Journal of Development Economics*, 74, pp. 199-250.

Schuman, Howard and Stanley Presser (1979) The open and closed question, *American Sociological Survey*, 44(5), pp. 692-712.

Schünemann, Holger, Jan Brożek, Gordon Guyatt, and Andrew Oxman eds. (2013) *Handbook for Grading the Quality of Evidence and the Strength of Recommendations Using the GRADE approach*, updated October 2013 version, Retrieved May 31, 2019, from <https://gdt.gradeapro.org/app/handbook/handbook.html>.

Seligman, Martin E. (1990) *Learned Optimism: How to Change Your Mind and Your Life*, Pocket Books.

Seligman, Martin E. (2011) *Flourish: A Visionary New Understanding of Happiness and Well-being*, Free Press.

Seligman, Martin E. and Mihaly Csikszentmihalyi (2000) Positive psychology: An introduction, *American Psychologist*, 55(1), pp. 5-14.

Sen, Amartya (1985) *Commodities and Capabilities*, North-Holland. (鈴木興太郎訳 [1988] 『福祉の経済学—財と潜在能力』岩波書店.)

センゲ, ピーター・M (守部信之訳) (1995) 『最強組織の法則—新時代のチームワークとは何か』徳間書店.

Siilasmaa, Risto (2018) *Transforming NOKIA: The Power of Paranoid Optimism to Lead Through Colossal Change*, McGraw-Hill Education.

Skoufias, Emmanuel, Benjamin Davis, and Sergio de la Vega (2001) Targeting the poor in Mexico: An evaluation of the selection of households into PROGRESA, *World Development*, 29(10), pp. 1769-1784.

- Slovic, Paul, Melissa L. Finucane, Ellen Peters, and Donald G. MacGregor (2007) The affect heuristic, *European Journal of Operational Research*, 177, pp. 1333-1352.
- Smith, Richard and Drummond Rennie (2014) Evidence based medicine: An oral history, *The Journal of the American Medical Association*, 311(4), pp. 365-367.
- Soman, Dilip and Min Zhao (2011) The fewer the better: Number of goals and savings behavior, *Journal of Marketing Research*, 48, pp. 944-957.
- Stiglitz, Joseph E., Amartya Sen, and Jean-Paul Fitoussi (2009) *Report by the Commission on the Measurement of Economic Performance and Social Progress*, Retrieved July 1, 2019, from <https://ec.europa.eu/eurostat/documents/118025/118123/Fitoussi+Commission+report>.
- Subotnik, Rena F., David E. Karp, and Elizabeth R. Morgan (1989) High IQ children at midlife: An investigation into the generalizability of Terman's Genetic Studies of Genius, *Roeper Review*, 11(3), pp. 139-144.
- 須藤修 (1988) 『ノイズと経済秩序－資本主義の自己組織化』 日本評論社.
- 須藤修 (1995) 『複合的ネットワーク社会－情報テクノロジーと社会進化』 有斐閣.
- Sunstein, Cass R. (2013) *Simpler: The Future of Government*, Simon & Schuster. (田総恵子訳 [2017] 『シンプルな政府－“規制”をいかにデザインするか』 NTT 出版.)
- Sunstein, Cass R. and Reid Hastie (2014) *Wiser: Getting Beyond Groupthink to Make Groups Smarter*, Harvard Business Review Press. (田総恵子訳 [2016] 『賢い組織は「みんな」で決める－リーダーのための行動科学入門』 エヌティティ出版.)
- スタックラー, デヴィッド, サンジェイ・バス (橘明美・臼井美子訳) [2014] 『経済政策で人は死ぬか?－公衆衛生学から見た不況対策』 草思社.
- 鈴木興太郎 (2004) 「センの潜在能力アプローチと福祉国家システムの構想」, 塩野谷祐一・鈴木興太郎・後藤玲子編 『福祉の公共哲学』 公共哲学叢書 5, 第 5 章所収, 東京大学出版会, pp. 73-100.
- Syed, Matthew (2015) *Black Box Thinking: Marginal Gains and the Secrets of High Performance*, Portfolio. (有枝春訳 [2016] 『失敗の科学－失敗から学習する組織、学習できない組織』 ディスカヴァー・トゥエンティワン.)
- 障害福祉研究会編 (2002) 『ICF 国際生活機能分類－国際障害分類改訂版』 中央法規出版.
- 高橋由紀 (2015) 「地方自治体における男女共同参画政策の評価」, 内藤和美・山谷清志編

- 『男女共同参画政策—行政評価と施設評価』晃洋書房, 第6章所収, pp. 98-116.
- 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(5), pp. 254-266.
- タレブ, ナシーム・ニコラス (望月衛監訳, 千葉敏夫訳) (2017) 『反脆弱性—不確実な世界を生き延びる唯一の考え方』, 上巻, ダイヤモンド社.
- 田中啓 (2014) 『自治体評価の戦略—有効に機能させるための16の原則』東洋経済新報社.
- 丹後俊郎 (2002) 『新版メタ・アナリシス入門—エビデンスの統合をめざす統計手法』朝倉書店.
- Terman, Lewis M. and Melita Oden (1959) *The Gifted Group at Mid-Life: Thirty-Five Years' Follow-Up of the Superior Child, Genetic Studies of Genius, Volume V*, Stanford University Press.
- Tetlock, Philip E. (2005) *Expert Political Judgment How Good Is It? How Can We Know?*, Princeton University Press.
- Tetlock, Philip E. and Dan Gardner (2015) *Superforecasting: The Art and Science of Prediction*, Crown. (土方奈美訳 [2016] 『超予測力—不確実な時代の先を読む10カ条』早川書房.)
- Thaler, Richard H. (2015) *Misbehaving: The Making of Behavioral Economics*, W. W. Norton & Company. (遠藤真美訳 [2016] 『行動経済学の逆襲』早川書房.)
- Thaler, Richard H. (2018) Nudge, Not Sludge, *Science*, 361(6401), p. 431.
- Thaler, Richard H. and Cass R. Sunstein (2008) *Nudge: Improving Decisions About Health, Wealth, and Happiness*, Yale University Press. (遠藤真美訳 [2009] 『実践行動経済学—健康、富、幸福への聡明な選択』日経BP社.)
- 遠山真世 (2001) 「障害者雇用政策の3類型—日本および欧米先進国の比較を通して」, 『社会福祉学』, 42(1), pp. 77-86.
- 遠山信幸・百瀬ひろこ・水上由美子・井野隆史 (2007) 「医療の質向上への取り組み: ピアレビューとしての事例検討会の意義」, 『医療の質・安全学会誌』, 2(2), pp. 146-151.
- Townsend, Peter Brereton (1979) *Poverty in the United Kingdom: A Survey of Household Resources and Standards of Living*, Penguin Book.
- Tricomi, Elizabeth, Antonio Rangel, Colin F. Camerer, and John P. O'Doherty (2010)

- Neural evidence for inequality: Averse social preferences, *Nature*, 463, pp.1089-1091.
- 常木淳（2000）『費用便益分析の基礎』東京大学出版会.
- 津富宏（2003）「系統的レビューに基づく社会政策を目指して：キャンベル共同計画の取り組み」．『日本評価研究』, 3(2), pp. 23-39.
- 津谷喜一郎（2000a）「コクラン共同計画とは—日本的展開へ向けて—」, 厚生省健康政策局研究開発振興改良技術情報推進室監修『わかりやすい EBM 講座』厚生科学研究所, pp. 120-135.
- 津谷喜一郎（2000b）「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—」, 『保健医療科学』, 49(4), pp. 313-319.
- Tversky, Amos and Daniel Kahneman（1973）Availability: A heuristic for judging frequency and probability, *Cognitive Psychology*, 5, pp. 207-321.
- Tversky, Amos and Daniel Kahneman（1974）Judgment under uncertainty: Heuristics and biases, *Science*, 185(4157), pp. 1124-1131.
- Tversky, Amos and Daniel Kahneman（1981）The framing of decisions and the psychology of choice, *Science*, 211, pp. 453-458.
- 内田良（2011）「柔道事故と頭部外傷—学校管理下の死亡事例 110 件からのフィードバック」, 『愛知教育大学教育総合開発機構紀要』, 1, pp. 95-103.
- 内田良（2013）『柔道事故』河出書房新社.
- 内田良（2015）『教育という病—子どもと先生を苦しめる「教育リスク」』光文社.
- 上山信一（1998）『「行政評価」の時代—経営と顧客の視点から』NTT 出版.
- 上山信一（2002a）『行政の経営改革—管理から経営へ』第一法規.
- 上山信一（2002b）「行政評価の本質と「科学革命」：わが国自治体の行政評価を手掛かりに」『公共政策研究』, 2(2), pp. 70-81.
- 上山信一・伊関友伸（2003）『自治体再生戦略—行政評価と経営改革』日本評論社.
- Watts, Geoff（2009）Peter Brereton Townsend, *Lancet*, 374, p. 112.
- Weiss, Carol H.（1998）*Evaluation: Methods for Studying Programs and Policies*, 2nd edition, Prentice Hall.（佐々木亮監修, 前川美湖・池田満監訳 [2014]『入門評価学—政策・プログラム研究の方法』日本評論社.）
- Wholey, Joseph S., Harry P. Hatry, and Kathryn E. Newcomer eds.（2004）*Handbook of Practical Program Evaluation*, 2nd edition, John Wiley & Sons.

- 薬師寺明子・清水智美（2013）「「障がい者千人雇用」をめざす総社市における障害者就労支援システムの提案」, 『美作大学・美作大学短期大学部紀要』, 58, pp. 29-40.
- 山口慎太郎（2016）「差の差法で検証する「保育所整備」の効果—社会科学における因果推論の応用」, 岩波データサイエンス刊行委員会編『IWANAMI DATA SCIENCE』, 3, pp. 112-128.
- 山本勲（2015）『実証分析のための計量経済学—正しい手法と結果の読み方』中央経済社.
- 山本徳郎（2013）『教育現場での柔道死を考える—「子どもが死ぬ学校」でいいのか』かもがわ出版.
- 山岡秦幸（2015）『ここまでわかる施策満足度調査—行政の経営品質向上をめざして』ブイツーソリューション.
- 山谷清志（1997）『政策評価の理論とその展開—政府のアカウンタビリティ』晃洋書房.
- 山谷清志（2006）『政策評価の実践とその課題—アカウンタビリティのジレンマ』萌書房.
- 山谷清志（2008）「特集「自治体評価の10年」について」, 『日本評価研究』, 8(1), pp. 1-2.
- 山谷清志（2012）『政策評価』, BASIC 公共政策学第9巻, ミネルヴァ書房.
- 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), pp. 31-45.
- Zeitlin, Marian, Hossein Ghassemi, and Mohamed Mansour (1990) *Positive Deviance in Child Nutrition: With Emphasis on Psychosocial and Behavioural Aspects and Implications for Development*, The United Nations University Press.