

プログラム評価の手法と総合評価の実施状況

東 信 男*

(会計検査院事務総長官房上席研究調査官)

I はじめに

我が国では、平成13年1月に行われた中央省庁等改革を契機に、政策評価制度が各府省に導入されており、その実効性を高めるため、14年4月に行政機関が行う政策の評価に関する法律（以下「政策評価法」という。）が施行されている。この政策評価は、各府省がその所掌する政策について自ら評価を行うことが基本となっており、事業評価方式、実績評価方式、総合評価方式等を選択的に用いることとされている。このうち総合評価は、特定のテーマについて、当該テーマに係る政策効果の発現状況を様々な角度から掘り下げて分析し、政策に係る問題点を把握するとともにその原因を分析したりなどする評価方式¹⁾である。この総合評価は、制度として我が国に初めて導入された評価方式であるが、14年度から政策評価法に基づいた評価書が府省の一部で作成されている。

一方、米国では、1960年代から各連邦省庁を中心にプログラム評価（Program Evaluation）が行われている。このプログラム評価は、社会科学の手法を用いて政策の効果を体系的に分析する評価方式で、我が国の総合評価に相当すると考えられている。米国では、法律により各連邦省庁にランダム実験モデル等の分析手法を用いてプログラム評価を行うことを義務付けている政策があるため、体系的な評価理論に基づいた評価手法や定量的な分析手法が開発されるとともに、実務でも採用されている。

総合評価では、政策効果の発現状況を具体的に明らかにすることが求められているため、プログラム評価の手法は、各府省が総合評価を行う上で参考になると考えられる。本稿は、プログラム評価の手法を紹介するとともに、各府省が14年度と15年度に行った総合評価の実施状況を、プログラム評価の手法に対応させて整理したものである。（本稿は、すべて筆者の個人的見解であり、筆者が属する会計検査院の公式見解を示すものではない。）

*1956年生まれ。1980年横浜国立大学経済学部卒業、1986年ロチェスター大学経営大学院修士課程終了（MBA）。1980年会計検査院採用、その後、文部検査第2課総括副長、通商産業検査課総括副長、大蔵検査課決算監理官などを経て2001年より現職。この間、1990～1993年在ニューヨーク総領事館出向、2003年名古屋大学経済学部講師併任。

1) 「政策評価に関する基本方針（平成13年12月28日閣議決定）」（別紙）〔総合評価方式〕。

II プログラム評価の手法²⁾

1 プログラム評価の理論的背景

プログラム評価の手法を理解するためには、その理論的背景を知ることが不可欠である。プログラム評価では、各府省の政策を次のようにとらえている。

(1) 政策体系

各府省の政策は、国民生活や社会経済に発生した課題を解決するために実施していることから、目的手段の関係としてとらえると「政策（狭義）→施策→事務事業」から構成されている。ここで①政策（狭義）とは特定の行政課題を解決するための基本的な目標、②施策とは政策（狭義）を達成するための具体的な目標で、組織的な行政活動のまとめり、③事務事業とは施策を達成するための具体的な政策手段で、個々の行政活動のことである。各府省の政策は、多くの場合、一つの政策（狭義）に複数の施策があり、また、一つの施策に複数の事務事業があるため、ピラミット型の3層構造になっている。このように政策（狭義）、施策及び事務事業は、相互に目的と手段の関係を保ちながら、全体として一つの政策体系を構成している。

(2) セオリー

各府省の政策は、その目標を達成するまでの過程が整合性の取れた理論（セオリー）に基づいて設計されている。このセオリーとは、原因と結果が連鎖上に連なる「仮定」のことである。つまり、原因1が結果1を引き起こすと、この結果1は次の原因2となって次の結果2を引き起こしていく。さらに、この結果2は、次の次の原因3となって次の次の結果3を引き起こしていく。政策が目標を達成するまでの過程は、この連鎖状に連なる「仮定」に基づいて設計されており、この仮定のどこかがうまく機能しなければ、政策は途中で破綻し、目標とする改善効果までたどり着かないことになる。

(3) ロジック・モデル

各府省の政策は、セオリーによる因果関係としてとらえると「投入（インプット）→活動→結果（アウトプット）→（外部要因）→成果（アウトカム）」の各過程を経て国民生活や社会経済に影響を及ぼす。この一連の流れをプログラム評価では、ロジック・モデルと呼んでいる。ここで①インプットとは行政活動を実施するために必要な資源の投入、②活動とはアウトプットを提供するための組織的な行政活動、③アウトプットとは国民生活や社会経済に発生した課題を解決するために提供する行政サービス、④外部要因とはアウトカムに影響を及ぼす当該アウトプット以外の要因、⑤アウトカムとはアウトプットが国民生活や社会経済に及ぼす改善効果のことで、個別的なもの（下位アウトカム）と包括的なもの（上位アウトカム）がある。このようにインプット、活動、アウトプット及びアウトカムは、相互に原因と結果の関係を保ちながら、全体として一つのロジック・モデルを構成している。

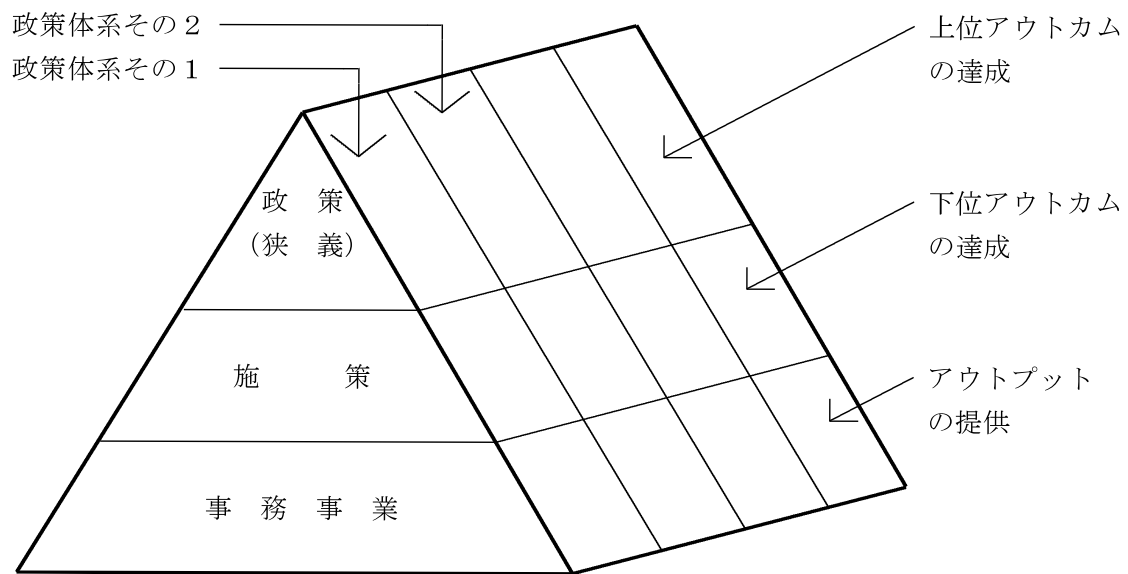
(4) 政策体系とロジック・モデル

各府省の政策を、政策体系とロジック・モデルの関係からとらえると、事務事業は個々の行政活動であ

2) IIについては、龍慶昭・佐々木亮（2004）に負うところが大きい。

り、施策を達成するための手段となっていることから、アウトプットは事務事業が提供する行政サービス、下位アウトカムは施策が目標とする改善効果に対応している。一つの施策に複数の事務事業がある場合、下位アウトカムを達成するために複数のアウトプットが提供されることになる。また、施策は政策（狭義）を達成するための手段となっていることから、上位アウトカムは政策（狭義）が目標とする改善効果に対応している。一つの政策（狭義）に複数の施策がある場合、上位アウトカムを達成するために複数の下位アウトカムが国民生活や社会経済に改善効果を及ぼすことになる（政策体系とロジック・モデルの関係については、図1参照）。

図1 政策体系とロジック・モデルの関係



2 プログラム評価の評価手法

施策³⁾の効果が出ていない場合、その原因として①施策が整合性の取れた形で設計されていないこと、②施策が設計通りに実施されていないこと、③施策が有効な政策手段を用いていないことが考えられる。また、施策がその成果以上の費用を投入していることも考えられる。プログラム評価では、このような事態を明らかにするため、次の4つの評価手法を採用している。

(1) セオリー評価

セオリー評価は、各府省が施策を立案するに当たり、政策目的と政策手段の関係を明確にした政策体系を設計しているかどうか、また、原因と結果の連鎖関係について、整合性の取れたロジック・モデルを設計しているかどうかということの評価する手法である。この評価手法は、ロジック・モデルが整合性の取れた形で設計されなければ、アウトカムは達成されないという考え方に基づいている。なぜなら、インプットからアウトカムに至る仮定の連鎖がうまく機能しなければ、施策はたとえ実施されたとしても途中で破綻し、目標とする改善効果までたどり着けないからである。セオリー評価でマイナスの評価が出た場合、ロジック・モデルに関して見直しを行うことになる。

(2) プロセス評価

プロセス評価は、各府省が施策を実施するに当たり、事前に設計したロジック・モデルに従って、計画した量・質のアウトプットを計画したタイミングで提供しているかどうかということの評価する手法である。この評価手法では、ロジック・モデルのうち「インプット→活動→アウトプット」までの過程に焦点が当てられる。この評価手法は、ロジック・モデルが整合性の取れた形で設計されたとしても、計画された量・質のアウトプットが計画されたタイミングで提供されなければ、アウトカムは達成されないという考え方に基づいている。プロセス評価でマイナスの評価が出た場合、インプットの量・質・タイミングや行政活動の方法・内容に関して見直しを行うことになる。

(3) インパクト評価

インパクト評価は、各府省の実施した施策が国民生活や社会経済に改善効果（インパクト）を与えたかどうかということの評価する手法である。この評価手法では、ロジック・モデルのうち「アウトプット→(外部要因)→アウトカム」までの過程に焦点が当てられる。この評価手法は、計画された量・質のアウトプットが計画されたタイミングで提供されたとしても、アウトプットがアウトカムを達成する手段として有効でなければ、アウトカムは達成されないという考え方に基づいている。インパクト評価でマイナスの評価が出た場合、アウトプットの量・質・タイミングに関して見直しを行ったり、政策手段であるアウトプットそのものの見直しを行うことになる。

(4) コスト・パフォーマンス評価

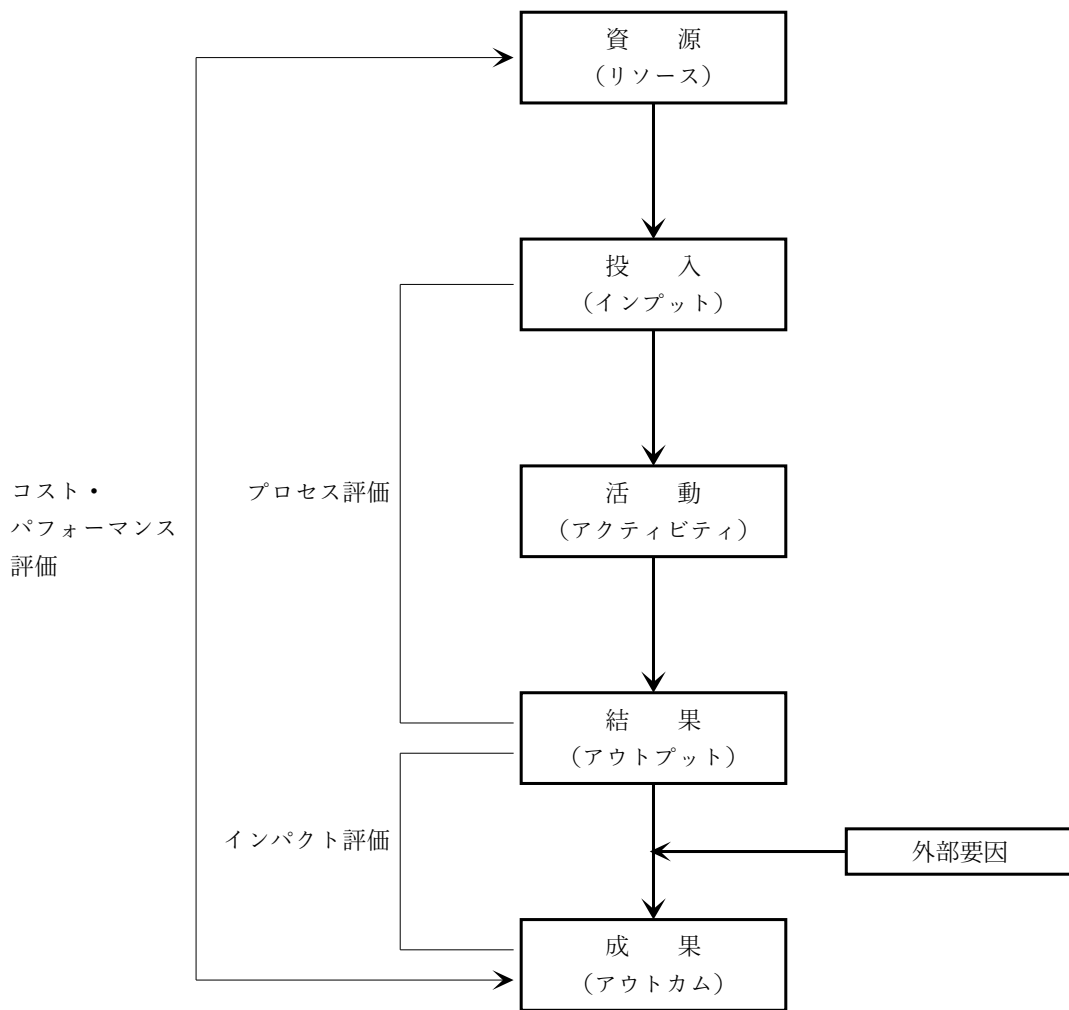
コスト・パフォーマンス評価は、各府省の実施した施策が投入した資源以上の改善効果を国民生活や社会経済に与えたかどうかということの評価する手法である。この評価手法は、アウトプットが国民生活や社会経済にインパクトを与えたとしても、それに見合う以上の費用が掛かったのであれば、費用対効果の観

3) プログラム評価は、政策体系を構成しているどの階層の政策も評価の対象にできるが、Ⅱの2以下では、事務事業を政策手段とする施策を評価の対象とする場合を想定している。

点からは優れた政策手段とはいえないという考え方に基づいている。コスト・パフォーマンス評価でマイナスの評価が出た場合、アウトプットの量・質・タイミングに関して見直しを行ったり、政策手段であるアウトプットそのものを見直しを行うことになる（ロジック・モデルと評価手法の関係については、図2参照）。

以下、各評価手法ごとに、それぞれの評価項目及び分析手法について具体的に紹介することとしたい。

図2 ロジック・モデルと評価手法



(注) セオリー評価は、ロジック・モデルそのものを具体的に把握し、その整合性を検証することが、評価の内容となる。

3 セオリー評価の評価項目

セオリー評価では、「インプット→活動→アウトプット→(外部要因)→アウトカム」で構成されるロジック・モデルを具体的に把握し、原因と結果の連鎖関係について、整合性の取れたロジック・モデルが設計されていたかどうかということを検証することが、評価の内容となる。セオリー評価では、以下の項目を計画ベースで把握し、フローチャートを作成することが不可欠である。

(1) アウトカム

政策目的と政策手段の関係を明確にした政策体系が設計されていることを確認するとともに、評価の対象となる施策の具体的な目標を把握する。プログラム評価では、最終的に業績指標を用いて施策の達成状況を定量的に測定するため、施策は定量的に測定可能なレベルまで特定化かつ具体化されている必要がある。また、施策の達成状況を定量的に測定するための業績指標（アウトカム業績指標）を把握する。このアウトカム業績指標とは、国民生活や社会経済の変化に着目した定量的な評価指標のことである。

(2) アウトプット

施策の政策手段となる事務事業の全体像を計画ベースで把握する。具体的には、政策手段となる事務事業の名称に加え、この事務事業が提供するアウトプットについて①実施主体、②実施地域、③実施期間、④受益者の資格と受益者総数、⑤量と質（アウトプット業績指標）を調査する。ここで「量」とはアウトプットの規模のことであり、「質」とはアウトプットの内容のことである。例えば、公共事業であれば「量」が処理能力・収容能力等で、「質」が規格・構造等のことである。また、融資事業であれば「量」が貸付件数・貸付総額等で、「質」が1件当たりの貸付限度額・貸付金利等のことである。複数の事務事業を政策手段とする施策については、それぞれの事務事業について全体像を把握する。

(3) 活動

アウトプットを提供するための組織的な行政活動を計画ベースで把握する。具体的には、事務事業の実施主体が、受益者に対してアウトプットを提供するために行う一連の業務の流れを調査する。例えば、社会保障事業であれば「保険料の徴収→給付申請の受付→資格審査→給付金の支給」、公共事業であれば「計画→設計→積算→契約→施工→施設の運用」、融資事業であれば「貸付財源の調達→貸付申請の受付→貸付審査→貸付実行→債権の管理」という業務が考えられる。

(4) インプット

アウトプットを提供するために必要な資源の量・質を計画ベースで把握する。具体的には、事務事業の実施主体が、受益者に対してアウトプットを提供するために投入する①事業費等の資金的資源、②職員等の人的資源、③勤務時間等の時間的資源、④施設等の物的資源、⑤受益者ニーズ等の情動的資源を調査する。ここで「量」とは資源の規模のことであり、「質」とは資源の内容のことである。例えば、資金的資源であれば「量」が事業費総額で、「質」が国庫補助金・自己負担金等の内訳である。また、人的資源であれば「量」がスタッフ総数で、「質」が役職・資格別等の内訳である。事業費等の資金的資源については、可能な限り直接事業費だけではなく、人件費、管理部門経費等の間接事業費も含めて包括的に把握することが重要である。

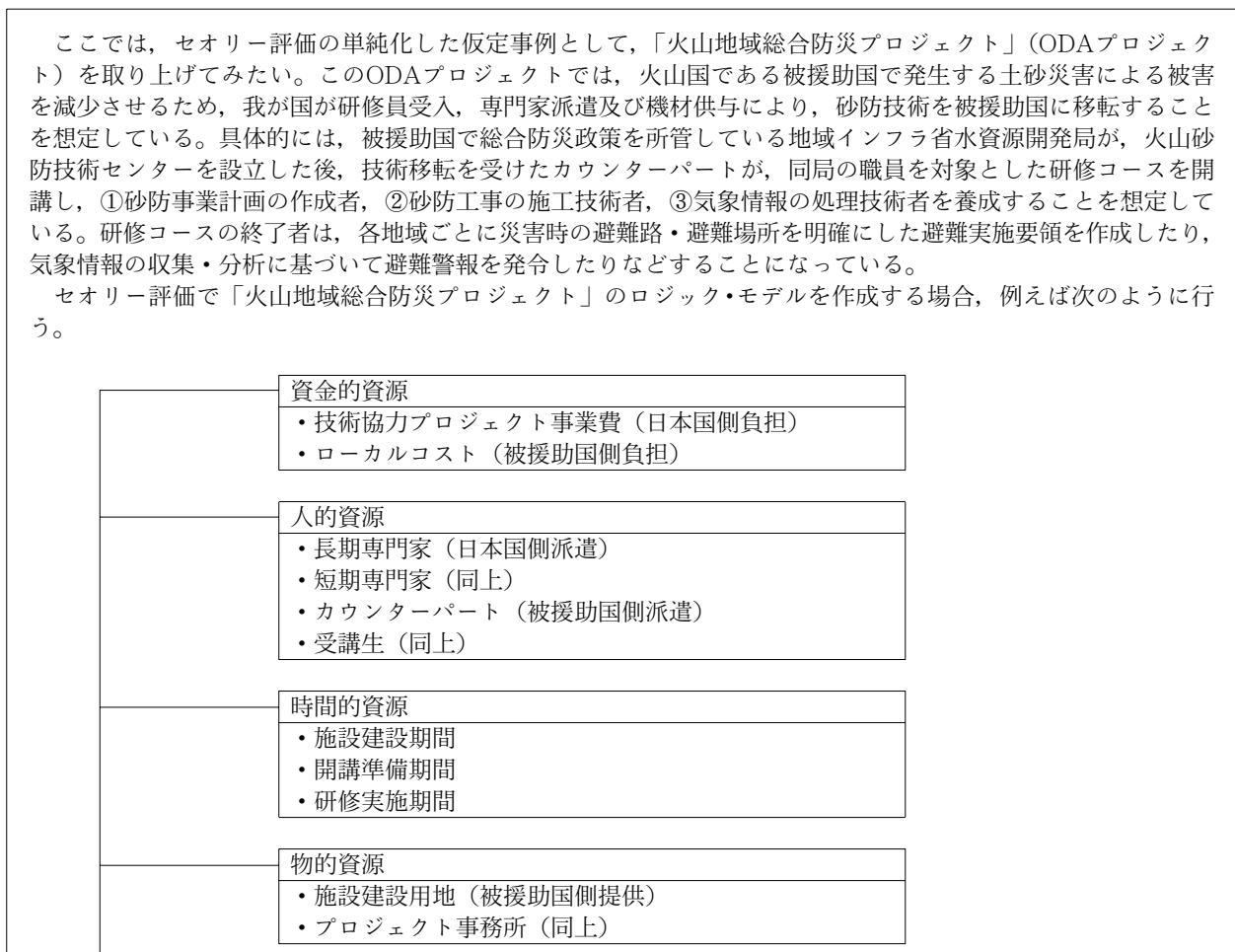
(5) プロセス・セオリー

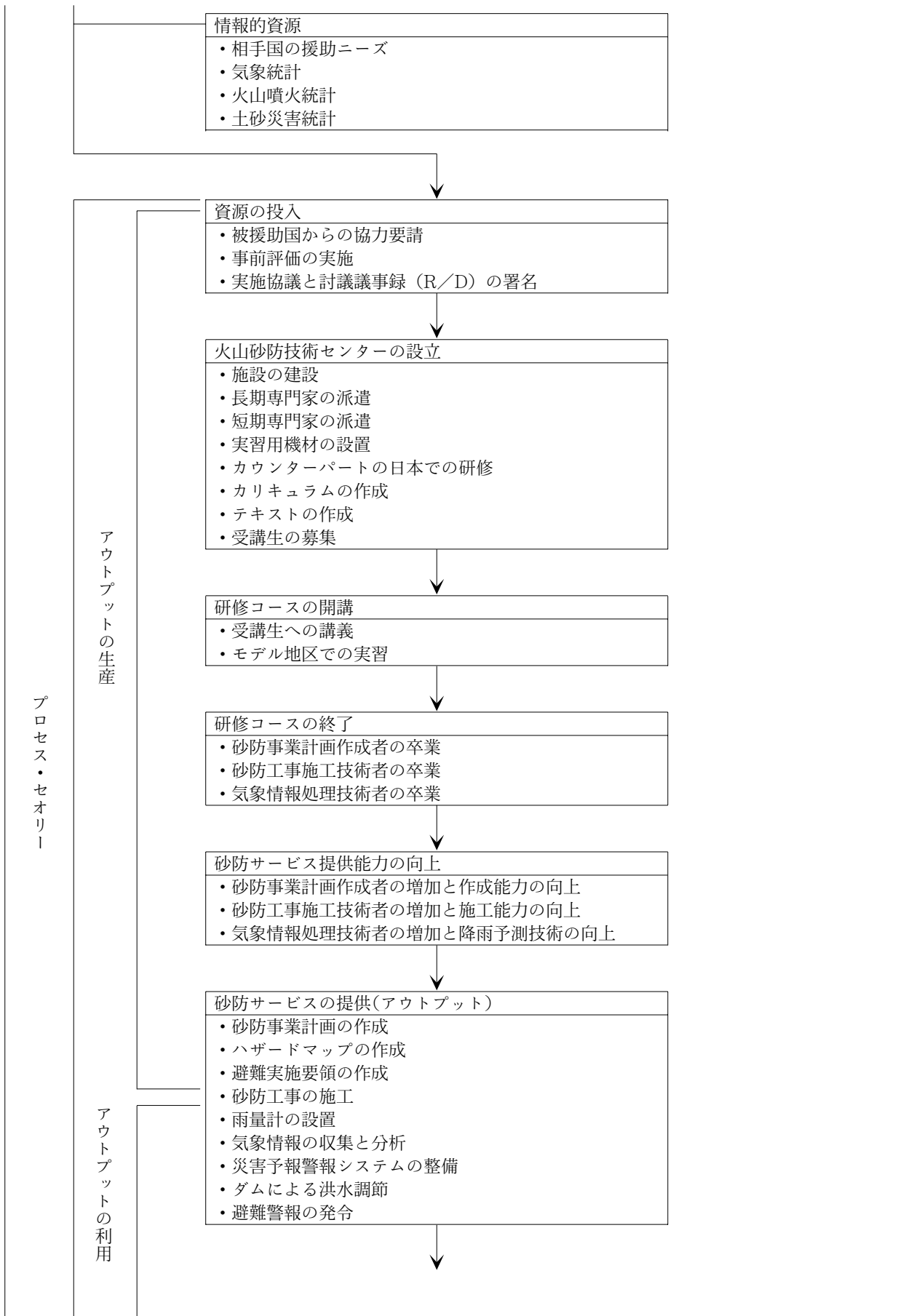
インプット、活動及びアウトプットの相互関係を把握する。具体的には、アウトプットの生産段階と利用段階に分けて調査する。アウトプットのプロセスとは、アウトプットを行政内部に生産するため、資源が「インプット→活動」の過程の中で、どのように組み合わせられ、どのような順序で、または、どのようなタイミングで投入されることになっていたのかということである。また、アウトプットの利用段階とは、アウトプットが「活動→アウトプット」の過程の中で、どのようなタイミングで、どのような経路をたどって受益者に利用されることになっていたのかということである。アウトプットは、行政内部に生産されたとしても、必ずしも受益者に利用されるとは限らないため、生産段階と利用段階を明確に分けることが重要である。

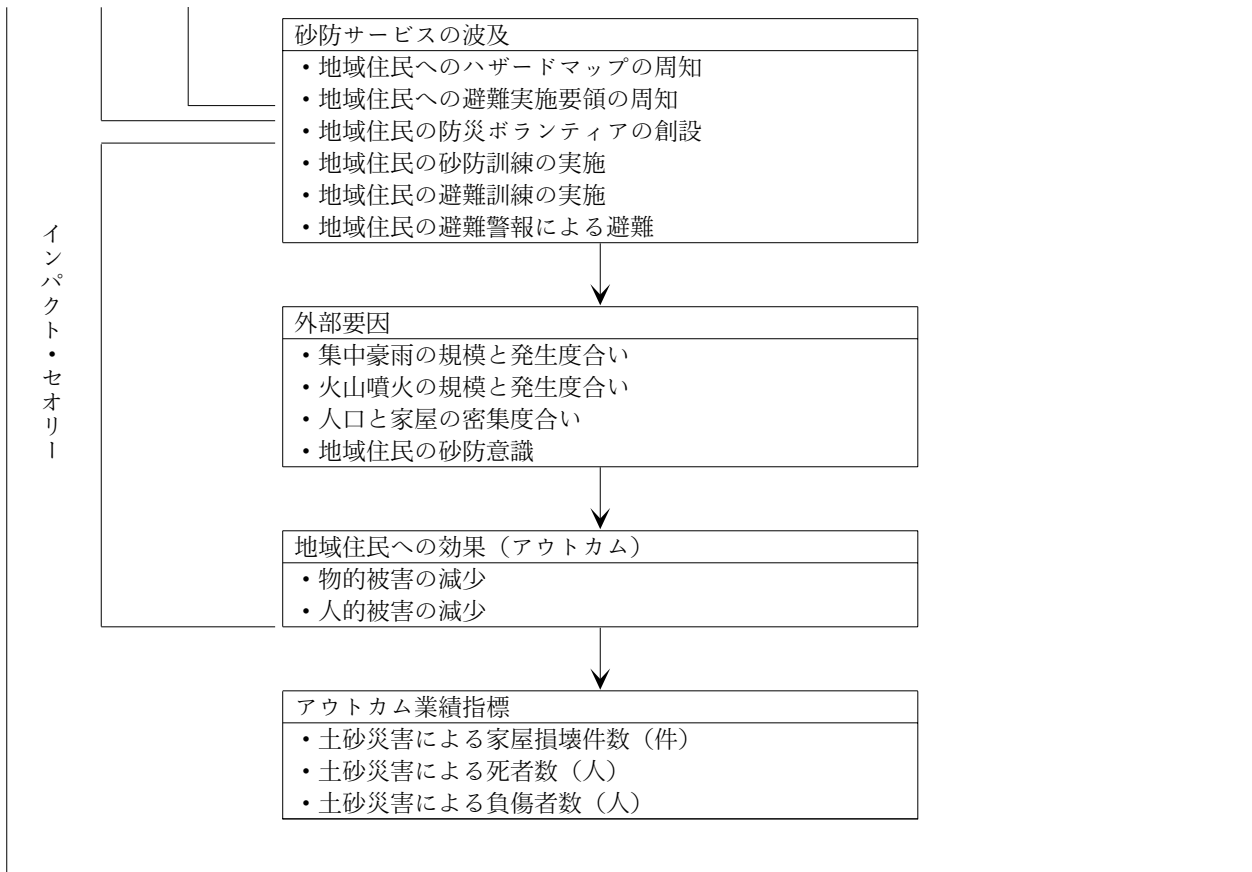
(6) インパクト・セオリー

アウトプットとアウトカムとの相互関係を把握する。具体的には、提供されたアウトプットが「アウトプット→(外部要因)→アウトカム」の過程の中で、どのような経路をたどって国民生活や社会経済に改善効果を及ぼすことになっていたのかということ調査する。また、改善効果に影響を及ぼす外部要因として、どのようなものが想定されていたのかということ調査する(ロジック・モデルをフローチャートにした仮定事例については、図3参照)。

図3 ロジック・モデルの仮定事例







4 プロセス評価の評価項目

プロセス評価では、ロジック・モデルのうち「インプット→活動→アウトプット」までの過程の実績を把握し、アウトプットが計画通りに提供されているかどうかということを検証することが、評価の内容となる。プロセス評価では、以下の項目を実績ベースで把握し、計画と比較することが不可欠である。

(1) インプットの投入結果

資源が、計画通りに投入されたことを検証する。計画した量・質の資源が計画したタイミングで投入されないと、計画した行政活動が行えないことになる。例えば、図3の仮定事例の場合、被援助国がローカルコスト分の財政資金を調達できなかったり、施設建設用地を提供できなかったりすると、火山砂防技術センターを建設できないことになる。

(2) 行政活動の結果

資源の投入後、実施主体の行政活動が、計画通りに行われたことを検証する。計画した行政活動が行われないと、計画した量・質のアウトプットを計画したタイミングで生産できないことになる。例えば、図3の仮定事例の場合、火山砂防技術センターが竣工しても、地域インフラ省が、研修コースの受講者を募集しなかったり、研修コースの修了者を他の局に配置換えしたりすると、各地域ごとに災害時の避難路・避難場所を明確にした避難実施要領を作成できないことになる。

(3) アウトプットの生産結果

行政活動の結果、アウトプットが、計画通りに生産されたことを検証する。計画したアウトプットが行政内部に生産されないと、受益者が計画した量・質のアウトプットを計画したタイミングで利用できないことになる。例えば、図3の仮定事例の場合、火山砂防技術センターの研修コース修了者が、避難実施要領を作成しなかったり、気象情報の収集・分析に基づいて避難警報を発令しないと、地域住民は災害時に砂防サービスを利用できないことになる。

(4) アウトプットの利用結果

アウトプットが生産された後、受益者が、計画通りに利用したことを検証する。受益者が、計画した量・質のアウトプットを計画したタイミングで利用しないと、国民生活や社会経済に改善効果を及ぼさないことになる。例えば、図3の仮定事例の場合、避難実施要領が作成されたり、気象情報の収集・分析に基づいて避難警報が発令されても、地域住民が災害時にこれらに従って避難しないと、土砂災害を減少できないことになる。

5 インパクト評価の分析手法

インパクト評価では、各府省の実施した施策が国民生活や社会経済に改善効果を与えたかどうかということを検証するため、ロジック・モデルのうち「アウトプット→(外部要因)→アウトカム」までの過程を分析してインパクトを測定することが、評価の内容となる。インパクト評価では、施策の対象となり得る母集団を、施策が実際に実施されたグループ(実施グループ)と施策が実施されなかったグループ(比較グループ)に分け、これらの二つのグループの間で差が生じているかどうかということ測定する。また、比較グループが存在しない場合、実施グループの実施前の状況と実施後の状況において差が生じているかどうかということ測定する。いずれの場合も、両者の差から外部要因の影響をいかに取り除くかがポイントになる。これらの差を測定する手法として、次のような分析手法がある(インパクト評価の分析手法については、図4参照)。

図4 インパクト評価の分析手法

分析手法	比較グループの有無	比較グループを設定する方法	比較グループを設定する時期	分析結果の信頼性
受益者評価	無	/	/	↓ 低い
単純事前・事後比較モデル	無			
パネル・モデル	無			
時系列モデル	無			
クロスセクション・モデル	無			
準実験モデル	一般指標モデル	有	統計的	↓ 低い
	統計的等化モデル	有	統計的	
	マッチング・モデル	有	統計的	
	回帰・分断モデル	有	統計的	
ランダム実験モデル	有	無作為	施策実施前	↓ 高い

出所) 龍慶昭・佐々木亮(2004) p51の表を参考に筆者が加筆修正した。

(1) 受益者評価

ア 手法の内容

受益者評価（Beneficiaries Judgement）は、実施グループの受益者に施策実施後、記憶に基づいて施策が実施されなかった場合の状況を想定させたり、アウトカム業績指標の予想値を設定させた後、これらを現在の状況や、アウトカム業績指標の実績値と比較させ、受益者が認識した差を施策のインパクトと見なす分析手法である。この分析手法は、受益者を対象としたアンケート調査、インタビュー等により実施される。

イ 技術的な限界

受益者評価は、簡便であるため広範囲にわたり採用されているが、受益者の記憶は、あいまいで不正確であることが多いため、分析手法としての信頼性は高くない。また、調査対象者の抽出方法、調査内容の設問方法、調査の回答方法等が変わると、分析結果が大きく異なることがある。

(2) 単純事前・事後比較モデル

ア 手法の内容

単純事前・事後比較モデル（Simple Before-After Comparison Model）は、同一の実施グループに関して、施策実施前後の各1時点においてアウトカム業績指標の実績値を測定し、2時点間の実績値の差を施策のインパクトと見なす分析手法である。測定の時点として、施策実施の直前と直後を設定するのが原則であるが、季節変動要因を考慮して前年同月比とする場合もある。

イ 技術的な限界

単純事前・事後比較モデルは、簡便であるため広範囲にわたり採用されているが、アウトカム業績指標に生じる外部要因の影響を取り除けないため、分析手法としての信頼性は高くない。この分析手法を採用する前提条件は、もし施策を実施しなければ、アウトカム業績指標の事前と事後の実績値が同じ水準になるであろうと十分に予想できる場合である。このような前提条件は、事前と事後の測定時点が比較的開いていない場合に満たされる。また、施策のインパクトは、実施直後に現れるとは限らないため、インパクトが現れるまで一定の期間を要すると見込まれる施策については、採用できない。

(3) パネル・モデル

ア 手法の内容

パネル・モデル（Panel Study）は、固定された同一の実施グループに関して、施策実施前後の一定期間においてアウトカム業績指標の実績値を測定し、実績値に現れる差を施策のインパクトと見なす分析手法である。この分析手法は、インパクトが施策の実施直後には現れないと見込まれる場合、採用されることが多い。

イ 技術的な限界

パネル・モデルは、アウトカム業績指標の実績値を一定期間にわたって測定する必要があるが、調査対象者が転居等で途中で脱落したりなどしてサンプル数が減少すると、測定結果に偏りが生じるため、分析結果の信頼性は低下してしまう。また、このモデルは、調査対象者の所在を一定期間にわたって把握することが前提となるため、大きな労力と高いコストが必要になる。さらに、施策が実施されないグループとの比較を行わないため、アウトカム業績指標に生じる外部要因の影響を取り除けない。

(4) 時系列モデル

ア 手法の内容

時系列モデル (Interrupted Time Series Model) は、同一の実施グループに関して、施策実施前後の長期間においてアウトカム業績指標の実績値を測定し、実績値のトレンドに現れる差を施策のインパクトと見なす分析手法である。実績値のトレンドは、表やグラフにより視覚的に分析したり、回帰分析により統計的に分析する。この分析手法は、アウトカム業績指標のトレンドや規則性を把握することにより、アウトカム業績指標の見込値を予測することができる。

イ 技術的な限界

時系列モデルは、アウトカム業績指標の実績値を長期間にわたって測定する必要があるが、各測定時点において、調査項目、調査方法、調査対象者等の測定方法が変わると、測定結果に影響が生じるため、分析結果の信頼性は低下してしまう。また、施策が実施されないグループとの比較を行わないため、アウトカム業績指標に生じる外部要因の影響を取り除けない。

(5) クロスセクション・モデル

ア 手法の内容

クロスセクション・モデル (Cross-Section Model) は、複数の実施グループに関して、施策実施後の1時点においてアウトプット業績指標とアウトカム業績指標の実績値を測定して両者の相関関係を把握し、相関度の高さを施策のインパクトと見なす分析手法である。相関度の高さは、表やグラフにより視覚的に分析したり、回帰分析により統計的に分析する。サンプルとなる実施グループ数が少なく、かつ、都市圏、農村圏など地域により人口数に変動があるため、規模の大きな実施グループのデータが分析結果に大きな影響を与える場合、アウトカム業績指標の実績値を一定の人口当たり (例えば1000人当たり) に揃えて相関関係を把握することがある。

イ 技術的な限界

クロスセクション・モデルは、実際に提供された行政サービスの量と質のばらつきを利用してインパクトを測定するため、一定数以上 (例えば25以上) の実施グループについてデータを収集する必要がある。また、このモデルは、政策手段と改善効果の相関関係を把握することはできるが、施策が実施されないグループとの比較を行わないため、政策手段と改善効果の因果関係を把握したり、改善効果から外部要因の影響を取り除けない。

(6) 一般指標モデル

ア 手法の内容

一般指標モデル (Generic Control Model) は、施策の対象となり得る母集団の平均値 (全国平均値、全県平均値、全市平均値等) を比較グループと見なしている。この母集団の中には、施策が実施されたグループと施策が実施されなかったグループが含まれている。そして、実施グループと母集団に関して、施策実施後の一定期間 (又は1時点) においてアウトカム業績指標の実績値を測定し、実施グループの実績値と母集団の平均値に現れる差を施策のインパクトと見なす分析手法である。この分析手法は、母集団が外部要因から影響を受けているということは、実施グループも同様の影響を受けているはずであるという仮定に基づいている。

イ 技術的な限界

一般指標モデルは、全国平均値、全県平均値等の一般指標が存在している行政分野でのみ、採用可能である。また、比較グループとした母集団が、外部要因についてどれだけ実施グループと近似しているかどうかということの詳細に分析していないため、分析手法としての信頼性は高くない。さらに、実施グループ全体の母集団に占める割合が高くなるにつれ、実施グループの実績値と母集団の平均値に現れる差は縮小していくため、施策が広範囲に実施されるにつれ、分析結果の信頼性は低下していく傾向にある。

(7) 統計的等化モデル

ア 手法の内容

統計的等化モデル (Statistically Equated Model) は、実施グループと比較グループを設定するため、施策の実施後に施策が実施された母集団と施策が実施されなかった母集団から、それぞれ共通の条件を有する部分を統計的に抽出して実施グループと比較グループに割り当てる作業を行う。そして、実施グループと比較グループに関して、施策実施後の一定期間 (又は1時点) においてアウトカム業績指標の実績値を測定し、両グループの実績値の差を施策のインパクトと見なす分析手法である。

イ 技術的な限界

統計的等化モデルは、比較の対象として施策が実施されないグループを必要とするため、全国で一律に行われる施策については、採用できない。また、実施グループと比較グループは、完全に等質ではないため、外部要因の影響を完全に取り除けない。

(8) マッチング・モデル

ア 手法の内容

マッチング・モデル (Matching Model) は、実施グループと比較グループを設定するため、施策の実施前に施策を実施する実施グループを特定するとともに、この実施グループと共通の条件 (マッチング指標) を有する比較グループを特定し、この比較グループには施策を実施しない。そして、実施グループと比較グループに関して、施策実施後の一定期間 (又は1時点) においてアウトカム業績指標の実績値を測定し、両グループの実績値の差を施策のインパクトと見なす分析手法である。この分析手法では、施策の実施前に実施グループと比較グループが、アウトカム業績指標とマッチング指標の両方について近似値を有していることを確認する必要がある。

イ 技術的な限界

マッチング・モデルは、比較の対象として施策が実施されないグループを必要とするため、全国で一律に行われる施策については、採用できない。また、実施グループと比較グループは、完全に等質ではないため、外部要因の影響を完全に取り除けない。さらに、施策の実施前に実施グループと比較グループを設定しなければならぬため、既に実施された施策については、採用できない。

(9) 回帰・分断モデル

ア 手法の内容

回帰・分断モデル (Regression-Discontinuity Model) は、実施グループと比較グループを設定するため、施策の実施前に施策の対象となり得る母集団を、アウトカム業績指標の実績値に基づいて一定の基準値 (カットオフ・ポイント) より大きいグループと小さいグループに分ける。次に、カットオフ・ポイン

トより小さいグループはアウトカムの改善が必要で、カットオフ・ポイントより大きいグループはアウトカムの改善が必要ないと見なして、小さいグループ（実施グループ）に施策を実施し、大きいグループ（比較グループ）には施策を実施しない（アウトカム業績指標の性格により、逆の場合もある）。そして、実施グループと比較グループに関して、施策実施後の1時点においてアウトカム業績指標の実績値を測定し、カットオフ・ポイントにおける両グループの実績値の回帰線の差を施策のインパクトと見なす分析手法である。この分析手法では、施策の実施前に実施グループと比較グループのアウトカム業績指標の実績値の回帰線がカットオフ・ポイントで連続していることを確認する必要がある。

イ 技術的な限界

回帰・分断モデルは、比較の対象として施策が実施されないグループを必要とするため、全国で一律に行われる施策については、採用できない。また、実施グループと比較グループは、完全に等質ではないため、外部要因の影響を完全に除外できない。さらに、施策の実施前に実施グループと比較グループを設定しなければならないため、既に実施された施策については、採用できない。

(10) ランダム実験モデル

ア 手法の内容

ランダム実験モデル (Randomized Experimental Model) は、実施グループと比較グループを設定するため、施策の実施前に施策の対象となり得る母集団から無作為にサンプル集団を抽出するとともに、このサンプル集団を無作為に施策を実施する実施グループと施策を実施しない比較グループに割り当てる作業を行う。実施グループと比較グループへの割り当ては、無作為に行われるため、実施グループと比較グループは、施策がこれから実施されるかどうかを除いて、ほぼ等質になっている。そして、実施グループと比較グループに関して、施策実施後の一定期間（又は1時点）においてアウトカム業績指標の実績値を測定し、両グループの実績値の差を施策のインパクトと見なす分析手法である。この分析手法は、施策の実施後に二つのグループに生じた違いが、すべて施策実施の有無によって引き起こされたと判断できるため、最も信頼性の高い分析手法とされている。

イ 技術的な限界

ランダム実験モデルは、比較の対象として施策が実施されないグループを必要とするため、全国で一律に行われる施策については、採用できない。また、施策の実施前に実施グループと比較グループを設定しなければならないため、既に実施された施策については、採用できない。さらに、実施グループへの参加について潜在的な受益者の同意が得られなかったり、実施グループの参加者が施策実施中に脱落したりなどしてサンプル数が減少すると、測定結果に偏りが生じるため、分析結果の信頼性は低下してしまう。

6 コスト・パフォーマンス評価の分析手法

コスト・パフォーマンス評価では、各府省の実施した施策が投入した資源以上の改善効果を国民生活や社会経済に与えたかどうかということを検証するため、施策の費用対効果を測定することが、評価の内容となる。この場合、施策の改善効果をいかに貨幣価値に換算するかがポイントになる。施策の費用対効果を測定する手法として、次のような分析手法がある。

(1) 費用効果分析

ア 手法の内容

費用効果分析（Cost-Effectiveness Analysis）は、施策の実施により発生した社会的効果と社会的費用を特定した後、必ずしもすべてを貨幣価値に換算しないで社会的効果と社会的費用を算定し、様々な単位（金額、人数、件数、時間等）で表示された効果と費用の比較を行う分析手法である。この分析手法には、費用のみを貨幣価値に換算して効果一単位当たりの費用又は費用一単位当たりの効果を算出する方法、効果と費用の双方について貨幣価値以外の表示も含めた数値指標を用いる方法がある。また、効果項目や費用項目が複数ある場合、それらの項目に適当な重み付けを行って一つの指標に統合する方法もある。施策にとって、複数の政策手段を相互に比較して他の条件が同一であれば、効果一単位当たりの費用の最小なもの又は費用一単位当たりの効果の最大なものが、費用対効果の観点から最も優れた政策手段ということになる。

イ 技術的な限界

費用効果分析は、効果と費用が必ずしも貨幣価値に換算されないため、分析結果だけで当該施策が投入した資源以上の改善効果を国民生活や社会経済に与えたかどうかを検証することができない。施策を達成するための政策手段が複数あった場合、これらの政策手段相互間の相対的なコスト比較や効果比較ができるだけである。

(2) 費用便益分析

ア 手法の内容

費用便益分析（Cost-Benefit Analysis）は、施策の実施により発生した社会的便益と社会的費用を特定した後、すべての便益項目と費用項目を貨幣価値に換算して比較する分析手法である。便益や費用が一定期間にわたって継続的に発生する場合、便益や費用を現在価値に割り戻した上で比較を行う。比較の方法には、社会的便益の社会的費用に対する比率を用いる方法、社会的便益から社会的費用を差し引いた純社会的便益を用いる方法等がある。施策は、その社会的便益が社会的費用を上回った場合、費用対効果の観点から社会的合理性を有すると見なされる。

イ 技術的な限界

費用便益分析は、施策の実施により発生した社会的便益と社会的費用を貨幣価値に換算する必要があるが、施策の実施により発生する社会的便益と社会的費用の中には、貨幣価値に換算することが困難なものもあるため、すべての便益項目と費用項目を算定できないことがある。特に社会的便益は、行政サービス特有の性格により貨幣価値への換算に困難が伴いがちで、計測の前提条件、計測の対象項目、計測の手法等が変わると、分析結果が大きく異なることがある。

Ⅲ 総合評価の実施状況

1 総合評価への取組状況

我が国では、13年1月に行われた中央省庁等改革を契機に、政策評価制度が各府省に導入されたが、その実効性を高めるため、13年6月に政策評価法が制定された。これを受け、政府は13年12月に政策評価に関する基本方針（以下「基本方針」という。）を閣議決定し、この中で各府省は評価方式として、事業評価方式、実績評価方式、総合評価方式等を選択的に用いることとされた。その後14年4月に政策評価法が施行され、各府省は政策評価を行った場合、評価の対象とした政策、用いた評価方式、政策効果の分析手法

等について記載した評価書を作成することとなった。この作成された評価書⁴⁾に基づき、各府省の14年度と15年度における総合評価への取組状況について現状と課題をまとめると、次のようになっている(表1参照)。

表1 総合評価の実施件数

(単位：件)

府 省	基本計画	14年度	15年度	計
内閣府	○	1	—	1
宮内庁	—	—	—	—
公正取引委員会	○	1	1	2
国家公安委員会・警察庁	○	—	—	—
防衛庁	○	16	11	27
金融庁	○	—	—	—
総務省	○	—	1	1
公害等調整委員会	—	—	—	—
法務省	○	1	—	1
外務省	○	120	126	246
財務省	○	1	—	1
文部科学省	○	2	—	2
厚生労働省	○	1	3	4
農林水産省	○	1	—	1
経済産業省	—	—	—	—
国土交通省	○	11	8	19
環境省	—	—	—	—
計	13府省	155	150	305

注)「基本計画」欄の○印は、基本計画で総合評価を評価方式の一つと位置付けている府省である。

4) 評価書は、各府省のWebサイトで公表されている。Ⅲの1以下は、筆者が各府省のWebサイトから入手した評価書に基づいて整理したものである。

(1) 現 状

- (ア) 政策評価法の対象となっている17府省のうち、それぞれの基本計画⁵⁾で総合評価を評価方式の一つと位置付けているのは、13府省⁶⁾となっている。
- (イ) 実際に総合評価を実施したのは、14年度が10府省、15年度が6府省となっている。このうち、14年度又は15年度に総合評価を実施したのは11府省で、2年度続けて総合評価を実施したのは5府省となっている。
- (ウ) 総合評価の全体の実施件数は、14年度が155件、15年度が150件で、両年度の合計は305件となっている。
- (エ) 総合評価の各府省ごとの実施件数は、ばらつきが大きく、14年度と15年度の合計305件のうち、外務省が246件（80.7%）、防衛庁が27件（8.9%）、国土交通省が19件（6.2%）で、他の8府省はそれぞれ4件以下となっている。
- (オ) 総合評価は、各府省が自ら行う場合とその全部又は一部を外部のシンクタンク等に委託する場合がありますが、総合評価の委託件数は、14年度が8件、15年度が5件で、両年度の合計は13件となっている。

(2) 課 題

ア 総合評価の実施の必要性

大部分の府省では、基本方針に規定された3つの評価方式のうち、実績評価を中核的な評価方式と位置付けている。この実績評価は、あらかじめ政策効果に着目した達成すべき目標を設定し、これに対する実績を測定したりなどして目標の達成度合いについて評価する評価方式⁷⁾である。しかし、この実績評価では、目標が達成されていないと評価されても、その原因について、ロジック・モデルの整合性が取れていないことにあるのか、政策が計画通りに実施されていないことにあるのか、または、有効な政策手段が用いられていないことにあるのか、明らかにすることができない。さらに、目標が達成されたとしても、投入された資源以上の改善効果を国民生活や社会経済に与えたのかどうか、明らかにすることができない。このような分析は、総合評価を行って初めて解明できるため、総合評価を実施していない又は実施件数の少ない府省については、総合評価を積極的に実施する必要があると考えられる。

イ テーマの重点化の必要性

府省の中には、総合評価の対象として、テーマの重点化を図ることなく毎年度、所掌する政策を網羅的に取り上げているところがある。その評価内容を見ると、評価対象とした政策について行政内部の活動状況を記述しているのが大部分であり、掘り下げた分析はほとんど行われていない。総合評価は本来、特定のテーマについて、当該テーマに係る政策効果の発現状況を様々な角度から掘り下げて分析し、政策に係る問題点を把握するとともにその原因を分析したりなどする評価方式であり、毎年度、所掌する政策を網羅的に取り上げることは予定されていない。これは、総合評価でセオリー評価、インパクト評価、コスト・パフォーマンス評価等の評価手法を用いて掘り下げた分析を行うためには、かなりのコストと時間を要するからである。したがって、総合評価の対象として毎年度、所掌する政策を網羅的に取り上げている府省については、特定のテーマを重点的に取り上げ、セオリー評価、インパクト評価、コスト・パフォーマンス

5) 各府省が、政策評価法第6条に基づき、基本方針を踏まえながら、評価の目的、評価の方式、評価の観点など、政策評価に関する基本的事項を定めたものである。

6) 国土交通省では、総合評価に対応するものをプログラム評価と称している。

7) 基本方針（別紙）〔実績評価方式〕。

ス評価等の評価手法を採用して掘り下げた分析を行う必要があると考えられる。

ウ 政策を取り上げる必要性

府省の中には、総合評価の対象として、政策そのものではなく、組織運営のあり方を取り上げているところがある。具体的には、評価の対象として、再任用制度等の人事、官舎の整備等の福利厚生、秘密保全体制等の服務規律等を取り上げている。しかし、政策評価法は、政策を評価の対象と規定しており⁸⁾、組織運営のあり方そのものについては、評価の対象には含めていない。これは、組織運営のあり方が、個別の政策を実施する上でのインフラとなっているため、最終的には政策効果の発現状況に反映されることになるからである。したがって、総合評価の対象として、組織運営のあり方を取り上げている府省については、政策そのものを取り上げる必要があると考えられる。

2 総合評価の評価手法

各府省が総合評価で採用した評価手法は、厳密な意味で本来のプログラム評価の評価手法と同一ではないものもあるが、ロジック・モデルのどの過程に焦点を合わせているかによって、プログラム評価の評価手法に近似的に置き換えることが可能である。そこで、各府省が14年度と15年度に行った305件の総合評価で採用した評価手法を、Ⅱの2で紹介したプログラム評価の評価手法に対応させて整理すると、次のようになっている(表2参照)。

表2 総合評価の評価手法の内訳

(単位：件)

評価手法	14年度	15年度	計
セオリー評価	18	29	47(15.4)
プロセス評価	153	131	284(93.1)
インパクト評価	34	28	62(20.3)
コスト・パフォーマンス評価	6	3	9(3.0)
計	211	191	402

注1) 1件の総合評価で複数の評価手法を採用している場合があるため、横計は表1の横計と一致しない。

注2) () は表1の総合評価の実施件数の計305件で除した割合で、単位は%である。

1件の総合評価で複数の評価手法を採用している場合があるため、合計は100%にならない。

8) 政策評価法第2条第2項。

(1) 現 状

- (ア) セオリー評価は、ロジック・モデルの整合性に焦点を当てた評価手法であるが、47件（15.4％）の総合評価で採用されている。ただし、これら47件の評価内容は、政策目的と政策手段の関係をフローチャートや表にただけのものが大部分であり、Ⅱの3で紹介したように、「インプット→活動→アウトプット→（外部要因）→アウトカム」で構成されるロジック・モデルを具体的に把握し、原因と結果の連鎖関係について、整合性の取れたロジック・モデルが設計されていたかどうかということを検証したものは、ほとんど見受けられない。
- (イ) プロセス評価は、ロジック・モデルへの準拠性に焦点を当てた評価手法であるが、284件（93.1％）の総合評価で採用されている。ただし、これら284件の評価内容は、インプットの投入結果、行政活動の結果、アウトプットの生産結果等を記述しただけのものが大部分であり、Ⅱの4で紹介したように、アウトプットが計画通りに提供されているかどうかということを検証したものは、ほとんど見受けられない。
- (ウ) インパクト評価は、政策手段の有効性に焦点を当てた評価手法であるが、62件（20.3％）の総合評価で採用されている。これら62件のインパクト評価では、Ⅱの5で紹介した分析手法のいずれかが採用されている。
- (エ) コスト・パフォーマンス評価は、政策の費用対効果に焦点を当てた評価方式であるが、9件（3.0％）の総合評価で採用されている。これら9件のコスト・パフォーマンス評価では、Ⅱの6で紹介した分析手法のいずれかが採用されている。
- (オ) 総合評価では、セオリー評価、プロセス評価、インパクト評価及びコスト・パフォーマンス評価が同時に採用されることが理想であるが、6件（2.0％）の総合評価でこれら4つの評価手法が同時に採用されている。

(2) 課 題

ア セオリー評価の採用と充実の必要性

セオリー評価は、ロジック・モデルの整合性を検証するとともに、その過程で他の評価手法にとって不可欠な情報を生み出すため、プログラム評価の第一段階と位置付けられている。なぜなら、ロジック・モデルの整合性が取れていないと、政策は計画通りに実施されたとしても、目標とする改善効果までたどり着けないからである。また、プロセス評価では実績と計画の比較が行われるが、この計画に関する情報は、セオリー評価により把握できるからである。さらに、インパクト評価ではアウトカム業績指標を用いて改善効果が測定されるが、このアウトカム業績指標に関する情報は、セオリー評価により把握できるからである。したがって、各府省は、ロジック・モデルの整合性を検証するとともに、他の評価手法にとって不可欠な情報を把握するため、総合評価において、セオリー評価の採用割合を高めるとともに本来のセオリー評価を実施する必要があると考えられる。

イ プロセス評価の充実の必要性

プロセス評価は、ほとんどの総合評価で採用されているが、その評価内容を見ると、インプットの投入結果、行政活動の結果、アウトプットの生産結果等を記述しただけのものが大部分である。Ⅱの4で紹介したように、インプットの量・質・タイミングについて実績と計画を比較したり、実施主体の行政活動について実績と計画を比較したりなどして、最終的にアウトプットが計画通りに提供されているかどうかということを検証したものは、ほとんど見受けられない。このように各府省のプロセス評価の大部分は、本

来のプロセス評価にはなっていないため、評価とは呼べないものとなっている。したがって、各府省は、ロジック・モデルへの準拠性を検証するため、総合評価において、本来のプロセス評価を実施する必要があると考えられる。

ウ インパクト評価の採用の必要性

インパクト評価は、政策が国民生活や社会経済に改善効果を与えたかどうかということの評価するために採用される手法である。これは、政策が仮に計画通りに実施されたとしても、有効な政策手段が用いられていなければ、国民生活や社会経済に改善効果を与えることができないからである。つまり、プロセス評価だけでは、政策が計画通りに実施されたことを検証できても、政策の効果が発現していることを保証することはできないのである。したがって、各府省は、政策手段の有効性を検証するため、総合評価において、インパクト評価の採用割合を高める必要があると考えられる。

エ コスト・パフォーマンス評価の採用の必要性

コスト・パフォーマンス評価は、政策が投入した資源以上の改善効果を国民生活や社会経済に与えたかどうかということの評価するために採用される手法である。これは、政策が投入した資源以上の改善効果を国民生活や社会経済に与えていなければ、費用対効果の観点から社会的合理性を有していないからである。また、一つの政策目的に複数の政策手段がある場合、限られた財源の範囲内で政策の効果を高めるためには、費用対効果のより高い政策手段により多くの資源を投入する必要があるからである。つまり、インパクト評価だけでは、政策の効果が発現していることを検証できても、当該政策の継続性を保証することはできないのである。したがって、各府省は、政策の費用対効果を検証するため、総合評価において、コスト・パフォーマンス評価の採用割合を高める必要があると考えられる。

3 インパクト評価の分析手法

各府省が総合評価のインパクト評価で採用した分析手法は、プログラム評価のインパクト評価で採用される分析手法と同一になっている。そこで、各府省が14年度と15年度に行った62件のインパクト評価で採用した分析手法を、Ⅱの5で紹介した分析手法に対応させて整理すると、次のようになっている（表3参照）。

表3 インパクト評価の分析手法の内訳

(単位：件)

分析手法		14年度	15年度	計
受益者評価		16	13	29 (46.8)
単純事前・事後比較モデル		11	13	24 (38.7)
パネル・モデル		—	2	2 (3.2)
時系列モデル		13	11	24 (38.7)
クロスセクション・モデル		—	1	1 (1.6)
準実験モデル	一般指標モデル	4	2	6 (9.7)
	統計的等化モデル	1	4	5 (8.1)
	マッチング・モデル	—	—	—
	回帰・分断モデル	—	—	—
ランダム実験モデル		—	—	—
計		45	46	91

注1) 1件のインパクト評価で複数の分析手法を採用している場合があるため、横計は表2のインパクト評価の件数と一致しない。

注2) () は表2のインパクト評価の計62件で除した割合で、単位は%である。1件のインパクト評価で複数の分析手法を採用している場合があるため、合計は100%にならない。

(1) 現 状

- (ア) 受益者評価は、29件（46.8%）のインパクト評価で採用されている。この受益者評価は、分析結果の信頼性が最も低いと見なされているが、インパクト評価で採用される割合が一番高くなっている。
- (イ) 単純事前・事後比較モデルは、24件（38.7%）のインパクト評価で採用されている。
- (ウ) パネル・モデルは、2件（3.2%）のインパクト評価で採用されている。
- (エ) 時系列モデルは、24件（38.7%）のインパクト評価で採用されている。表やグラフによる視覚的分析は21件で、回帰分析による統計的分析は3件となっている。
- (オ) クロスセクション・モデルは、1件（1.6%）のインパクト評価で採用されている。これは、グラフによる視覚的分析となっている。
- (カ) 一般指標モデルは、6件（9.7%）のインパクト評価で採用されている。この一般指標モデルは、全国平均又は全産業との比較となっている。
- (キ) 統計的等化モデルは、5件（8.1%）のインパクト評価で採用されている。
- (ク) 一般的に、比較グループを設定しない分析手法の採用割合が高くなっている。また、比較グループを設定する分析手法のうち、政策実施前に比較グループを設定する分析手法の採用は、皆無となっている。

(2) 課 題

インパクト評価では、アウトカム業績指標に関して、実施グループと比較グループに現れる差、または、実施グループの政策実施前後に現れる差が、政策手段の有効性と見なされている。しかし、この差の中には、外部要因の影響が含まれることがあるため、分析結果の信頼性を高めるためには、この外部要因の影響を取り除く分析手法を採用することが重要になる。

ア 比較グループの設定の必要性

外部要因の影響を取り除く第一段階は、可能な限り、比較グループを設定する分析手法を採用することである。なぜなら、実施グループは、政策実施の有無を除いて他の要素を長期間にわたって等質に保つことができないからである。各府省のインパクト評価の分析手法を見ると、全国で一律に行われていない政策を評価対象にしている場合でも、つまり、比較グループを設定する分析手法が採用可能な場合でも、比較グループを設定する分析手法は、一部のインパクト評価でしか採用されていない。したがって、各府省は、分析結果の信頼性を高めるため、全国で一律に行われていない政策については、比較グループを設定する分析手法を採用する必要があると考えられる。

イ 政策実施前の設定の必要性

外部要因の影響を取り除く第二段階は、可能な限り、政策実施前に比較グループを設定する分析手法を採用することである。なぜなら、政策実施後では、比較グループが政策実施前の状態において、実施グループと等質になっていたことを確認できない場合が多いからである。各府省のインパクト評価の実施主体は、外部のシンクタンク等に委託されている場合を除き、ほとんど政策実施部局で実施されている。このため、政策実施前に比較グループを設定する分析手法を採用することが可能な状況になっている。したがって、各府省は、分析結果の信頼性を高めるため、政策実施部局がインパクト評価を行う場合、政策実施前に比較グループを設定する分析手法を採用する必要があると考えられる。

4 コスト・パフォーマンス評価の分析手法

各府省が総合評価のコスト・パフォーマンス評価で採用した分析手法は、プログラム評価のコスト・パフォーマンス評価で採用される分析手法と同一になっている。そこで、各府省が14年度と15年度に行った9件のコスト・パフォーマンス評価で採用した分析手法を、IIの6で紹介した分析手法に対応させて整理すると、次のようになっている(表4参照)。

表4 コスト・パフォーマンス評価の分析手法の内訳

(単位：件)

分析手法	14年度	15年度	計
費用効果分析	1	1	2 (22.2)
費用便益分析	6	2	8 (88.9)
計	7	3	10

注1) 1件のコスト・パフォーマンス評価で複数の分析手法を採用している場合があるため、横計は表2のコスト・パフォーマンス評価の件数と一致しない。

注2) () は表2のコスト・パフォーマンス評価の計9件で除した割合で、単位は%である。1件のコスト・パフォーマンス評価で複数の分析手法を採用している場合があるため、合計は100%にならない。

(1) 現状

(ア) 費用効果分析は、2件(22.2%)のコスト・パフォーマンス評価で採用されている。

(イ) 費用便益分析は、8件(88.9%)のコスト・パフォーマンス評価で採用されている。

(2) 課題

4つの評価手法のうちコスト・パフォーマンス評価の採用割合が最も低くなっている理由の一つは、政策の社会的便益を貨幣価値に換算することが困難であるため、費用便益分析を採用することができないことにあると考えられる。しかし、政策の社会的便益を貨幣価値に換算できなくても、インパクト評価で複数の政策手段が同一の改善効果をもたらすことを検証できれば、これらの政策手段について費用効果分析を行うことは可能である。これにより、費用対効果のより高い政策手段により多くの資源を配分することが可能になる。したがって、各府省は、インパクト評価で複数の政策手段が同一の改善効果をもたらしたことを検証できた場合、費用対効果のより高い政策手段を把握するため、コスト・パフォーマンス評価で費用効果分析を行う必要があると考えられる。

IV おわりに

我が国の政策評価では、事業評価方式、実績評価方式、総合評価方式等を用いることとされている。このうち実績評価については、比較的簡易な評価方式であるため、各府省の所掌する主要な政策を網羅的に取り上げることができることから、各府省の政策評価において、中核的な評価方式に位置付けられている。しかし、実績評価は、目標の達成度合いについての情報を提供することはできても、目標が達成されていない場合の発生原因についての情報を提供することはできないし、政策手段の費用対効果についての情報

を提供することもできない。これらは、総合評価でセオリー評価、プロセス評価、インパクト評価及びコスト・パフォーマンス評価を行うことにより、解明できるのである。

このように総合評価は、実績評価を補完する評価方式として重要であると考えられるが、各府省の実施状況を見ると、プロセス評価が中心で、しかも本来のプロセス評価に該当しないものがほとんどである。したがって、各府省は、プログラム評価の手法を参考にしながら、総合評価の充実を図る必要があると考えられる。

[参考文献]

佐々木亮 (2003) 「プログラム評価の実践例」『NIRA』2003 VOL.16 No.5

佐々木亮 (2003) 『政策評価トレーニング・ブック』多賀出版

田辺智子 (2002) 「政策評価の手法 - アメリカの評価理論と実践をもとに -」『季刊行政管理研究』No.97

田辺智子 (2003) 「プログラム評価の手法概観」『NIRA』2003 VOL.16 No.5

山田治徳 (2000) 『政策評価の技法』日本評論社

龍慶昭・佐々木亮 (2004) 『「政策評価」の理論と技法 (増補改訂版)』多賀出版

Rossi, Peter H., Lipsey, Mark W. and Freeman, Howard E. (2004) Evaluation: A Systematic Approach, 7th edition, Thousand Oaks: Sage Publications

United States General Accounting Office (1991) Designing Evaluations (GAO/PEMD-10.1.4)

United States General Accounting Office (1998) Performance Measurement and Evaluation-Definitions and Relationships (GAO/GGD-98-26)

Weiss, Carol H. (1998) Evaluation: Methods for Studying Programs and Policies, 2nd edition, Upper Saddle River: Prentice-Hall

Wholey, Joseph S., Hatry, Harry P. and Newcomer, Kathryn E. eds. (1994) Handbook of Practical Program Evaluation, San Francisco: Jossey-Bass

