

## 政策評価と社会政策実験

藤本幸生\*

### Policy Evaluation and Social Policy Experiments

Sachio Fujimoto\*

It has a general tendency that the political process of Japan, judging from the management cycle, has made more light of Evaluation process than Planning and Practice process. However, the policy evaluation will be strengthened in all governmental organizations setting the period for the reorganization of the Ministries and the Agencies on January 2001. On the other hand, the social policy experiments are grown from the natural science field, but there are less examples in Japan than in U.S.A. and European countries, because of this experiments have the ethical problems.

In this paper, I will discuss about the concept and some techniques of the social policy experiments with a few examples, and the relation between the policy evaluation and the policy experiments.

#### Key Words (キーワード)

Policy evaluation (政策評価), Social policy experiments (社会政策実験), Efficiency evaluation (効率性評価), Effectiveness evaluation (有効性評価), Randomized true experiments (無作為実験法)

#### 1 はじめに

政策評価は、社会政策の成果について信頼性のある有効な情報を提供するという意味で重要な社会情報過程であると考えられる。しかし、マネージメント・サイクルとしてみた場合にわが国の政策過程においては、企画立案や実施の段階に比べて評価段階が軽視されてきたことは否めない。<sup>1)</sup>

行政の簡素化・効率化を目指して制定された「中央省庁等改革基本法（平成10年法律第103号）」に則り、2001年1月の省庁統廃合に向けて現在準備が行われているが、基本法第4条（改革の基本方針）の6には「国民的視点に立ち、かつ、内外の社会経済情勢の変化を踏まえた客観的な政策評価機能を強化するとともに、評価の結果が政策に

適切に反映されるようにすること。」が盛り込まれた。そこで、この期に合わせて、新組織の全府省に政策評価を導入する方向で検討が進められている。現在検討されている政策評価は各府省による所管事業に関する評価と、新総務省による政府全体の統一的、総合的な評価に分けることができる。この内、各府省における政策評価の指針として「政策評価に関する標準的ガイドラインの案（平成12年7月31日）」が公表され、平成12年度末を目処に最終案が確定される予定であるが、この案で示された政策評価の目的は、(1) 国民に対する行政の説明責任を徹底すること、(2) 国民本位の効率的で質の高い行政を実現すること、(3) 国民的視点に立った成果重視の行政への転換を図ること、とされている。<sup>2)</sup>

\*呉大学社会情報学部 (Faculty of Social Information Science, Kure University)

本稿では、こうした流れを踏まえて、政策評価手法について概観するとともに、政策評価の有力な手法ではあるが、わが国ではあまり実施例が見られない社会政策実験について、各手法の考え方およびその政策評価との関連について検討する。

## 2 政策評価手法の概要

### 2.1 政策評価の視点とその評価手法

政策を「評価する」ということは、その政策を実施することによって意図した目的が達成されたか、もしくは意図した目的にどれほど近づいたか、について計測し、検討することを意味している。言い換えれば、その政策を実施した場合と実施しない場合の「状況の変化」の度合いを測定・評価することによって、政策の成果を把握することである。したがって、政策評価では「状況の変化」をいかに捉えられるかが課題であり、それを捉える方法として、(1)政策の実施費用との関連において、費用に見合った成果が得られたかどうかという観点から、その効率性 (efficiency) に着目した手法 (以下では「効率性評価」という。)、(2)政策が実施された場合とされなかった場合の状況を計測・比較して期待された成果が達成されたかどうかという観点から、その有効性 (effectiveness) に着目した手法 (以下では「有効性評価」という。)、(3)その他の手法、に大別することができる。

### 2.2 効率性評価手法

効率性評価とは、実施された政策がその実施費用に見合うだけの十分な便益をもたらしたかどうか、また同じ目的を達成するのに他の政策よりも単位当たりの費用が安いかどうか、などについて計測・評価することである。この種の政策評価の代表的な手法として費用便益分析 (Cost-Benefit Analysis) がある。公共政策では一般的に便益を特定することが難しいので、便益を測定する方法が開発できるかどうかがこの評価手法の実用性を大きく左右することになる。

便益とは、政策を実施することによって一定期

間に生じる効用 (経済的側面および非経済的側面を含む。) を可能な限り貨幣タームに換算したものである。一方、費用は同一期間に投入される金額であるが、機会費用の概念がその基本となる。便益も費用も客観的な貨幣タームという形で評価することが重要であり、どこまで客観化できるか、さらに測定が難しい2次的波及効果 (外部性) をいかに把握し、便益として換算するかという課題がある。

政策実施がもたらす便益度に着目した効率性評価の場合には、社会全体としての効用の最大化を目的とする経済的効率性を追求した基準によって評価される。すなわち、この評価法は、道路、治山治水、土地改良などの公共事業における経済的効果の計測のように便益が量的に把握できる政策に応用される場合が多い。

政策は通常、単年度というよりは多年度にわたって実施される場合が一般的であるし、その効用が生じる期間も長期にわたる場合が多い。しかし、政策評価では政策の策定時あるいは実施後すぐに、多年度に及ぶ便益と費用を把握して比較しなくてはならない。そのために通常、割引現在価値基準、費用便益比率法、内部収益率法などの方法が用いられる。

費用便益分析によって政策の評価を行う場合には、一般的に次のような条件が必要とされている。

- 1) 政策が単独もしくは別々に資金調達されており、費用が他の活動から区別できること。
- 2) 政策の純便益が十分に大きいこと。
- 3) 政策の効果およびその大きさが認識され、推定可能であること。
- 4) 便益を貨幣タームで評価できること。
- 5) 政策の実施担当者が、単に既存政策を継続するか否かを考えるのではなく、他の政策と比較することの重要性を認識していること。

一方、便益が簡単には貨幣タームに置き換えることができない政策に関しては、直接的便益評価法、あるいは、間接的便益評価法が用いられる。これらの方法は、政策の選択もしくは評価において一定の条件をつけた評価法であり、総称して仮

想評価法 (Contingent Valuation Method : CVM) と呼ばれる。

直接的便益評価法の代表的な手法には、支払意思額法 (Willingness To Pay:WTP) とか、受入補償額法 (Willingness To Accept : WTA) があり、政策対象者からの正直な意思表示を前提とした調査に基づいている。例えば、前者は「大気汚染を減少するという提案に関して、あなたが自発的に支払う最大額はいくらですか」、後者は「大気汚染が増加するという事態に直面した場合に、あなたは最小限いくら貰いたいですか」、または「この提案を阻止するためにあなたが支払う最大額はいくらですか」などのように金額そのものを聞く方法で、結果として疑似市場需要曲線を得ることができる。上記のように直接金額を聞くのではなく、順序付けやランク付けを聞く方法を仮想ランク付け法 (Contingent Ranking Method: CRM) と言い、結果として推定効用関数を得ることができる。

間接的便益評価法とは、観察される需要関数や費用関数、財や投入要素の価格変化、非市場活動における変化、などの要因を用いて便益を推定する方法である。例えば、レクリエーションのような非市場活動の便益を家計支出や旅行費用などから間接的に推定する方法である。代表的な手法としてはヘドニック価格法 (Hedonic Price Method) と旅行費用法 (Travel Cost Method) がある。ヘドニック価格法は、環境の質など貨幣タームで表すことが難しい公共財の価値を不動産価格など市場で観察される価格から推定しようとする方法である。また、旅行費用法は、例えばレクリエーション施設を無料の環境財と考えて、この施設に来るための旅行費用から環境需要額を推定しようとする手法である。

### 2.3 有効性評価手法

政策の有効性に着目した評価においては、本来ならば、「状況の変化」は同一の個人あるいは集団が政策というインパクトの対象となった場合と対象とならなかった場合との影響の大きさの差に

よって政策の成果が表される。政策対象となった場合の影響の大きさは事後的に把握することは可能であるが、政策対象となった同じ個人または集団が、政策対象とならなかったと仮定した場合の影響の大きさを把握することは通常不可能であり、これを如何に推定するかが有効性評価の中心的課題になる。

この場合の有効性とは政策毎に設定された評価尺度 (必ずしも1つとは限らないし、定性的尺度も含む。) に従って測定される概念である。有効性評価が対象とする政策は、影響の大きさを量的に把握できる政策はもちろんであるが、むしろ社会保障、福祉、教育、環境など成果の数量化が比較的難しい社会政策においてその真価が発揮される。

有効性評価は、一般的に特定された個別の政策について、その影響を可能な限り掘り下げて把握することを目的としており、対象とする政策の種類、政策実施対象の種類・範囲・選定可能性、その他社会的・経済的条件によって多様な評価手法が研究されている。有効性評価は、基本的には、(1) 実際に政策実験を行う方法と、(2) それ以外の方法とに大別することができる。実施された政策を実体験した人達からその効果に関して直接的な評価を得ることが可能であれば、政策評価手法としては極めて精度の高い結果を得ることができ、政策評価のための有効な判断情報となり得る。もちろん、この方法をすべての政策に適用することには限界があり、その補足的な手段としてさまざまな方法が開発されている。

例えば、政策のもたらす影響の範囲がある特定の個人またはグループに限定できる場合には実験的方法が適用できるし、その影響の範囲が対象者全体に及んで特定が困難な場合には実験以外の方法が用いられる。前者の実験的方法においては政策効果を把握するために政策の対象者と対象とならない者との対比に基づく比較分析を行うことになる。この場合、政策の対象者の中から実験対象グループと比較対照グループとを無作為に選定できることが理想であるが、それが困難な場合には

作為的に両グループを設定する方法が用いられる。

後者の実験以外の方法においては、政策が実施される前後の時間軸上で、政策対象グループの実施前と実施後の状況を対比する方法が一般的に用いられる。

本格的な政策実験が可能か否かは求められる評価の精度が大きく関わっているため、政策評価を行う場合に用いる手法の決定は、対象政策の特性、要求される分析精度、分析に許される時間などを十分配慮して決定しなければならない。本格的な政策実験もしくはその他の方法を採用した場合でも、結果をより客観的に捉えるために、統計的解析手法が補完的手法として用いられる。

### 3 有効性評価の考え方とデザイン

#### 3.1 有効性評価の考え方

有効性評価では、政策が実施されたことにより、政策対象者の状況が所期に意図した方向へどれだけ変化したかを計測・評価する。したがってこの場合には、あらかじめ具体的に目標が設定され、成果の評価基準も規定しておく必要がある。例えば、水質汚濁の改善、子供の栄養状態の改善、犯罪発生率の低下などの幅広い社会的問題がその目標として設定される。

これらの課題に対する政策の有効性評価を行うには、まず実施された政策によってそれらの状況がどれだけ改善されたか、すなわち政策が意図した目標にどれだけ近づいたか、もしくは目標をどれだけ達成できたか、といった状況変化を具体的に把握するために必要なデータを収集する必要があるが、このようなデータ収集のためのアプローチ方法（以下では「デザイン」という。）を吟味・作成する必要がある。

有効性評価の代表的な手法としては次のようなものがある。<sup>3)</sup>

1) 対照実験法：政策を実施する「実験集団」と、実施しない「対照集団」を設け、当該政策以外の条件を同等にして比較を行うことにより、政策の効果を測定する。

2) 統計解析法：政策の実施にあたり、制御可能な要因と不可能な要因（外部要因）との関係、目標と実績の乖離が生じた要因等を回帰分析や計量経済モデル等を用いて分析する。

3) ベンチマーク法：目標となる水準（ベンチマーク）を設定し、評価対象の対応する指標と比較対照する。

社会政策に関する有効性評価における視点は、社会実験が可能か否かを問わず、ある政策が実施された場合と実施されなかった場合の政策対象者に与える影響の差をいかに統計的手法によって正確に推計するかにある。

#### 3.2 有効性評価のデザイン

##### 3.2.1 デザインの概念

上述のように政策効果を把握するには、さまざまな政策の影響を受けた結果として生じる対象者の状況変化の中から、当該政策の効果だけを抽出するために、政策の実施前と実施後においてその政策の対象となったグループに関する政策効果を把握し易いように実験計画を設定（デザイン）することが極めて重要である。

この「デザイン」とは、評価対象である政策の価値を判定するための情報を収集することが基本的な目的であり、「誰」を対象として、「いつ」の時点で、計測を行うのかを示すものである。言い換えれば、デザインとは、もし政策が実施されなかった場合、あるいは別の政策が代わりに実施された場合にはいかなる結果となるのかを予測する場合に政策の影響度を効果的に計測するために行うための設計である。

##### (1) デザインにおける「誰」とは

デザインにおける「誰」とは、効果を測定したい政策の実験対象グループ（以下では「実験グループ」という。）と、その政策効果を効果的に測定するために用いられる比較対照グループ（以下では「対照グループ」という。）のことである。一般に、対照グループは効果を測定したい政策の対象者とならないグループではあるが、その属性については実験グループと同一または類似のグルー

プを設定する必要がある。また、この対照グループはグループの選定を行う際に、無作為抽出が可能であるか否かによって2種類に分かれる。

- 1) 実験グループ
- 2) 対照グループ（無作為抽出）
- 3) 対照グループ（作為的抽出）

(2) デザインにおける「いつ」とは

デザインにおける「いつ」とは、評価を行う時点のことであり、政策の実施前、実施中、実施後および実施前から実施後にかけて繰り返して行われる場合などがある。

- 1) 事前評価（実施前）
- 2) 中間評価（実施中）
- 3) 事後評価（実施後）
- 4) 時系列評価（実施前から後まで繰り返し）

評価手法として、定量的な方法を用いるか、定性的な方法を用いるかにかかわらず、このデザインこそが政策評価の基礎データとなるサンプルの性格を決定し、データの価値を左右することになる。このデザインが妥当であれば、それを基盤としてさまざまな分析手法を利用することができる。

実際に、政策効果を把握する場合には、政策の特性やデータ収集状況に応じて次のような手法が適用される。

- 1) 実験グループと対照グループとに無作為に分割し、それぞれの効果を比較する。

- 2) 何らかの基準によって作為的に分割した実験グループと対照グループとを比較する。
- 3) 実験グループに対して、政策の実施前と実施後とを比較する。
- 4) 3)の方法を時系列的あるいはクロス・セクショナル的に比較する。
- 5) 政策の効果を既定の基準（ベンチマーク）に照らして評価する。
- 6) 政策の効果についての専門家、政策担当者、政策対象者による判断によって評価する。

### 3. 2. 2 デザインの組合せとその特徴

以下では、デザインの組合せおよびその特徴について検討するが、実験対象者および評価時点としてそれぞれ次の3つの場合を想定する。

〔実験対象者〕

- 1) 実験グループと無作為抽出対照グループ
- 2) 実験グループと作為的抽出対照グループ
- 3) 実験グループのみ

〔評価時点〕

- 1) 事前評価と事後評価
- 2) 事後評価のみ
- 3) 時系列評価

これらの実験対象者と評価時点の組合せによる政策評価デザインは表1のようにまとめることができる。<sup>4)</sup>

表1 政策評価デザインの組合せ

実験対象者	事前・事後評価	事後評価のみ	時系列評価
実験グループと無作為抽出対照グループ	デザイン1	デザイン4	(デザイン5)
実験グループと作為的抽出対照グループ	デザイン2	該当なし	デザイン5
実験グループのみ	デザイン3	該当なし	デザイン6

(1) デザイン1：実験グループと無作為抽出対照グループを用いた事前・事後評価デザイン（True Control Group Pretest-Posttest Design）

これは自然科学で行われる本格的な実験方法を政策評価に適用したものであり、サンプル集団を政策を実施する実験グループと実施しない対照グ

ループに無作為に分割する。対照グループには、まったく政策を実施しないか、あるいは代替的な政策を実施する。両グループのスタート時点における状況を把握するために、事前評価のスコア（評価点）を計測しておく。政策の終了時点に行う事後評価において、実験グループの平均スコア

(事後評価スコア－事前評価スコア)が対照グループの平均スコアよりかなり高い場合には、その差は当該政策の効果によるものと判断することができる。そのような結果が得られるときには、この政策は有効であると判断できる。また、当該政策と代替的な政策を比較する場合にも有効な手段であるといえる。

(2) デザイン2：実験グループと作為的抽出対照グループを用いた事前・事後評価デザイン (Non-Equivalent Control Group Pretest-Posttest Design)

これはグループを作為的に抽出することを除けば、デザイン1と同じである。このデザインは、例えば、学校において生徒個人を対象として実験グループと対照グループに無作為に割り当てることはできないが、クラス毎に割り当てることができるような場合に有効である。すなわち、政策の実験対象となるクラスに対して、政策の対象とならないクラスを対照グループとして設定するが、両クラスの属性は異なっている場合である。

(3) デザイン3：対照グループを用いない事前・事後評価デザイン (Before and After Design)

事前評価を受けた1つのグループに対して、政策を実施した後に事後評価を行う。この結果を別途定められる何らかの評価基準で比較する。ただし、この手法は他の状況が変化しない場合（例えば、事前評価と事後評価の間隔が短い場合など）に有効な手法である。

(4) デザイン4：実験グループと無作為抽出対照グループを用いた事後評価のみのデザイン (True Control Group Posttest Only Design)

これも本格的な実験と呼ばれるが、デザイン1で事前評価を行わない場合である。すなわち、事前評価が不可能であったり、長期間を要したり、政策効果を阻害するなどの理由で、事前評価を行うことが困難な場合に、サンプル集団を実験グループと対照グループに無作為に割り当てることによって、同等のグループを確保して比較を行う方法である。この方法は、主として態度の変化を計測する場合に用いられる。

例えば、サンプル数がかなり大きいときには、いろいろな特性において同等になるような無作為分割を行うことができるので、どんな事後テストをするかの決定を政策が終了する時点まで保留できるし、また、最初に計画していなかったテストを事後的に加えることも可能である。デザイン4はデザイン1と一緒に用いることも容易であり、例えば、基本的には事前・事後評価を行うが、態度の評価に関しては事後評価だけを用いるといった選択もできる。

(5) デザイン5：実験グループと作為的抽出対照グループを用いた時系列評価デザイン (Time Series with a Non-Equivalent Control Group)

これはサンプル集団を作為的に実験グループと対照グループに分割し、両グループについて政策実施の前と後に時系列的に評価する方法である。時系列デザインでは、何らかの外生的・無関連な事象が政策の実施とともに偶然に生じた場合、その効果は両グループに共に現れるので、それを識別することが可能である。対照グループを無作為に抽出した時系列デザインは理論的には存在するが現実には実現が難しいデザインである。

(6) デザイン6：対照グループを用いない時系列評価デザイン (Single Group Time Series Design)

これは実験グループ自身を対照グループとして利用する方法である。政策実施前に一定の間隔毎に同じグループに対して同じ評価尺度による評価を複数回にわたって実施して、時系列的に見た評価尺度の傾向を示すモデルを作成しておき、政策実施後における予測値を推定しておく。一方、同一グループに対して政策を実施した後の成果指標を計測し、その差を政策の影響とする方法である。このデザインは事前・事後評価デザインと同様に、1つのグループを用いて測定する方法であるが、事前・事後評価デザインに比べて他の政策や情勢変化を除外できる点で優れている。

時系列デザインの利点は、政策の実施前後の一連の測定値を得ることができる点であり、これは実施前後の1時点だけのデータしか取れない場合に比べてより正確な評価ができることである。時

系列デザインの問題点は、政策の実施によって明かな変化が観察された場合でさえ、この政策による変化なのか、他の要因によってたまたま起こった変化なのかを見極めることができない点である。

上述のようなデザインを選択する場合の優先順位は、通常次のようになっている。

- 1) 可能であるなら、無作為抽出の対照グループを用いたデザインを選ぶ。
- 2) 無作為抽出対照グループが選べないなら、作為的抽出対照グループを用いたデザインを選ぶ。
- 3) 作為的対照グループも選べないならば、時系列デザインを用いる。
- 4) どれも駄目なら、1つのグループによる事前・事後評価デザインを用いる。

#### 4 社会政策実験のデザインと事例

社会政策実験は自然科学における実験方法を社会政策評価の分野に応用したものであるが、その主眼は、新しい政策の導入を検討する場合に適切な判断情報を提供することにある。新規の政策がいかなる形態を取るべきかをあらかじめ判断することは極めて困難であるが、個別の政策実験を積み重ねることによって、成果の一般化を促進することも可能である。さらに、さまざまな政策形態について検証することで、より効果的で経済効率性の高い政策立案に関する情報を得ることも可能となる。

欧米諸国においては社会政策の効果・影響を把握するために、実際に上記のようなデザインによってグループを設定した政策実験がしばしば行われている。しかし、わが国では主として倫理的・人道的理由等から具体的な社会政策実験を行うことは敬遠される傾向にある。以下では、このような

デザインに基づいた社会政策実験の手法について検討するが、後に示す事例では実験グループおよび対照グループのデザインの仕方を中心にみていくことにする。

##### 4.1 無作為実験法

###### (1) 無作為実験法の考え方

無作為実験法(Randomized True Experiments)は、社会政策評価の領域に実験科学の原理を適用した方法であり、その目的は無関係な要因を取り除いて政策の影響だけを抽出することにある。また比較対象を設定した実験技術を利用することによって特定の政策についての影響度の評価結果を一般化することも重要な視点であり、統計的有意性の検討とともに重視される。これらが確保されるならば、この実験法による評価結果は政策の意思決定や運営・管理にとってより高い信頼性のある情報を提供することができる。

ただし、この無作為実験法を現実の社会政策に適用する場合、理論的には厳密な科学的実験の条件を適用することが基本的な要件であるが、この手法は必ずしもいつでも可能であるとは限らない。現実にはそれが困難な場合には疑似実験法や統計的手法を用いて、無作為実験法に準拠するアプローチを利用することになる。

無作為実験法は、無作為に分類された実験グループと対照グループに対して実際に政策実験を行うことによって、現実には政策を実施したり、取り止めることの影響を把握する方法である。実験にかかる費用は、実施する実験の規模や期間などによって極めて幅があり、例えば、カリフォルニア州で行われた負の所得税政策のように3年から10年に及ぶものもある。通常は無作為実験法による純効果は、次のような簡単な形で表される。

$$\text{純効果} = \left( \begin{array}{c} \text{無作為抽出の実験グループに対する政策実施前後の評点の差} \end{array} \right) - \left( \begin{array}{c} \text{無作為抽出の対照グループに対する政策実施前後の評点の差} \end{array} \right) \pm \left( \begin{array}{c} \text{誤差} \end{array} \right)$$

純効果の値は統計上の誤差を含むので、分散分析、t検定などによって統計的検定を行う必要がある。ただし、政策によっては事前評価が不可能で、事後評価による評価しか意味のない場合もあるので、政策実施後の効果の把握はいずれのグループにとっても重要である。

## (2) 実験対象の選択

実験の対象となるグループ単位（個人、家計、学校や企業などの組織、行政区分など）を適切に選択することは政策実験の基本的な要件である。社会政策実験を行う場合には、まず政策実験の対象となるグループ単位を確定する必要があるが、グループ単位の確定が極めて難しい場合もある。例えば、住宅補助政策は、低所得者住宅の質的向上を目的とすると同時に、地域社会の住宅ストックの増加についても寄与する政策である。この場合、「家計」という単位と「地域社会（市区町村や町丁目など）の住宅ストック」という単位の各々ごとの純効果を把握できるようなグループ単位の設定が必要となる。

さらに、政策の純効果を推定する場合の重要な要素は、実験グループと対照グループの選択・認定である。その際、両グループを比較するには、グループ設定において次の点に留意することが必要である。

- 1) グループの構成：両グループは、構成員の属性が同様であること
- 2) グループの経歴：両グループは、同程度の成熟度であること
- 3) グループの性向：両グループは、政策に対して同様の性向を持っていること

無作為実験法において、グループを選択・認定する場合、どんな分類基準を採用したとしても何らかのバイアスが含まれることは避けることはできない。

## (3) 事例1：ニュージャージーとペンシルベニアにおける所得保障実験<sup>5)</sup>

米国経済機会局(OEO)は、1960年代後期に福祉政策の一環として行われた所得保障政策の効果に関して、貧困家庭およびそれに準じる家庭に対

する所得保障政策とその労働意欲に与える影響を検討することを目的として、大規模な政策実験を行った。この実験は1968年からスタートして3年間実施された。政策の対象は18才から58才までの世帯主で、所得水準が貧困水準の150%以下の家計である。所得保障の手段としては、貧困水準に相当する保障を与えるものから、対象家計の所得水準に応じて段階的に保障を与えるものなど8種類が準備された。例えば、まったく所得のない家計には貧困水準の125%が支給され、また稼得所得に応じて支給率を30%から50%の間で支給額を決める方法が取られた。

この政策実験は、ニュージャージー州の4つの市とペンシルベニア州の1つの市で実施され、実験対象として適格と判断された家計をそれぞれの同意のもとで無作為に2つのグループ、つまり政策を実施する実験グループとその比較のために政策を実施しない対照グループとに分類された。実験グループの家計は毎月の稼得所得額を報告し、適格であれば保障額が支給された。

各家計は政策実施前に詳細なインタビューを受け、実施期間中には3か月毎にインタビューを受けた。インタビュー内容は、雇用、所得、消費、健康、心理的側面の問題などが含まれている。これらの資料は、月別所得の報告と一緒に対照グループと対比しながら、受取支給額が労働意欲を減退させているかどうかを判定するために分析された。

実施上の問題としては、最初に実験対象となった家計の協力を長期にわたって維持することが困難であったことである。いずれの実験でも共通していることは、所得保障は十代の青少年および小さい子供を持つ母親（つまり二次的労働力）の労働意欲を低下させることであった。

## (4) 事例2：住宅費支給に対する需要実験<sup>6)</sup>

貧困家庭に相応の住居を供給する最善の方法については長い間課題とされていた。この課題に関しては、公共住宅を建設して直接提供したり、低所得者住宅の建設業者に補助金を与えたりするこれまでの方法が適切でないことも認識されていた。米国住宅・都市開発省(HUD)は、1970年代後期

に相応の住居取得を促進させることを目的として貧困家庭に直接所得補助を行う政策実験を行った。この政策効果に関しては、低所得者への直接補助は既存の住宅市場を刺激する政策効果を持っていること、および地主や家主に価格値上げの誘因を与える恐れがあることが指摘されていた。そこで、HUDは、次のような3種類の政策実験を実施した。

- 1) 需要実験—直接補助が良質の住居取得を促進するか否かを判定する。
- 2) 供給実験—住宅市場が補助政策に対して価格や家賃の引き上げといった対応をするか否かを判定する。
- 3) 行政面の実験—地域の住宅局が住宅費補助政策をどれほど効率的に管理できるかを判定する。

このうち需要実験では、貧困水準が125%以下の所得層を対象に、フェニックスで1,001家計、ピッツバーグで1,240家計をサンプルとして、17の実験グループと2つの対照グループとに無作為に抽出し、あるグループには一定の水準に見合う住居を取得するために必要な補助金額を支給する、他のグループには家計所得の25%までを補助額として支給する、市場価値での家賃支払いを前提とする家賃補助を行う、などさまざまな要件の違いに合わせた17種類の政策が実施された。給付され

る金額は基本的には、給付額決定時に住宅の質の相違を考慮するか否か、家賃の比率を考慮するか否か、によって決定された。住宅の質を考慮することは、ある家計が住居費として支払い可能と思われる額と平均的な住居費用との差額を給付額として算定するもので、次式によって表せる。

$$P = C^* - bY$$

$P$ : 可能最高給付額

$C^*$ : 一定規模、一定構成の家計にとって支払い可能な推定住居費用

$b$ : 家計が負担すると予想される住居費用の対所得比率

$Y$ : 家計所得

給付水準が変化するのは、 $C^*$ の値をどのような大きさとするかに依存している。ここでは、 $C^*$ を80%、100%、120%という3つの値を、また、 $b$ には15%、25%、35%の3つの値を用いた。この値が大きいほど、家計の受取給付額は大きくなる。上記の方式を用いた結果、住宅の一定の質的条件を満たした11のグループで給付が認められた。その条件とは最大居住者数、適切な設備、材質など住居の物理的な質的条件のことであり、入居時およびその後の定期的な検査を受けることが要件となっている。

表2 需要実験における対策とサンプル規模

A. 住居の相違に応じた対応策 ( $P = C^* - bY$ )

bの値	C*水準	住居条件		
		最低基準	最低家賃 低い=0.7C*	最低家賃 高い=0.9C*
所得の15%	100%	1[81]		
	120%	2[63]	6[58]	9[90]
所得の25%	100%	3[77]	7[89]	10[88]
	80%	4[82]	8[79]	11[78]
所得の35%	100%	5[75]		

B. 家賃比率 (P=aR)

a=60%	a=50%	a=40%	a=30%	a=20%
13[49]	14[190]	15[179]	16[176]	17[11]

C. 給付なし

住居情報あり	住居情報なし
18[296]	19[307]

(注) 各数字は、政策内容の相違による分類番号。ただし、12はまったく制約がなく、どのような住宅でも自由に選択できるグループと定義されている。また、[ ]内の数字は各政策の対象家計数である。

最低家賃による分類は、住居費の最低水準を保障するための間接的方法である。この場合には、家計は一定水準の住居支出を維持することが要件である。実験対象となったグループには、C\*の90%もしくは70%の給付が認められた。

この政策実験では、結果として住居給付費を受給可能な家計のうち3分の1の家計だけが実際に給付を受け、3分の2は給付額と住居基準との条件に合わずに支給を受けられなかった。このことから分かることは、住宅の建築基準として、この政策に参加する意欲を甚だしく低下させるほど高い水準に設定されているわけではないこと、また、単純な家賃補助は運よく良質の住宅にめぐり会った人とそうではない人との間に不公平が生まれていること、などである。したがって、この政策は

住宅を慎重に選ぶ努力を怠った人に有利に働いたことになる。

4. 2 擬似実験法

(1) 擬似実験法の考え方

擬似実験法(Quasi-Experiments)は、無作為実験法が適用できない場合に用いられる。この実験法は、何らかの基準によって抽出されたサンプル・グループのうちから作為的に抽出した実験グループと対照グループとを比較する方法である。したがって、政策の純効果を推計する方法は無作為実験法と似てはいるが、純効果を十分推計できるかどうかは対照グループの特性が実験グループの特性にどれほど近似しているかに依存している。

$$\text{純効果} = \left( \begin{array}{l} \text{作為的に抽出した実験グループに対する政策実施後の評点} \end{array} \right) - \left( \begin{array}{l} \text{作為的に抽出した対照グループに対する政策実施後の評点} \end{array} \right) \pm \left( \begin{array}{l} \text{誤差} \end{array} \right)$$

(2) 事例1：教育切符政策<sup>7)</sup>

教育切符(Educational Voucher)政策は、教育プロセスに自由市場の概念を取り入れようとするものであり、両親は子供の教育費用に相当するクレジットもしくは切符を受け取り、子供の学校を自由に選択して、選択した学校に切符を納付する仕組みである。このような教育資金融資の政策は、

学校間の競争を促し、学校が生徒のニーズにより責任を持つようにさせることで教育の質を改善することが期待できる。比較の対象とされた学校は作為的に選定された次の3つの学校群である。

- 1) 伝統的な勉学志向の学校
- 2) 革新的なカリキュラムを行う学校
- 3) 政策が実施されない学校 (対照学校)

3)の対照学校は、同じ学区から選択され、人種的構成、社会経済的構成、福祉状況などの特性を照合して類似と考えられる学校が選ばれた。

得られたデータについて得点分析や分散分析によるデータ分析が行われた。教育切符政策の結果として、政策実施の最初の年にはこの政策に参加した6つの公立学校で生徒の読解力の低下が見られたが、その後の評価では、伝統的な学校ではなく、革新的なカリキュラムを行う数校においてだけ読解力の低下が見られた。しかし、この研究の目的はこうした観察結果を再検討して政策策定のために有用な情報を提供することである。

### (3) 事例2：エルサルバドルの教育テレビ導入評価<sup>8)</sup>

エルサルバドルでは、教育改革政策の一環として、中学校への教育テレビの導入政策が行われているが、この政策の効果を評価するために、エルサルバドルと米国スタンフォード大学の共同研究のスタッフは、一般学力、読解力のテストを行い、また次のように作為的に設定された3つのグループの生徒に対してインタビュー調査を行った。

グループA：中学に入学した902人の生徒のうち581人の生徒に教育テレビを用いた改革カリキュラムを受講させ、207人の生徒には改革前の従来のカリキュラムを受講させた。

グループB：中学に入学した707人全員に改革カリキュラムを受講させ、そのうち482人には教育テレビを用いたが、残りの225人は用いなかった。

グループC：600人の生徒に改革カリキュラムを受講させ、そのうち467人が教育テレビを用いたが、133人には用いなかった。

各グループの性別、居住地区、家族の社会的地位などの属性の相違は統計的に一定になるように工夫した上で、教育テレビを用いたか否かによって、グループ間およびグループ内での影響の比較を行った。

比較の結果、改革カリキュラムを受講した生徒は、従来のカリキュラムを受講した生徒よりも一般的学習能力でより向上し、また、教育テレビを

用いた生徒は用いなかった生徒よりも向上が見られた。グループAとBについては特定の課題について習得した知識を測定するアチーブメント・テストの結果を使って2年間にわたって比較が行われたが、その結果、教育テレビの導入効果は初年度だけに顕著に見られたが、それ以降はあまり明確な効果は現れなかった。

## 4.3 事前・事後比較分析

事前・事後比較分析(Before and After Studies)は、同じ政策対象グループを2時点間で比較するもので、政策実施前の時点での評点と実施後の時点での評点の差を政策の純効果の推定値として捉える方法である。

政策実施の事前と事後において、その政策対象グループの評点の変化を比較する場合、次の点を考慮する必要がある。

- 1) 事前と事後の間が長期にわたる場合には、政策の効果を見誤る恐れがある。また、政策の実施時期とその効果が現れる時期とにズレがある場合には、この方法による比較分析は適切といえない場合がある。
- 2) 事前と事後の間の長期的傾向に変化のある場合には比較分析は困難である。例えば、雇用訓練政策は、経済状況が悪化しているときや長期に及ぶ失業の増加時期にはあまり効果的とは言えない結果になることが考えられるし、犯罪率の減少が警察の取締り強化時期と偶然一致した場合には効果的に見えることもあり得る。
- 3) 通常あまり起こらないが、事前と事後の間に起こり得る一回限りの出来事が結果に影響を与えることがあり得る。例えば、自然災害、政治的危機、伝染病の蔓延などが考えられる。

## 4.4 不連続回帰分析

### (1) 不連続回帰分析の考え方

不連続回帰分析は、ある政策の対象者を選択するにあたって明確なルールが存在する場合に有効である。例えば、ある大学の奨学金政策において学生が標準テストで得た点数に基づいて対象者が

決められる場合とか、食糧切符政策で家計所得の額によって補助額が決定される場合である。もしこのような選択ルールが十分に遵守されるならば、ルールに基づいて選択された対象者に注目して、当該政策の純効果を統計的手法によってかなり正確に推定することが可能である。この分析には、選択ルールが画一的に適用され、影響度に関する有効な測定手法が存在することが必要条件である。

この手法の一般的アプローチ法は「選択バイアス・モデル」として知られている。政策対象者の選択プロセスが既知である場合に用いられるものと同じ推論に依拠しており、無作為実験法ではできない政策評価が期待できる。

#### (2) 事例：住宅と人口に関するゾーニング規制の効果推定<sup>9)</sup>

シカゴ市内およびその郊外地区に関してシカゴ標準人口調査区域 (SMSA) から抽出したセンサス・サンプルを用いて、ゾーニング規制の効果について実験が行われた。ゾーニング規制とは、広

い面積に1家族だけしか認めないという極めて厳しい利用形態から、商業地域と工場地域を含めていかなる土地利用も認めるという緩い利用制限形態にまで広範にわたる制限政策である。

住宅と人口についての1960年と1970年センサス値を用いて、ゾーニング規制指標をベースにしてシカゴの郊外地区における1970年の特徴を予測する回帰分析を行った。ゾーニング規制は、規制のない場合に考えていたよりも成長率を変化させ、その地域の社会的経済的およびライフ・サイクル上の状況変化に影響を与えた。また、回帰分析では、シカゴ郊外の住宅密集度に対してゾーニング規制が行われた場合の効果についても検討された。回帰分析の結果は、1970年の実質住宅密度と1980年ベースで予想されたものとの差を示している。そのほかの変数のすべては、1960年の土地利用形態を制約したゾーニング規制の効果を計測する尺度となる。回帰分析の結果は表3のとおりである。

表3 ゾーニング規制の回帰分析結果

独立変数	回帰係数	標準誤差
1960年住宅密度 (1マイル平米あたり)	.95	.03
排除的ゾーニング指数	-253.01	50.61
商業利用に規制された面積比率	-52.18	3.93
工業利用に規制された面積比率	-4.18	2.93
規制指標	2325.7	413.58
決定係数=0.94, サンプル数=144		

#### 4.5 シミュレーション・モデルによる評価

近年、シミュレーション手法は社会科学の分野で広く利用されるようになったが、政策評価においても応用される。社会政策評価においては、関係する要素数が多いために、いかに実際のデータを用いたとしても現実の世界を完全に再現することはできないが、シミュレーション手法を用いることによってすべての構成要素を制御して、それらの要素を体系的に操作することができるので、

ある特定の問題や仮定がいかに成果に影響を与えるかを直接的に理解することが可能である。

ここでは、前述した手法のうちから、無作為実験法、不連続回帰分析、事前事後比較の3つの社会政策実験のデザインについていかにシミュレーションするかを検討する。<sup>10)</sup>

##### (1) 無作為実験法モデル

事前評価の計測値  $X$  は、真のスコア ( $t$ ) と誤差項 ( $e$ ) からなる加算型関数であると仮定する。

$$X = t + e_x$$

実験グループおよび対照グループに対して、無作為に  $t$  と  $e$  を設定し、事前評価値とする。次に、両グループを示すダミー変数  $Z$  を以下のように設定する。なお、 $r$  は平均が 0、標準偏差がある  $\sigma$  値に等しく分布する任意の変数で、他のすべての変数から独立である。

$$r \leq 0 \text{ の場合には, } Z = 1 \text{ (実験グループ)}$$

$$\text{それ以外の場合には, } Z = 0 \text{ (対照グループ)}$$

政策の事後計測値  $Y$  は次のように  $t$ ,  $e$ ,  $gZ$ , の加算型として設定するが、対照グループでは  $Z = 0$  であるから  $gZ$  も 0 となるので、 $gZ$  は実験グループだけに加算される。なお、 $g$  は政策の影響の大きさを示す係数である。

$$Y = t + e_r + gZ$$

### (2) 不連続回帰分析モデル

不連続回帰分析においては、事前計測値として一定値を選択すること、および評価の性質上、実験グループに対照グループより高いスコアを割り振るのが特徴である。

$Z$  に関しては、次のように設定する。

$$X \leq \text{一定値の場合, } Z = 1$$

$$\text{それ以外は, } Z = 0$$

事前計測値  $X$  および事後計測値  $Y$  は、無作為実験法モデルと同じである。

$$X = t + e_x$$

$$Y = t + e_r + gZ$$

### (3) 事前事後比較モデル

このモデルでは、片方のグループに作為的に一定の条件を与えて、結果として 2 つのグループは政策の事前計測値に一定の差を生じさせることが特徴である。このシミュレーションにおいては、一方のグループの事前・事後の計測値に一定の値を加算することで不均等を作り出す方法をとる。まず、不連続回帰分析モデルと同様に、 $Z$  を設定する。

$$X \leq \text{一定値の場合, } Z = 1$$

$$\text{それ以外は, } Z = 0$$

$X$  については、次のように設定する。

$$X = t + e_x + dZ$$

$d$  は実験グループに加算される定数（実験グループが対照グループより勝るか劣るかのいずれを望むかによって、正もしくは負のいずれともなる）で、ダミー変数  $Z$  (0, 1) に乗じる。事後計測値  $Y$  は次のようになる。

$$Y = t + e_r + dZ + gZ$$

### (4) シミュレーションの実施

これら 3 つのモデルは、まずパラメータを規定する 6 つの制約値 (K1~K6) の特定化が行われる。これらの制約値を変更することで、政策効果の大きさ、不均等の度合い、計測値の信頼度、サンプルの規模などを変更することができる。

さらに、3 つのモデルに関する変数  $X$ ,  $Y$ ,  $Z$  が定義され、不連続回帰モデルでは事前計測における判定値としてゼロ値が使われる。

3 つのモデルはともに次のような ANCOVA 回帰モデルを使って分析され、各モデルについてグループの平均値と標準偏差を計測する。

$$Y_i = b_0 + b_1 X_i + b_2 Z_i + e_i$$

ただし、

$Y_i$  : ケース  $i$  の事後評価のスコア

$b_0$  : 定数あるいは中止パラメータ

$b_1$  : パラメータに関する傾き

$X_i$  : ケースの事前評価のスコア

$b_2$  : 政策効果のパラメータ

$Z_i$  : ケースのグループ・ダミー変数

$e_i$  : ケースの誤差

### (5) シミュレーションの実行

シミュレーション・モデルの有効性を証明するために、次の 2 つのシミュレーションが実施された。

ケース 1 : 高信頼度の場合

すべてのモデルに関する政策効果を 5 ポイントとし、グループ間の不均等を修正するために事前事後比較モデルの実験グループには 3 ポイントを加えられた。不連続回帰モデルの判定値としてゼロを用い、計測値の信頼度は 0.9 とした。

ケース 2 : 低信頼度の場合

すべてのシミュレーション・パラメータはケー

ス1と同じとし、計測値の信頼度を0.5とした。計測値の信頼度は、正しい値と誤った値との標準偏差の相対的大きさを変えることで設定した。

真の値=5、1回を100ケースとし、50回のシミュレーションを行った結果を表4に示す。なお、係数の有意性は  $b_2 \pm 2 * SE(b_2)$  の範囲を外れた値によって決定される。一般に、推定値がバイアスを持たないと見なすのは、平均値の真の値からの誤差がプラス・マイナス標準誤差の2倍以上にならない場合である。つまり、0.05有意水準においては、計測値が  $b_2 \pm 2 * SE(b_2)$  の間に収まるということである。例えば、無作為実験法モデルの場合、低信頼度のシミュレーションでは、平均ポイントは4.890、標準誤差は0.081であり、したがっ

て真のポイント5は、 $4.89 \pm 2 * 0.081$  の間に収まっているので、これらの条件に関してバイアスはないと考えられる。

シミュレーションの結果から次のような方法論上の重要な原理を読み取ることができる。すなわち、無作為実験法と不連続回帰分析はともにバイアスのない推定結果になっている。また、事前事後比較モデルでは、高信頼度・低信頼度のどちらもバイアスのある推定結果であることを示している。さらに、無作為実験モデルと事前事後比較モデルでは高信頼度・低信頼度のどちらのシミュレーションにおいても平均ポイントに対して同程度の標準偏差であるが、不連続回帰分析モデルの標準誤差はかなり大きいことを意味している。

表4 シミュレーション結果

		$b_2$	$SE(b_2)$	$Min(b_2)$	$Max(b_2)$
高信頼度(0.9)	無作為実験法	4.973	0.041	4.251	5.578
	不連続回帰分析	5.252	0.044	4.448	5.884
	事前事後比較	5.032	0.066	3.702	5.951
低信頼度(0.5)	無作為実験法	4.890	0.081	3.663	6.401
	不連続回帰分析	6.344	0.094	4.521	7.930
	事前事後比較	5.030	0.180	2.280	7.870

#### 4.6 計量経済モデルによる影響度評価<sup>11)</sup>

1970年代のオイル・ショック以来、多くの電力会社ではピーク時の電力需要を削減することで、高価な設備増設を避けることを目的として需要管理政策を実施している。この政策を実際に顧客に対して実施する前に、政策の実施上の問題を取り除き、政策の影響を計測するために、南カリフォルニア・エジソン会社は1984年から86年にかけて一般住宅用の電力負荷管理政策についての長期の政策実験を行った。

この実験は、会社からピーク期間であることを通知された場合、政策対象者である顧客は同意した水準まで電力需要を減少させなければならない、というものである。もし顧客が契約水準まで使用

量を削減するのに失敗した場合には電力供給はストップされるが、その場合は顧客は需要水準を下げることを条件としてリセット・ボタンで電力供給サービスを回復できる。

この政策の影響度を計測するために、顧客を実験グループと対照グループに無作為に分類された。電力需要データは実験グループの住宅に設置されたデコーダを通して15分間隔で収集された。さらに、両グループから人口学および行動学的なデータが収集された。分析の焦点は、政策が実施される夏期期間に電力需要がどの程度削減できるかにおかれた。削減された電力需要を計測するために、次のモデルが使われた。

[実験グループ]

$$KW = b_0 b_1 T + b_2 D + b_3 DT + b_4 Y + b_5 DY$$

[対照グループ]

$$KW = b_0 + b_1 T + b_4 Y$$

[電力需要の減少]

$$KW = -b_2 D - b_3 DT - b_5 DY$$

ただし、

$T$  : 実施日の温度

$D$  : 実験グループは1, 対象グループは0のダミー変数

$DT$  : 温度と各グループ間の相関

$Y$  : 顧客調査から得られる人口学的な住宅についての変数

$DY$  : グループ別の人口学的な住宅についての変数の相関

#### 4.7 判定評価

社会政策実験においてその効果を推定する場合、実験グループと対照グループをデザインし比較すること、あるいは長期的な影響の変化を捉えることは容易なことではない。そこで、ある程度の妥当性をもって政策効果を推定できる専門家グループ（例えば、政策評価者、政策実施担当者、政策参加者など）を何らかの観点から選定し、その人達の判定によって評価する方法として判定評価（Judgment Assessments）がある。この種の専門家達には、一般的に、政策評価にあたって利己的動機に基づく判断をすることなく、より客観的視点からの判断を下すことが期待される。

例えば、「2か月の大型トラック運転手養成政策は、90%の成功率を納めている」と言われる。これから推論できることは、「この種の政策が技術習得の目標を達成するのに有効なものである」ということである。一般の人からみると「トラック運転手になりたいと思っている未熟練者が2か月の養成政策によってほとんどの人がその目的を達成できる」とはとても考えられないことであるが、専門家は運転技術や試験についての知識に基づいて判断した結果として、上記のような判断を下すことができるのである。そのような高度な専

門的な知識がその基礎となっているという点に専門家による判断には価値がある。もちろん、この種の判断が曖昧性を持っていることも事実である。上記の例で、目的達成者が30%程度しかいない場合には、政策によるものか、別な要因によるものかは明確には判断することは困難である。

専門家による判断を利用する場合には、詳細な判断基準が明示されることが重要である。すなわち、専門家の判断が依拠する基盤を可能な限り明確にしておくことである。もし、専門家自身の経験がその判断の基礎となっている場合には、その経験の範囲や状況ごとの違いを明確にすることが必要である。

専門家としてその役割を期待される人は、以下のようなデータを考察しておくことが必要である。

##### 1) 管理記録

専門家は、次のような項目に関して管理記録から情報収集をしなければならない。

- ・ 政策の規模
- ・ 選択された参加者のタイプ
- ・ 参加者の経験度合
- ・ 参加者の参加後の経緯
- ・ 参加者当たりの政策費用
- ・ 政策目標との関連での参加前後の状況変化

##### 2) 政策運営の観察

家庭訪問、クラスルーム会合など参加者の活動を求める政策では、専門家がその活動を直接観察する必要がある。

##### 3) 参加者へのインタビュー

実験グループと対照グループの双方あるいは一方に対して非公式にインタビューする場合には、少なくとも自発的に申し出た者を対象とし、以下の問題を取り上げる。

- ・ 参加者の応募状況
- ・ 参加者の動機
- ・ 参加者の政策に関する満足度
- ・ 政策目標達成にむけての参加者の進歩

##### 4) 関係者とのインタビュー

地域の役人、競合的な政策の管理者、学校管理者や警察責任者など地域制度における重要な役職

者、地域の制度に大きな影響力を持つ個人あるいは組織（大地主、銀行家、政治家など）にインタビューして、次の問題を聞き出す必要がある。

- ・ 政策の価値
- ・ 政策が社会にもたらす利害の程度
- ・ 実験期間後の政策継続の利害の程度

## 5 ま と め

政策評価手法の概要、および社会政策実験の各手法の考え方・特徴と事例を示してきたが、手法はここに示したものがすべてではなく、政策の内容や実験環境の違いに応じてさまざまな方法が研究されている。

このようは社会政策実験には総じて次のような考慮すべき問題点がある

- 1) 自然科学の実験では同一条件における実験結果の再現性が重視されるが、社会政策の場合には同一条件の状況を作り出すこと自体不可能であり、したがって過去に行った実験を完全に再現することもできない。それゆえ、政策実験を用いた政策評価の結果は個別具体的になりがちで、一般化が非常に困難である。
- 2) 政策実験に参加する人は、自分が実験の一部になっていることや観察されていることを認知しているために、実験に参加していない時とは異なった行動をとる可能性がある。また、実験の規模の大小によって、参加者の行動は異なるものになる場合が多い。通常の人間の行動は家族・友人など身近な知人の行動を見習ったり、影響を受けて次の行動を決める場合が多いが、例えば、全国規模の実験では全体の参加者像が見えないために普段よりも慎重な行動をとるようになる。
- 3) 社会政策実験では、実験グループとされた一部の人に公共サービスを提供し、対照グループになった人にはサービスを提供しないことによって政策の影響の大きさを捉えようとしている。この場合、対象となった政策が生命に関わるような公共サービスである場合には、倫理的また

は政治的問題となる可能性がある。

- 4) 社会政策実験には、多額の費用と長期の時間を伴う場合が多い。また、補助金や給付金など現金が支給される場合が多く、その金額も恣意的に決められるので、実験グループと対照グループ間で不公平が問題になる場合がある。
- 5) この実験法は政策過程の初期段階では成果が十分に生じない恐れがあるためにあまり成果を期待できない。

以上のような問題点を抱えてはいるが、社会政策実験は社会政策の有効性を改善し、政策効果の「ある」、「なし」をかなり正確に結論付けることができるという点で有用な手法であるといえる。特に、過去に行ったことがないような新規の政策を実施する場合に、本格的に実施する前にその政策の影響や効果について知見する上で有効な手法であると考えられる。

## 注

- 1) 藤本(1998A), p39, 藤本(1998B), p106など
- 2) 参考文献(25)より
- 3) 参考文献(25)資料13
- 4) 参考文献(10)
- 5) 参考文献(14)
- 6) 参考文献(18)
- 7) 参考文献(21)
- 8) 参考文献(14)
- 9) 参考文献(15)
- 10) 参考文献(19)
- 11) 参考文献(12)

## 参 考 文 献

- (1) Basilevsky, A.D. Hum, 1984, *Experimental Social Programs and Analytic Methods: An Evaluation of the U.S. Income Maintenance Projects*, New York: Academic.
- (2) Bentkover, J.D., V.T. Covello & J. Mumpower, 1986, *Benefits Assessment*, D.Reidel Publishing

- Company.
- (3) Bennett, C.A. & A.A.Lumsdaine, 1975, *Evaluation and Experiment*, New York: Academic.
- (4) Berk, R.A. & D.Rauma, 1978, Capitalizing on Non-Random Assignment to Treatment: A Regression Continuity Analysis of a Crime Control Program, *Journal of the American Statistical Association*, March, 21-28.
- (5) Burnstein, L., H.Freeman & P.H.Rossi, 1985, *Collecting Evaluation Data: Problems and Solution*, Beverly Hills, CA: Sage Publications.
- (6) Campbell, D.T., 1969, Reform as Experiment, *American Psychologist*, **24**, April, 409-429.
- (7) Conner, R.F., 1977, Selecting a Control Group: An Analysis of the Randomization Process in Twelve Social Reform Programs, *Evaluation Quarterly*.
- (8) Hirst, E., 1990, Progress And Potential In Evaluating Energy Efficiency Programs, *Evaluation Review*, **14**, No.2, April, 192-205.
- (9) Moffitt, R., 1991, Program Evaluation with Non-experimental Data, *Evaluation Review*, **15**, No.3, June.
- (10) Morris, L.L., & C.T.Fitz-Gibbon, 1987, *How to Design a Program Evaluation*, SAGE.
- (11) Patton, M.Q., 1987, *How to Use Qualitative Methods in Evaluation*, SAGE.
- (12) Ridge, R.S., G.A.Stern & R.E.Watt, 1990, Econometric Model Evaluation: Implications for Program Evaluation, *Evaluation Review*, **14**, No.13, June.
- (13) Rossi, P.H.(ed.), 1982, *Standard for Evaluation Practice*, New Direction for Program Evaluation.
- (14) Rossi, P.H. & H.E.Freeman, 1985, *Evaluation : Systematic Approach*, SAGE.
- (15) Shlay, A., & P.H.Rossi, 1981, Keeping Up the Neighborhood : Estimating Net Effect of Zoning, *American Sociological Review*, Dec.
- (16) Smith, V.K., 1993, Non-market Valuation of Environmental Resources : An Interpretive Appraisal, *Land Economics*, **69**, No.1, February, 1-26.
- (17) Stecher, B.M. & W.A.Davis, 1987, *How to focus on Evaluation*, SAGE.
- (18) Struych, R.J. & M.Bendick(eds.), 1981, *Housing Vouchers for the Poor : Lesson from a National Experiment*, The Urban Institute Press.
- (19) Trochim, W.M.K. & J.E.Davis, 1986, Computer Simulation for Program Evaluation, *Evaluation Review*, **1**, No.5, October.
- (20) Wilner, D.M., R.P.Walkely, T.C.Pinkerton & M.Tayback, 1962, *The Housing Environment and Family Life*, Johns Hopkins Univ. Press.
- (21) Wottman, P.M., C.S.Reichardt & G.St.Pierre, 1978, The First Year of the Education Voucher Demonstration: A Secondary Analysis of Student Achievement test Scores, *Evaluation Quarterly*, **2**, May.
- (22) 藤本幸生, 1998A, 経済情報学の創生とその目指すべき方向, 社会情報学研究 (日本社会情報学会), No.2.
- (23) 藤本幸生, 1998B, 社会情報研究の視点と複雑系研究, 社会情報学研究 (呉大学社会情報学部), No.4.
- (24) 宮川公男, 1994, 政策科学の基礎, 東洋経済新報社.
- (25) 中央省庁等改革推進本部事務局・総務庁行政監察局, 2000, 政策評価に関する標準的ガイドラインの案
- (26) 通商産業省大臣官房政策評価広報課, 1998, 政策評価の現状と課題—政策評価研究会中間報告—