

フィールド研究における準実験

—— 単一事例実験の関連方法 ——

田 中 潜次郎

目 次

序 論

一 準実験の概要

二 非等価対照群型

三 中断時系列型

結 語

序 論

準実験 (quasi-experimentation) は、はじめ政府や自治体の行う政策が問題解決に有効に作用したか否かを評価する観点として考案され^(一)、のちには広く実験的条件統制を行うのが難しいフィールド領域において、因果推論 (causal inferences) を行うために体系化された方法である^(二)。「準実験」という名称は、外見的には正規の実験法とは異なる形式をとりながら、実質的には実験法に準じて原因と結果の関係を明らかにする機能を備えていること^(三)を示す。日本では最近、公共政策の評価法として紹介されている^(四)。

(一) Campbell, D. T. Reforms as experiments. *American Psychologist*, 24, 409-429, 1969.

(二) Cook, T. D. & Campbell, D. T. Quasi-experimentation: Design & analysis issues for field settings. Houghton and Mifflin, 1979.

(三) 準実験を示す用語は、'quasi-experimentation' の他に、以下の文献の中で、'quasi-experiments' & 'quasi-experimental designs' が使われている。① Cook, T. D. & Campbell, D. T. The design and conduct of quasi-experiments and true experiments in field settings. In Dunnette, M. D. (Ed.) *Handbook of industrial and organizational psychology*, Chapter 7 (pp. 223-326). Rand McNally, 1976. ② West, S. G. Beyond the laboratory experiment: Experimental and quasi-experimental designs for interventions in naturalistic settings. In Karoly, P. (Ed.) *Measurement strategies in health psychology*, Chapter 7 (pp. 183-233). Wiley-Interscience, 1985.]°

(四) 薬師寺泰蔵「公共政策」(八〇〜九七頁)。東大出版会・一九八九年。

一 準実験の概要

(一) 基本特徴

(1) 問題

準実験は、正規の実験法にある特徴の中で、次の二点を欠く一群の経験的研究から成ると規定されている。^(二) 第一に、正規の実験は実験室で行われるのが普通であるが、準実験が実験室で行われることは極めて少ない。第二に、正規の実験で変数の効果を比較対照するために必須とされる無作為配置の手順が、準実験には欠けている。

以上の二点以外のすべての点で、準実験は正規の実験に似ている。準実験は実験と同様に、独立変数と従属変数の間の因果関係を実証することを目的とする。ただし、準実験における独立変数は単一であるとは限らないし、従属変数の測度も複数であることが少なくない。^(三)

準実験は、フィールドないしは自然の条件下で行われることが多い。その結果、次の二つの問題が生じる。第一は、変数の効果を評価するために必要な比較対照の基準すなわちコントロールが欠けているから、これに準じる手順をいかに整えるか、ということである。第二は、変数の条件統制が不完全であるから、本来の実験法とは異なる解釈基準をいかに設定するか、である。

(2) 由来

準実験の成立経緯は、大略次の通りである。基本的な発想は、すでに一九五〇年代にあらわれている。^(三) しかし、

「準実験」という名称が使われるようになったのは、六〇年代に入ってからである。^(三) 最初は、対照群に「準」じる群 (quasi-controls) をいかに設定するかが最大の問題であったから、初期の準実験は、現在の非等価対照群型〔4〕を参照〕を意味していた。長期間の時系列データを対象にした現在の中断時系列型〔4〕を参照〕にも共通の特徴を認めて、これを非等価対照群型と並ぶ準実験として位置づけるようになったのは比較的遅く、七〇年代に入ってからではないかと思われる。

(3) 適用対象

① 観点 方法の適用を例示すると、次のようになる。タバコが有害であるという認識を国民の中に浸透させるために、「健康のため吸いすぎに注意しましょう」と表示することが全国一斉に義務づけられた。この有害性表示とタバコ消費量の間に因果関係があるか否かという問題は、準実験的な考察の対象とすることができる。もし、その後タバコの消費量が減少したとする。減少の原因が他の要因よりも有害性表示にあることを示すためには、比較対照の基準として、表示をしなかった場合のタバコ消費量のコントロール・データが必要である。しかしながら、表示は全国一斉に始められたので、コントロール群はない。問題は、このような場合にコントロールに準じるデータをいかに整えるか、ということである。

② 実例 一九五五年のアメリカ・コネチカット州で、交通事故による死亡件数が記録的に増加して三二四件になった。当時の知事が速度違反に対して厳罰処分を導入すると、その翌年には二八四件に減少した。^(四) 問題は、死亡件数減少の原因が新政策の実施にあるといえるかどうかである。

一九七〇年代のアメリカ・カリフォルニア州で、降水量が激減して平年同時期の三分の一になった。地域の自治体は市民に対して、水不足の実態についての情報提供と水節約の要請を行った。自治体によっては、水の超過使用

者から罰金をとる制度を導入した。この間に水使用量は平年同時期の六〇%まで減少した。問題は、使用量減少の原因が罰金制度の導入にあるといえるかである。^(五)

(4) 方法の種類

以上の実例に対しては、次のような分析が可能である。

① 非等価対照群型 交通事故の防止策を実施した州や、水の超過使用に罰金を課した自治体を実験群とみなし、特に対策を講じなかった他の州や自治体を対照群のコントロール・データとして使うことにより、政策の効果が評価される。ただし、この場合に対照群となる州や自治体は、実験群の州や自治体と等質な集団ではないから、本来の実験における対照群 (control group) ではなく、対照群に準じる集団 (quasi-control group) にすぎない。

この型は、群の構成が等質ではないから、「非等価対照群型」の準実験と呼ばれる(二を参照)。

② 中断時系列型 利用できるデータが政策を実施した州や自治体のものであって、対照群のデータが欠けている場合に、この方法が使われる。政策の効果を評価するためには、比較対照のコントロール・データが必要なることは当然であるから、その州や自治体における政策実施前の比較的長い期間の状態がコントロール・データとなる。非等価対照群型のデータは、原則として政策実施の事前と事後の単発型でよいのに対して、この型では、連続測定による時系列データが必要である。この型は、一連の時系列を政策導入の前と後に分断する形式で分析が行われるから、「中断時系列型」の準実験と呼ばれる(三を参照)。

最初に述べたタバコの有害性表示と消費量の関係については、中断時系列型の分析は可能であるが、有害性の表示が全国一斉に始められたので、非等価対照群型は適用できない。

(二) データ解釈の視点

準実験は、最初から計画的にこれを実施する場合よりも、正規の実験として計画された後で、条件が不備であったことが指摘され、計画実施後に別の視点を追加する必要が生じた場合や、実験として計画されることなく生じた事象に対して実験的な分析を行う場合の方が多いと考えられる。この場合の結果を解釈する際には、正規の実験よりも多くの配慮が必要となる。

次に示すのは、準実験の結果に対して、主要な変数以外の妨害要因 (threats to internal validity) として影響を与える可能性が高いものである。妨害要因に対処するための視点は、準実験研究の中で独自に考案されたのではなく、従来の実験的研究の中で蓄積されてきた考え方が、準実験という新しい形式において整理されたと考えべきであろう。なぜなら、正規の実験でも理想的な形でこれが実施されることはむしろ少ない。実験の結果には、多くの妨害要因が影響を及ぼしている可能性がある。したがって、これに対処するための方法や視点が以前から必要とされてきたのである。準実験は、今まで非系統的に使われてきた次善の手段に対して、合理的な根拠を与えることを目的とする体系的な試みであるといえよう。

(1) 対象そのものに関わる要因

① 歴史 (history) —— 外部要因 独立変数の事象とは無関係であるが、これと同時に生起して従属変数に影響を及ぼす事象を指す。^(六) 実験室の実験では、目的とする変数以外の要因、たとえば視覚刺激と聴覚刺激は暗室と静室を使うことによって統制される。このように被験者を外部刺激から隔離 (insulate) すると共に、外部からの影響を受けない従属変数を選ぶことによって、妨害要因の影響が排除される。^(七)

これに対してフィールドの準実験では、変数以外の要因を統制することは事実上不可能であるから、結果の解釈にあたっては、変数以外の何が結果に影響を及ぼしたかを点検することが必要である。

② 成熟 (maturation) —— 内部要因 歴史が外部的な妨害要因であるのに対して、対象そのものの内部的な変化・成長・衰弱が結果に影響を及ぼす場合がある。この内部的な妨害要因は、成熟と呼ばれる。とくに、人間を対象にして長期間にわたる調査を続ける場合には、この要因の影響に注意しなければならない。

(2) 従属変数の測定値に関わる要因

③ 回帰 (regression) 社会政策の領域では、交通事故多発に対して取締りを強化する場合に見られるように、問題が忍耐の限度をこえた時に、対策が実施される。その後で問題が解決したとしても、対策の実施がその原因なのか、それとも極端な値が単に本来の水準に戻った (regression-toward-the-mean) にすぎないのかは不明である。これを明らかにするには、対策実施前の本来の水準に関する時系列データをとる必要がある。

④ 変動 (instability) 問題に対する介入によって事態がある程度改善されたとしても、その改善の幅が事前の測定値の変動の範囲内に留まるのであれば、改善の原因が介入にあるとはいえない。介入効果は、データ本来の変動との関連において相対的に評価しなければならぬ。

(3) サンプリングに関わる要因

⑤ 選抜 (selection) 比較対照される二群の間に差があったとして、その差が生じた原因が介入そのものにあるのではなく、介入に先立って設定された群構成の非等質性にある場合である。群構成については、次のような問題もある。一年間の平均欠席数がわずか二日の児童を集めて構成した健康児の群は、その一年後には児童が属する母集団本来の平均水準に回帰して、欠席数が増加する可能性がある。この例のように、極端な値を集めて群を構

成する場合には、この問題が生じる可能性が大きい。

⑥ 脱落 (mortality) 介入以前にはほぼ等質な構成であった二つの群が、その後で各群に特有の形で被験者が脱落していく結果、介入以後の両群が異質な集団となる場合である。^(七) 選抜が事前の妨害要因であるのに対して、脱落は事後に生じる妨害要因である。

以上は非等価対照群型の例であるが、単一の群に対して継続的な調査を行う中断時系列型においても、調査途中の脱落者・欠席者・無回答者は、母集団からの無作為標本ではなく、むしろ特定の傾向をもつ被験者である恐れがある。この場合には、時系列の前半と後半は異質な集団であるから、両者を比較検討することには問題がある。^(一〇)

(4) 実施手順に関わる要因

⑦ 検査効果 (testing) 検査の実施自体が被験者の行動に影響を及ぼす場合である。たとえば、タバコの本数を問われた喫煙者はその後喫煙を抑制し、体重を測定された肥満者は食事を控え、健康チェックリストに記入した慢性疾患の患者はそれ以後、リストに記された症状に注意するようになる可能性がある。^(九) 研究目的が検査そのものの効果に関するものではなく、検査が介入効果の測定手段にすぎない場合に、この問題が生じる。

⑧ 記録条件 (instrumentation) 長期間の時系列データでは、その間に記録法が変化することが少なくない。この場合の事前と事後のデータを比較することには問題がある。たとえば、警察機構が整備された結果、整備後の窃盗件数が整備前の数倍に激増したという例がある。^(七) これは、機構整備が窃盗を増加させたのではなく、窃盗の摘発と記録を増加させたことを意味するのである。

(5) 選抜の要因と他の要因の相互作用^(七)

⑨ 選抜×成熟 (selection-maturation)

成長率が群によって異なる場合である。たとえば、中産階級と下

層階級の児童に対して、時間をおいて二回の学力試験を実施したとする。恵まれた環境下にある中産階級の児童は、学力の平均値が上位になるだけではなく、一定期間中に急勾配の向上を示す傾向がある。これは、その間に行われた学校教育の内容とは関係がない。

⑩ 選抜×歴史 (selection-history) 複数の群がそれぞれ異なる地域から抽出された場合に、各群はその地域特有の事象に接する可能性がある。これが妨害要因となって結果に影響を与える場合である。これは、別の名で地方史 (local history) と呼ばれる。

⑪ 選抜×記録条件 (selection-instrumentation) 二つの群が平均値と尺度上の得点分布において著しく異なる場合にありうる誤差要因である。

(三) 方法的特徴

(1) 意義——コントロールの重視

ある事象と他の事象の間に因果関係があることを実証するためには、一方があれば他方もあることを示すだけでは十分ではなく、一方がなければ他方もないことを示すためのコントロール手順が必要である。臨床領域の単一時例実験における基線と撤去の手順^(こて)や、社会領域の準実験においては、準対照群ないしは非等価対照群と、介入前の長い時系列データが、それぞれの領域で不十分であったコントロール手順に準じる機能をもつ。

準実験には、次の二つの意義がある。第一は、実験法を適用できなかった領域に、方法そのものではないにしても方法の論理を適用するという点で、実験法の適用領域を拡大しようとしたことである。第二は、その反面において、それぞれの領域に特有の事情に応じて方法そのものを修正することによって、従来は画一的であった実験法を

個別化・多様化しようとするものである。

(2) 実験法としての位置

正規の実験法は、真正実験 (true experiments)⁽¹¹⁾、または無作為配置型実験 (randomized experiments)⁽¹²⁾と呼ばれる。準実験は、無作為配置 (randomization) の手順を欠くという点で、真正実験とは区別される。

一方、目的とする変数以外の要因の影響を排除するためのコントロール手順を欠く方法は、実験としての条件をみたしていないから、前実験 (pre-experiments)⁽¹³⁾と呼ばれる。準実験は、無作為配置の手順を欠くけれども、比較対照のためのコントロール・データをもつ点で、前実験とは区別される。

したがって準実験は、前実験と真正実験の中間に位置する方法であるが、形式的には前実験に近いとしても、内容的には真正実験に近い性質をもつ。準実験の方法的意義を考えるならば、これを前実験と混同することは避けなければならない。⁽¹⁴⁾なお、単一事例実験を伝統的な事例研究から区別する際にも、伝統的な方法は無対照事例研究 (uncontrolled case studies) と呼ばれて、前実験の形式に分類されている。⁽¹⁵⁾

(3) 開放系の方法

科学方法論を閉鎖系 (closed system) と開放系 (open system) という観点から分類する場合には、準実験は開放系の方法として、次のように位置づけられている。⁽¹⁶⁾

① 閉鎖系 社会科学研究の一部は、物理学や化学の閉鎖系モデルに基礎をおいている。社会心理学や人格心理学の実験的研究でも、条件統制によって、因果関係の実証に不適切な要因を排除する方法をとることもある。

しかしながら、人間が時間経過と共に内発的に変化すること、本来の環境 (ecology) から人間を分離する際に生じる問題、個人差が大きいことを考慮に入れるならば、物理学や化学で行われる実験的な隔離を人間に対して直

接的に適用することには問題がある。そこで、心理学ではこの点を補充するために、対照群や無作為配置のような農学の方法を借用しなければならなかった。

② 修正開放系 農学の研究法は、広い意味での開放系の方法である。なぜなら、種子の生育は、完全に条件統制された環境下ではなく、現実の生育条件に近似する環境下で観察されるからである。日照・降水・土壌の条件差のような妨害要因は、閉鎖系の方法のようにこれを除外するのではなく、むしろ妨害要因を各群に均等に分配することによって、問題を解決するのである。ただし、農事試験場は普通の畑と同じではなく、農学者は農夫と同じではない。したがってこの方法は、完全な開放系ではなく、修正開放系 (modified open system) の方法として位置づけるべきであろう。

社会領域における農学モデルの適用は、社会実験法 (social experimentation) という形で行われた。しかし、長期間にわたる社会政策の効果を無作為配置の手順によって実証することは、理論上は可能であるとしても、現実には倫理・政治・経費面の理由から非常に困難なことであった。もちろん、無作為配置の方法は、フィールド研究者にとって重要な手段であったが、開放系においては万能の方法 (panacea) ではなかった。

③ 開放系 準実験が学史的には後退した方法であるという指摘はまちがっていない。それは、多くの点で農事試験場のフィールド実験に似ているが、群の配置が無作為ではなく、自己選択や監督者の命令によって行われることが多いという点で、方法は明らかに後退している。しかし、他面において次のような独自性をもつ。

準実験の源が自然科学の閉鎖系モデルにあることはないし、実験心理学研究のように自然科学と農学の混合モデルにもない。基本的には、準実験が農学モデルに基礎をおく方法であるとしても、社会実験法のように無作為配置を重視して、自然の生態 (naturalism) を犠牲にすることはない。農学モデルのように模擬試験場を設定すること

はないし、対象が意識をもたないことを前提にすることもない。準実験では、問題が本来存在する場所で研究が行われるのであり、人間が能動的 (active) な存在であり、刺激に対して敏感に応答し (reactive)、対象に積極的に働きかける (proactive) 存在であることが重視される。^(二)

(四) 発展の方向

(1) 統一化と個別化

無作為配置型の実験が農学・医学・工学・心理学など多くの分野に共通に適用される統一的大理論 (グランド・セオリー) ないしは「超方法」ともいべきものであるとすれば、^(一五) 準実験は心理学や行動科学の対象特性を重視した個別的な方法である。統一化と個別化という対立的な発展方向は、心理学方法論史の中にしばしば見られるものであり、準実験もこの流れの中に位置づけることができる。

心理学の方法は、単一の方向に向かって発展したのではなく、科学一般の共通性を重視する方向と、心理学の独自性を重視する方向があった。前者が心理学から離れていく遠心的な動きがあるとするれば、後者は心理学に向かっていく求心的な動きである。前者が方法そのものの水準を高めることが方法の進歩であるとする考え方であれば、後者は対象それぞれがもつ独自の特性を明らかにすることが方法の目的であるとする考え方である。心理学方法論は、この二つの対立的な動きが相互に影響を及ぼしながら発展してきたと考えられる。

(2) 多変量解析の場合

多変量解析では、統一化に向かう動きが中心であったと思われる。変量は、それが経済学のデータであれ心理学のデータであれ、基本的には同じ特性をもつものとして処理される。^(一六) これに対して、個別性を重視する方法として

は、次の例がある。因子分析は、現在では多変量解析の一部になっているが、当初は心理検査の結果を分析する方法として開発され、心理学の中で独自の位置を占めていた。^(二七) 数量化理論は、世論調査を分析する過程で、質的なものを数量化する方法として、^(二八) 他の多変量解析とは異なる独自の発展をとげた。

(3) 尺度構成法の場合

心理測定家のサーストンは、知能を数量化するために無知能を前提にした絶対尺度を提案して、知能を表す係数を、自然科学でも通用する数値の水準に引き上げようとした。^(二九) これに対してステイヴンスは、心理学では零点の存在を前提にした尺度は実際的ではないと考え、^(三〇) 現在では統計学の基礎知識になっている尺度論を展開した。^(三一) 数学上の概念が零点・負数・無理数・虚数のように抽象的な方向へ発展したのとは反対に、ステイヴンスの尺度論の意義は、日常生活で使われる具体的な数値に着目し、それが尺度（間隔尺度・順序尺度・名義尺度）として有意味であることを明らかにしたことにある。サーストンが心理尺度を高い水準の方向に発展させようとしたのに対して、ステイヴンスは低い方向に心理尺度を発展させたのである。

この考え方は、他の心理測定家に受けつがれた。トীগソンは、零点をもつ両極性の順序尺度を提案した。^(三二) これによって、快・不快のような感情の特性がよく説明できる。クームズは、部分的にのみ整合的な部分順序尺度など低水準の各種尺度を提案した。^(三三) この尺度は、質問紙調査における回答傾向の分析に適合する。

(4) 発展の流動性

方法論を評価する場合に、それが完成された固定的性質をもつものと考えべきではない。方法は多様な発展方向をもち、つねに生成変化の途上にある。統一化・抽象化と個別化・具体化の対立は、異なる方法間の対立だけでなく、^(三四) 一つの方法内部の対立や、^(三五) 一人の研究者内部の矛盾という形で生じることもあるし、^(三六) 自覚的な対立もあ

ば、半自覚的・無自覚的な対立もあることに留意する必要がある。

(五) 他の方法との関係

(1) 単一事例実験との関係

ABA型の単一事例実験が基線と撤去という周到なコントロール手順を備えるのに対して、AB型は、事前の対照基準である基線をもつが、事後の対照基準である撤去の手順が欠けている^(一)。したがって、ABA型が無作為配置型の実験法に並ぶ方法であるのに対して、AB型は実験法に準じる方法 (quasi-experimental) と評価されること^(二七)がある。この用語法では、準実験は単一事例実験よりも水準の低い方法と考えられている。形式上では、中断時系列型準実験の基本型は、AB型と同じである^(三)の^(一)を参照^(二)。しかし、中断時系列型ではデータの時系列性を重視して、多くの観察点を設定することによって、コントロール・データを整備している。したがって本論では、準実験が単一事例実験と並ぶ方法であり、両者共に外見的には前実験に似ているが、実質的には正規の実験に準じる方法として位置づけた^(三)の^(二)を参照^(二)。準実験と単一事例実験は、方法の水準において異なるところはなく、社会と臨床という対象において区別されるのである。

(2) 時系列研究における位置づけ

データの時系列性を重視する方法を、「時系列研究 (time-series research)」という名称でまとめた場合^(六)に、単一事例実験は単一対象の方法 (single-N) として、中断時系列型の準実験は複数対象の方法 (multiple-N) として位置づけられている。多重基線型は単一事例実験の主要形式の一つである^(二)が、これを準実験に適用すると、中断時系列型と非等価対照群型を結合した形式となる。これは、時系列データを複数の群に分けて分析する方法である。

多重基線型は、単一事例実験よりもむしろ準実験に適合する方法ではないかと思われる。^(二八)

(3) 相関的方法および受動観察研究との比較

多変量解析は、主に従属変数間の相関関係に関わる (correlational methods)。これに対して、準実験と単一事例実験は、独立変数としての介入と、従属変数としての問題改善の間の因果関係に関わる (causal methods)。

また、多変量解析は、対象に対する働きかけなしに収集されるデータに対して適用されることが多いという点で、「受動観察研究 (passive observational research design)」と呼ばれることがある。^(二九) これに対して準実験と単一事例実験は、対象に対する積極的な介入の効果を分析する方法であるから、受動的な方法ではない。

ただし、相関的な方法や受動観察研究の中にも、分析の過程において因果的方法になるものもある。この場合には、「受動的観察から原因を推論」する方法 (inferring cause from passive observation) として位置づけられる。^(三〇) 社会調査のパス解析 (path analysis) では、たとえば、父の職業が子の職業に対して及ぼす因果的効果 (causal effect) を、子の学歴とつら因果経路 (causal path) を中間に設定して数理的な解析が行われる。

(1) Cook, T. D. Quasi-experimentation: Its ontology, epistemology, and methodology. In Morgan, G. (Ed.) *Beyond method: Strategies for social research*, Chapter 5 (pp. 74-94). Sage, 1985.

(1) Stouffer, S. A. Some observations on study design. *American Journal of Sociology*, 55, 355-361, 1950.
Campbell, D. T. Factors relevant to the validity of experiments in social settings. *Psychological Bulletin*, 54, 297-312, 1957. 必ずしも Cook & Campbell (1979) ・前掲 [序論の注 (1)] による。

(1) Campbell, D. T. From description to experimentation: Interpreting trends as quasi-experiments. In Harris, C. W. (Ed.) *Problems in measuring change*. University of Wisconsin Press, 1963. [Cook & Campbell (1979) による]。Campbell, D. T. & Stanley, J. C. Experimental and quasi-experimental designs for research. Rand McNally, 1963. (Barlow, D. H., Hayes, S. C., & Nelson, R. O. *The scientist-practitioner: Re-*

search and accountability in clinical and educational settings. Pergamon, 1984. による)。

- (四) Campbell (1969). 前掲「序論の注(一)」。田中潜次郎「単一事例実験の方法論的諸問題」(「医事学研究」第三号、一九八八年)。薬師寺(一九八九年)・前掲「序論の注(四)」。
- (五) Agras, W. S., Jacob, R. G., & Lebedeck, M. The California drought: A quasi-experimental analysis of social policy. *Journal of Applied Behavior Analysis*, 13, 561-570, 1980. 田中(一九八八年)・前掲(四)。
- (六) Kratochwill, T. R. (Ed.) *Single subject research: Strategies for evaluating change*. Chapter 1 (Foundations of time-series research). Academic, 1978.
- (七) Cook & Campbell (1979)・前掲「序論の注(一)」。
- (八) 回帰の概念を使う場合には、次の点に注意する必要がある。「平均への回帰 (regression-toward-the-mean) は、最初の検査と再検査の得点の相関に関連する現象である。相関が低いほど、測定誤差は大きく、平均への回帰も大きい。重要なことは、回帰という概念は、すべての極端な値(極値)が再検査では平均に逆戻りすることを意味しないし、個々の得点がすべて次の機会には極端でなくなることも意味しないということである。この現象は、一つのサンプルの部分(すなわち極値)についての変化を全体として示し、各部分が平均的にいかに応答しているかに言及するのである」(Kazdin, A. E. *Single-case research design: Methods for clinical and applied settings*, p. 79. Oxford University Press, 1982).
- (九) West (1985)・前掲「序論の注(三)」。
- (一〇) Corsini, R. J. (Ed.) *Concise encyclopedia of psychology*. Wiley, 1987. Experimental designs の項目。
- (一一) 田中潜次郎「臨床研究における単一事例の実験計画」(「医事学研究」第二号、一九八七年)。
- (一二) Cook & Campbell (1976)・前掲「序論の注(三)」。
- (一三) 準実験を前実験と混同した例として、次のような定義がある。すなわち、準実験 (Quasi-Experiment) は、「実験計画が不備であるために、その結果が妥当性を欠く実験(である)。準実験では、独立変数の把握とコントロールが不十分であるから、独立変数と従属変数の関係について、いかなる結論を得ることもできない」(Arnold, W., Eysenck, H. J., & Meili, R. (Hrsg.) *Lexikon der Psychologie*. Herder, 1980)。「また、準実験には、「疑似実験計画法」(H. J., & Meili, R. (Hrsg.) *Lexikon der Psychologie*. Herder, 1980)という訳語があてられる場合がある(「前掲・序論の注(四)」。「疑似」は、外見は似ているが実際はちがうという意味

であるが、準実験はその反対の特徴をもつ方法である。この訳語は、準実験本来の特徴を適切に表現していない。

(一四) Riecken, H. W. & Boruch, R. F. *Social experimentation*. Academic, 1974. [Cook (1985)・注(1)]。

(一五) 大理論 (grand theory) とは、社会学者ライト・ミルズが、「抽象的で規範的」な伝統的理論を、社会学の発展を妨げるものとして批判し、否定的な意味をこめて与えた名称である。

一方、科学的事実を重視する現代の方法論に対して批判を行い、問題を哲学的観点から再構成する立場を大理論の再生として位置づけ、これを肯定的に評価する考え方もある(クエンティン・スキナー「グランドセオリーの復権」産業図書・一九八八年)。すなわち、現代の大理論は、「自然科学が社会分野の学問のいとなみに十全にして適切なモデルを提供する」という前提に対して、「あらゆる成功的な説明は同一の演繹的なモデルに従わねばならない」という実証主義者の主張に対して批判を行い、「社会分野の学問のモデルを自然科学の伝統的なイメージに従って作る可能性にたいして哲学的懐疑」をもつ立場である。

心理学者グラウマンは、学派や主義と呼ばれた古典的な大理論が衰退したことを指摘して、従来これが果してきた役割について次のように述べている (Graumann, C. F. *The state of psychology. International Journal of Psychology*, 7, 123-134, 1972)。「心理学は自然科学でもなく精神科学でもないし、全体論でも要素論でも説明できない。古典的な意味での学派は死滅して折衷論が支配的になったが、古くからの争点の一つも解決していない。日常の研究は理論に憶病になって、これを避けている。小範囲や中範囲の理論と数学的モデル構成が、今日では△形而上学的▽といわれる大理論 (große Theorien) に取って代りつつある。しかし大理論の断念は、心理学の統一概念を断念することにつながりはしないだろうか」。

大理論の内容と評価については、以上のように多様な考え方があがるが、過去に大きな役割を果たした大理論が現代では中心的な位置を占めていないという認識は共通しているように思われる。本論ではこれに対して、規範的ではないが統一的で抽象的な大型の理論や方法が、現代心理学に大きな影響を及ぼしてきたと考え、その代表例として実験計画と多変量解析をあげた。それは、心理学が科学全体の中で占める位置を規定する際には有効であるが、心理学に固有の課題を扱う際には妨害的に働く恐れがある。

過去の大理論としての哲学的思想と、現代の大理論としての実験計画や多変量解析を比較すれば、次のようになる。前者は、対象に全体として接近するから、それが大理論であることが容易に理解できる。これに対して後者は、つねに

個別的問題を分析するために使われるから、外見的には小理論・小方法のように見える。その反面、多くの対象領域に広く浸透していることに着目した時に、それが大理論・超方法であることが理解できるのである。

過去の大理論は、グラウマンによれば心理学の統一概念に貢献したというが、それと同時に哲学と心理学の境界を不明瞭にしたのではないかと思われる。一方、現代の大理論としての実験計画や多変量解析は、自然科学と心理学の境界を不明瞭にしていると考えることができよう。前者が哲学的大理論であるのに対して、後者は科学的大理論である。

(一六) ケンドール「多変量解析」培風館・一九八一年。

(一七) 「心理学の分野で育った因子分析は、はじめ、能力とか性格に関する心理学的理論と密着し、ときには因子分析理論が心理学的現象の内容とさえかわりをもつようにみられたが、理論的検討が進むにしたがって、その統計的な性格が明らかになり、今日では完全に統計学的方法として位置づけられるに至った。とくに近年、電子計算機が発達普及するにしたがい、この方法の実用的利点が認識され、ひとり心理学ばかりでなく、他の社会科学・自然科学においても、もちいられるようになっていく」(芝 祐順「因子分析法」東大出版会・一九七二年)。

(一八) 林 知己夫・駒澤 勉「数量化理論とデータ処理」朝倉書店・一九八二年。

(一九) Thurstone, L. L. A method of scaling psychological and educational tests. *Journal of Educational Psychology*, 16, 433-451, 1925. [Stevens (1951) に引用]。

(二〇) 「たいいていの心理測定は間隔尺度を作ろうとし、それは時に成功する。その場合の問題は尺度の単位を等しくする操作を工夫することであるが、容易に解決できる問題ではない(中略)。△真▽の零点を定めることが必要なのは、心理学ではごくまれである。たとえば、知能は順序尺度上で測定し、これを間隔尺度に近似させる努力をすれば、それだけで有用なのであって、知能の無い状態 (zero intelligence) が何を意味するのかを定義することは必要ではない。ただしソーンダイクとサーストンと共に「これを試みてみる」[Stevens, S. S. (Ed.) *Handbook of experimental psychology*, p.27. Wiley, 1951.]」。

たしかに、知能の無い状態はもはや精神薄弱ではないし、人間関係の無い状態はもはや分裂病とはいえない。零点が何を意味するかについて永遠の議論を重ねるよりも、むしろこれを前提にしない形で数量化・尺度化の方向を探る方が現実的である。この場合に重要なことは、零点を欠く尺度でも十分価値があるという認識をもつことであろう。

(二一) 河口至商「多変量解析Ⅱ」(一～二頁)。森北出版・一九七八年。

- (一一一) Torgerson, W. S. *Theory and methods of scaling*. Wiley, 1958.
- (一一二) クームズはこの他に、名義尺度と順序尺度の間をさらに細分化する提案をしている (Coombs, C. H., Raiffa, H., & Thrall, R. M. Some views on methodological models and measurement theory. *Psychological Review*, 61, 132-144, 1954. [Torgerson (1958) を参照])。
- (一一三) 多次元尺度構成 (multidimensional scaling) やその隣接分野における基本概念としての距離 (distance) と近接性 (proximity) の係数は数多くある [岩坪秀一「数値分類法 (Numerical Taxonomy) について」(「心理学評論」一七巻・一二九～一四四頁、一九七四年)]。これら諸係数の概念設定の方向を一般化と個別化に分類すれば、次のようになる。ミンコフスキーの距離は、ユークリッドの距離をさらに一般化した概念である。これに対して、数量化三類で使われる近接性の係数は、質問紙調査の回答特性によく適合している点で、きわめて個別化した特徴をもつ。
- (一一四) 多次元尺度構成は、対象を間隔尺度から順序尺度へと水準を下げることにより (メトリックからノンメトリックへ)、具体的な方向へ適用を拡大する形で発展したのであるが、それと共に、ミンコフスキーの距離のような抽象化の方向での提案も行われている。
- (一一五) ステューヴンスは、低水準の尺度構成の方が時に有用であると論じたが、その一方で、零点をもつ比率尺度による感覚測定法も提案している [相場 覚「Stevens の新精神物理学」(八木・大山「講座心理学」第四巻・第八章、東大出版会・一九七〇年)。
- (一一六) Corsini(1987)・前掲(一〇) Single case research design の項目。
- (一一七) 多重基線型は、対象内部の諸活動間に相互作用がないことを前提にした方法である。これは準実験では適切な前提であるが、単一事例実験では問題がある [田中・前掲(一一)・四の(二)(多重基線型の特徴と問題点)を参照]。
- (一一八) Corsini(1987)・前掲書(一〇) Research methods の項目。
- (一一九) 青井和夫・直井 優「社会調査の基礎」サイエンス社・一九八三年。アッシュヤー「因果分析法」朝倉書店・一九八〇年。

二 非等価対照群型

非等価対照群型 (nonequivalent control group designs) の準実験は、対照群の被験者が介入群と全く異質ではないが、さりとて等しいともいえない集団 (nonequivalent) で構成される形式である。これ以外の点では、非等価対照群型は正規の実験法と同じである。正規の実験法と同様に、測定は単発型であって、原則として介入の前と後にそれぞれ一回だけ行われる。この点において、反復的測定が不可欠の手順である中断時系列型の準実験 (三を参照) とは異なる特徴をもつ。

ここで使う主要な用語は、次の通りである。実験計画でいうところの実験群は、問題解決のために介入を行う群という意味で、介入群 (treatment group) と呼ぶ。対照群は、非等価対照群ないしは無介入対照群 (untreated control group) と呼ぶか、あるいは単に対照群と呼ぶ。ただしこの場合には、正規の対照群を意味しない。

記号は次の通り。Xは、介入すなわち処置の導入 (implementing treatments) を表す。Oは、測定ないし観察 (observation) を表す。これが反復される場合には、 X_n は n 回目の介入を、 O_n は n 回目の観察を表す。

(一) 不完全な形式

ここで述べる方法は、実験以前の形式「一(三)の(2)を参照」とみなされており、外見的には準実験に似ているが機能的には準実験とはいえないものである。これは、独立変数と従属変数の関係とは別の因果説明の可能性を取り除く手順を欠いている。したがって、これによって「因果関係を推論することは原則として許されない」のであ

るが、この種の研究から「しばしば新しい着想がもたらされた」ということは認めなければならない。^(二)

(1) 事後測定^(三)のみの単一群型

これは、単一の集団に対して介入 X を行い、その後一回の測定 O を行う場合である。手順の図式は次の通り。



この形式には、次のような欠点がある。第一に、介入に先立つ事前の測定がない。もし介入の後で問題が改善されたとしても、介入がその原因であったのか、それとも集団内部に改善傾向があったためであるかが不明である。

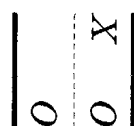
この形式は、成熟すなわち事前の改善傾向の影響を点検する手順を欠いている「(一)の(二)の(1)を参照」。

この例として、次のような場合が考えられる。^(二)ある学校で、幾何や算術よりも代数に重点をおく新しい数学カリキュラムが実施されたとする。課程終了後に試験を行ったところ、生徒の成績は、代数では全国平均を上回り、幾何と算術ではこれを上回らなかった。しかし、これにより新カリキュラムの実施とその後の成績の間に因果関係があるということはできない。実施前の内部傾向や偶然という要因の介入を点検する手順が欠けているからである。

実務を中心とする現場では、研究計画が整わないままに必要に迫られて介入が行われ、結果的にこの形式で研究が行われることが少なくないと考えられる。しかし、事前データに類するものがすべて欠如しているということも少ないのであって、実際にはその代用となる何らかの記録 (archival records) が残っていることがある。これを利用すれば、次善の準実験を行うことが可能となる「(三)を参照」。したがって、ここで述べた事後測定^(三)のみの単一群型は、現実にはありえない純粹化された架空の前実験であるか、^(二)そうでなければ、未熟な研究者が行う実証的研究だけにみられるものである。

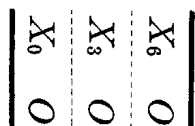
(2) 事後検査のみの非等価群型⁽³⁾

これは、前項(1)の形式に、無介入の群を比較対照のために付加した形式であるが、事前測定の手順を欠くという点では前項(1)の形式と同様である。図の中で、上段と下段の間にある点線は、両群が別の被験者の集団であるし、所属する母集団も異なることを表す。



方法の長所は、比較対照の群を付加して、介入以外の外部要因を除去することにある。一方、事前測定を欠くために、事後測定において両群の間に差があったとしても、その原因が介入の有無にあるのか、それとも群を構成する際に被験者を選抜する時点で両群に差があったのかが不明である。

この形式は、三群以上を比較する時にも使われる。たとえば、投薬量の差が患者に及ぼす効果に関する研究や、次の図に示すように、刑務所の仮出所者 (parolees) に対する出所後のカウンセリング期間の差 (六ヶ月・三ヶ月・なし) が再犯 (recidivism) の防止に及ぼす効果に関する研究などが考えられる。



この場合に、カウンセリングによる介入の期間と事後の再犯件数の間に相関があるならば、カウンセリングが再

犯防止の効果をもつといえる可能性は確かにある。しかし、長期間のカウンセリングに応じる人は、再び罪を犯すことの少ない人でもある可能性がある。つまり、事後の群間の差がカウンセリングという介入によって生じたのではなく、カウンセリングに応じた時点で群間の差がすでにあつたと考えることもできるのである。したがって、カウンセリングの効果を実証するためには、事前に群の間にある差を点検する手順が必要になる。

経済学や社会学では、事前測定の欠如という問題を克服しようとする試みが以前から行われてきた。たとえば、調査対象者の年齢・性・社会階層・人種・出生地・現住所などが、事前の傾向を推定するための次善の手段として使われてきたことはよく知られている。^(一)

(3) 事前事後測定付き単一群型^(四)

これは、社会科学でよく使われる形式の一つである。次の図に従ってみれば、調査対象とされた単一の集団に事前の測定 O_1 を行い、次に介入 X をした後、事後の測定 O_2 を実施する。これが前項(2)と異なるのは、事前測定を行うという長所をもつことと、比較対照のための群を欠くという短所をもつことである。

$$\begin{array}{|c|} \hline O_1, X, O_2 \\ \hline \end{array}$$

この例として、次のようなものが考えられる。工場の生産性を向上させるために、現場主任による監督体制を変更すると、変更後の生産性が変更前よりも明らかに向上したとする。これによって、監督体制の変更が生産性向上の原因となつたといえるか。この場合の問題点は、次の通りである。

第一は、歴史の要因である。現場の監督体制の変更と同時に、他の条件の変化はなかつたか。たとえば、給与表や組合対策の変更や、新しい訓練計画の導入はなかつたか。さらには、全国的に輸出力強化が提案されたり、暖か

い季節になって労働者が働きやすくなったということはないか。このような外部要因が生産性に影響を与えたという説明の可能性を除去するためには、これらが調査結果に影響を及ぼすことが論理的にありえないことを示すか、論理的にはあるとしても現実にはありえないことを示す必要がある。外部要因の影響を点検するための最も有効な手順は、対照群を設定することである。

第二は、回帰の要因である。監督体制は、何故に変更されなければならなかったのか。おそらく、生産性低下が忍耐の限度をこえるほどになって、早急に対策を講じる必要があったためであろう。ここで生じる問題は、事前測定時の低水準の生産性が恒常的な安定した水準を表しているのか、それとも不意に起ったストライキや資材調達の一時的な遅れによるものかである。状況が前者の場合であれば、問題はない。これに対して、もし後者が原因であるとすれば、生産性の低下は確率変動の結果であるから、監督体制の変更のような特別な介入を行わなくとも、生産性は本来の水準に回帰する傾向をもつのである。この状況は、一回の試験の成績が極端に悪い生徒を集めて実施した補習授業に似ている。^(五) 補習の有無にかかわらず、生徒の成績は母集団本来の水準に戻って、事前測定時よりも向上する可能性がある〔一(二)(3)を参照〕。

第三は、成熟の要因である。生産性が毎年安定した水準を保つことはむしろ少なく、変動するのが普通である。これには、前述の統計的回帰をもたらす確率変動だけではなく、系統的な変動も含まれる。労働者の経験年数とか機械の性能向上などによる生産性の持続的向上がなかったかどうかを点検する必要がある。このためには、単発型の事前測定では不十分であって、連続的な時系列測定の方が適切である。

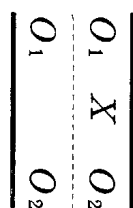
(4) 方法を評価する際の留意点

物理学者は、実験室の中で対象を隔離することによって、単一群による方法で因果関係を実証することができ

る。そこでは、歴史や成熟などの妨害要因は問題にならないからである。物理学者とは一見著しく異なる文化人類学者も、物理学者と同様の方法を使って因果関係を実証する。たとえば、斧のような新しい道具が未開部族に及ぼす影響を調査する際には、文明社会から隔離された密林で生活する部族を選び、その部族の事前の生活習慣が長い間変っていないことを確認する。このような物理学者や文化人類学者とは反対に、社会学者は、実験とは何の関係もない別の原因によって変化する対象について、それが現実に存在する場面の中で因果関係を説明しなければならぬのである。

(二) 標準型

事前事後測定付き無介入対照群型は、社会科学研究で最もよく使われる形式であり、非等価対照群型の標準型である。これは、介入を行わない群を比較対照のために設定する点で、事後検査のみの非等価群型〔(一)の(2)を参照〕と同じ特徴をもち、介入前に事前測定を行う点で、事前事後測定付き単一群型〔(一)の(3)を参照〕と同じ特徴をもっている。したがって、この二つの方法を結合したのが標準型の非等価対照群型であるといえよう。これを図示すれば、次のようになる。点線の上が介入群、下が対照群である。Oに添えた数字は観察の回数を表す。



この形式は、歴史・成熟などの基本的な妨害要因を除去することができるといって、前節で述べた前実験的形式よりも整備された方法である。しかしながら、すべての妨害要因がこれによって除去できるわけではないから、

準実験の結果を解釈するためには、前実験の場合に劣らない多様な観点が必要になるのであり、この点でも正規の無作為配置型実験とは異なっている。

なお、正規の実験法では妨害要因が事前に条件統制されていることになっているから、実験結果の解釈は、独立変数との関連において一様に行われる。これに対して準実験では、事前に妨害要因を完全に条件統制することは不可能であるから、実験結果の種類ごとに、これに関わる妨害要因を個別的に推定するための観点が必要になる。すなわち、「解釈可能性は、一つの研究で得られた特定の結果に依存する^(一)」のであるから、実験結果によって解釈の方法が変わるということに留意しなければならない。

また、準実験における介入群と対照群は、正規の実験法とは異なって、当該の問題に関して同等ではないから、すでに事前測定の段階で両群の間に平均値の差があることに注意しなければならない。たとえば、ある県で新しく制定された条例の効果を確かめるために、同様の条例をまだ制定していない県を比較対照の基準として利用したとする。この場合に、ごく限定された問題についても二つの県が同じ水準にあることはむしろ少なく、経済的・社会的・地理的・歴史的に異質な集団であることの方が一般的である。もちろん、著しく条件の異なる県を対照群とすることもありえないが、可能な限り類似する県を探したとしても、事前測定の段階で介入群と対照群の県の測定値が等しくなることはほとんどないと推定されるのである。

(1) 対照群に変化がなく、介入群だけが上昇する場合(図2-1)

事前測定において、介入群がすでに対照群を上回っている。事後測定になると、介入群がさらに上昇するのに対して、対照群は事前の水準を保っているから、事後の両群の差は事前よりも大きくなっている。問題は、その原因が介入の有無にあるかといえるかということである。

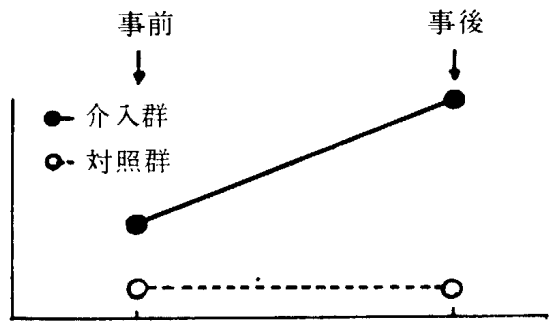


図 2-1 介入群だけが上昇する場合

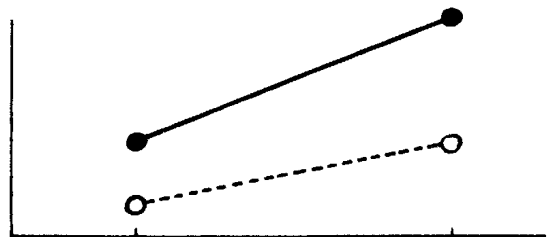


図 2-2 両群の上昇率に差がある場合

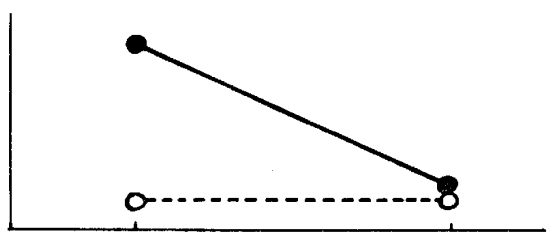


図 2-3 介入群の優位性が消滅する場合

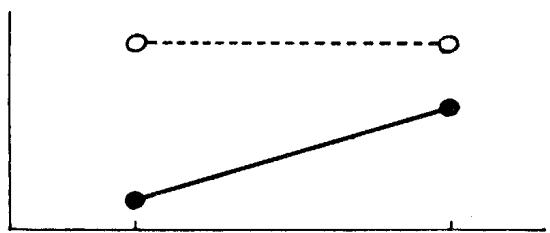


図 2-4 対照群の優位性が減少する場合

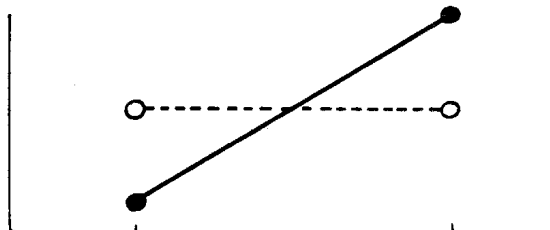


図 2-5 解釈が最も容易な場合

① 選抜×成熟の要因

これが結果に影響を及ぼすのは、両群の内部条件が異なる場合である。たとえば、一方の群が他方よりも経験があるか、疲れているか、退屈しているような場合に、この要因が作用する。もし実験が行われるかなり以前から、対照群が持続的な停滞傾向をもっていたのに対し、介入群は持続的な上昇傾向をもっており、図 2-1 における事前から事後までの上昇率はそれまでの傾向の延長線にあるにすぎなかったとする。この場合には、実験結果は介入の有無によって生じたのではなく、両群が本来的にもつ内部条件に原因があると考えなければならない。

② 記録条件の要因

群の平均値が尺度上の上限 (ceiling) か下限 (floor) にある場合、たとえば試験の平均点が満点か零点の近くにある場合には、測定値は尺度構成法でいうところの等間隔性に疑いがある。図 3-1 の

対照群の値が事前から事後に向かつて変化しなかった原因が無介入にあるのではなく、多くの被験者の得点が下限にあったために、それ以上の変化がなかったという可能性はある。この点は、得点の度数分布を群ごとにとって確かめる必要がある。

③ 選抜×歴史（地方史）の要因　意思決定参加という新しい業務管理方式を地上で働く労働者に対して実施したが、地下作業に従事する労働者には従来の管理方式を継続させたとする。意思決定参加方式が労働意欲に及ぼす影響を知るために、地上労働者を介入群、地下労働者を対照群として、春から夏にかけての気候が快適になる時期に実験が行われたとする。その結果が図3—1の通りであったとしても、地上労働者群の労働意欲向上の原因は新管理方式ではなく、気候の変化が地上の労働に対してのみ影響を及ぼすという地方史の要因によってこれが生じたという可能性がある。

④ 群によって差別的に生じる回帰 (differential statistical regression)　正規の実験の介入群（実験群）と対照群は、同一母集団から無作為抽出されるから、互いに等質な集団である。これに対して準実験の介入群と対照群は、当初から非等質な構成である。とはいえ、両群はできるだけ類似していることが実験計画上望ましいことであるから、先に設定された介入群の構成に近似するように、対照群を意図的に作りあげることがある。つまり、実験群が持続的に存在する安定した集団であるのに対して、対照群は事前測定の時点で急に作られた恣意的な集団である。この場合には、対照群の個々の測定値は本来の母集団の水準に回帰する傾向がある。

(2) 介入群と対照群の上昇率が異なる場合（図2—2）

① 選抜×成熟の要因　介入群と対照群の差が介入前よりも介入後に大きくなったとしても、その原因が介入の有無にあるとは必ずしもいえない。むしろ、各群が本来もっていた上昇傾向の程度が異なっていたために、両群

の差が時間経過と共に大きくなった可能性がある。これは、金持はますます金持になり、有能な者はますます有能になる、という格言の状況に似ている。

たとえば、生徒を成績によって二群に分けて、成績上位群を介入群として新しい教育法を実施する一方、成績下位群を対照群として従来の方式で教えると、図2-2の結果が得られたとする。選抜×成熟の要因がこの結果に作用した可能性があるが、これを分析するには次のような観点が有効である。つまり、上昇率の差が群間にあるならば、それは群内にもあるはずである。群内部で、より上位の生徒とより下位の生徒の上昇率に差があるなら、群内の得点差（群内分散）は、事前測定時よりも事後測定時に大きくなると予想される。これとは逆に、事前と事後の群内分散に差がなければ、選抜×成熟の要因は作用しなかったということになる。

② その他の要因 図2-2に対して影響を及ぼす可能性のある妨害要因は①の要因だけではなく、図2-1と同様に地方史や記録条件も考慮しなければならない。

(3) 介入群の優位性が事後測定時に消滅する場合（図2-3）

図2-1と図2-2では事前事後共に介入群が対照群よりも優位に立つのに対して、図2-3は、介入群が事前測定時にもっていた優位性が、事後測定時になって減少もしくは消滅する場合である。

その例として、学校統合が黒人生徒の学力の自己評価 (academic self-concept) に及ぼす効果に関する研究がある。^(七) 被験者は、小学校三年生・四年生・五年生の黒人生徒である。この中で、介入群は、事前測定時には黒人学校に通学し、学校統合によって事後測定時は統合学校 (integrated school) に通学するようになった生徒である。対照群は、介入群と同じ学区内の統合学校に事前測定時に引き続いて事後測定時も通学する生徒である。介入群の生徒は、事前には自己の学力を高く評価していたが、学校統合後には対照群と同程度になっている。

通常の実験計画における対照群は、実験的な操作や介入が行われないう群である。これに対して、この研究の対照群は無介入群ではなく、介入群よりも前から介入が行われ、それが継続されている群である。準実験では、このような形で対照群を設定することが少なくない。論点は図2-1や図2-2と同様であって、選抜×成熟の要因が主要な問題である。

(4) 対照群の優位性が事後測定時に減少する場合(図2-4)

その例として、学校において学力の低い生徒達に補習を行うことによって他の生徒の水準に近づける場合や、工場において生産効率の低い集団に対して生産性を向上させる対策を講じる場合が考えられる。

これに対する論点は、前項までとほぼ同様である。たとえば、対照群の測定値が変化しないのは、それが尺度の上限に達している(ceiling effect)からではないかという問題(記録条件の要因)や、両群が互いに異なる外部要因の影響を受けているのではないかという問題がある(地方史の要因)。しかし、このような一般的問題の他に、この結果に特有の問題といえるのが、次の二点である。

① 回帰の要因 群を構成する際に適切なマッチングが行われている場合や、群間の差が持続的に安定している場合には、回帰の問題は生じない。これに対して、図2-4の介入群がつねに低い点数をとる被験者の集団ではなく、一回だけの事前測定の点数が異常に低い人を集めて作られた集団である場合には、次回測定時の得点は母集団本来の水準に回帰する傾向がある。介入群の得点が事後測定で対照群に接近したのは、介入によってではなく確率変動によって生じた可能性がある。

② 選抜×成熟の要因 これは、標準的な非等価対照群型における主要な問題点であるが、図2-4の結果ではこの要因の影響が取り除かれている。なぜなら、金持はますます金持になるといのように、高得点群の方が低得

点群よりも高い上昇率を示す場合とは反対に、この結果では高得点の対照群は得点が上昇せず、むしろ低得点の介入群が上昇しているからである。もし、選抜×成熟の要因がこの結果に影響を及ぼしたと仮定しても、それにもかかわらず低得点群の水準に接近したということは、実験的介入がこの要因による逆方向の作用に勝るほどの強力な効果を及ぼした可能性があることを示している。

もちろん、すべての成長現象について、高得点群が低得点群よりも内発的に速く成長するという前提をおくのは適当ではない。この前提は、個別的には現実の事前測定にあらわれた成長傾向について、より一般的に見ても、ある特定領域だけに適用される原理として考えるべきものである。

(5) 解釈が最も容易な場合(図2-5)

この結果は、介入群が事前測定では対照群よりも低得点であったが、介入の後になると対照群を上回った場合を示している。これは、妨害要因が介在する恐れが最も少なく、したがって介入の有効性を実証することが最も容易な結果である。

① 上限ないし下限効果の問題 図2-1と図2-4の対照群が事前から事後に向かって変化しないのは、無介入が原因なのではなく、零点や満点のように尺度の下限か上限にあつて、そもそも変化不可能な記録条件があつたためであるという可能性がある。図2-5の対照群は、事前から事後に向かつて変化しない点では前二項と同じであるが、測定値が尺度の中間にあるから、上限・下限効果が介在したという可能性はない。なぜなら、仮にこの要因が関与していたとすれば、介入群は、事前測定で対照群を下回ることがなく、事後測定で対照群を上回ることもないからである。したがって、対照群の得点が変わらなかったのは、上限・下限効果のような記録条件が原因なのではなく、無介入が原因であることが示唆されるのである。

② 回帰の問題

群の配置が一回の事前測定の得点によって行われていたとしたら、その時点で低得点であった介入群は、事後測定では母集団の平均水準へと回帰して、得点が高くなる可能性がある。ただし、母集団の平均水準が対照群よりも上位の水準にあるとは考えられない。したがって、介入群が事後測定で対照群を上回ったことは、平均への回帰によっては説明できないのであって、介入によってこれが生じたといえる可能性がある。

③ 選抜×成熟の問題

テレビ番組「セサミ・ストリート」の教育効果に関する学力検査機関 (Educational Testing Service) のデータを再分析した研究がある⁽⁸⁾。これによれば、この番組を見るように指導された地区の児童 (介入群) は、指導が行われなかった児童 (対照群) よりも、指導以前の時点では学力が低かった。介入群に対する指導が行われた後の測定では、介入群の学力は対照群を上回った。この結果は、図2—5に類似している。

選抜×成熟の妨害要因がこの結果に影響を及ぼした可能性は少ない。仮にこれが介在していたとすれば、事前に低得点であった介入群の上昇率が対照群を上回ることを予測することはできない。しかし、介入群は高い上昇率を示し、対照群の得点を上回った。

しかしながら、この要因がもっと複雑な形で結果に影響を及ぼすのはありうることである。すなわち、介入群の児童が対照群よりも年少であり、知能水準が高かったとすれば、介入群は年少であるが故に初めは得点が低く、高い知能の故に向上率が高く、後に対照群を上回ったと説明することは可能であろう。しかし、この研究における介入群は、年齢では差がなく、介入前の知能 (cognitive aptitude) 検査でも差がなかったことが確認されている⁽¹⁾。

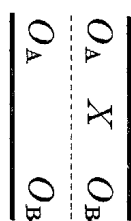
(三) 事前測定を変更した非等価対照群型

事後測定のための非等価群型〔(一)の(2)を参照〕と標準的な非等価対照群型〔(二)を参照〕を比較すれば、いずれも対

照群を配置しているという点で同格である。しかし、事前測定の有無という点から評価すれば、これを欠く前者の形式は、これを備える後者よりも不完全な方法である。ここでは、事前測定の手順を欠く非等価群型に対して、事前測定の代用となるようなデータを付加することによって、方法の機能を標準的な非等価対照群型の水準に引き上げるための手順について述べる。したがって、次に述べる方法は標準型の変法であるが、標準型をさらに発展させたものではなく、むしろこれに準じる機能をもつ方法である。

(1) 別種の測定を事前測定の代用とする無介入対照群型^(九)

事後測定と同じ手順で事前測定を実施することができない場合に、事前測定とは形式上異なるにしても実質的にはこれと関連のある事前データを両群の被験者から採し出して、正規の事前測定の代用データとして使う手順である。次の図のOに添えられたAとBは、事前と事後が異種のデータであることを意味している。



たとえば、代数の新しい教育法を実施して、その有効性を実証するために、介入群のクラスの生徒には新方法による授業を、対照群のクラスでは従来通りの授業を行ったとする。この実験では、事前測定を行うことができない。なぜなら、代数の授業をまだ受けていない生徒に、代数の試験を行うのは意味がないからである。そこで、事前に行われている一般的数学適性検査の成績を、事前の代数試験の代用(proxy for the pretest or proxy tests)として使うことができる。図によって説明すれば、一般的数学適性検査が O_A であり、介入後の代数試験が O_B にあたる。なお、事前測定の代用となる検査は、事後測定との間の相関が低いものであってはならない。たとえば、国語

の成績や体力検査の結果をこの場合の代用データとして使うことはできない。

また、ある会社で新任の管理職に対して、経営管理のための一年間の夜間訓練課程を開設したとする^(二)。新任管理職のうちの三〇人がこれを受講することになったが（介入群）、その他の五〇人は、これに興味がないとか自分には向かないなどの理由で、受講しなかった（非等価対照群）。年度末になって、八〇人の新任管理職は全員で、訓練課程の具体的内容とは別の経営管理能力検査（ O_B ）を受けた。その結果、夜間課程を修了した介入群の管理職の成績が対照群よりもよかったとする。しかし、その原因が夜間課程の受講にあるとはいえない。なぜなら、群の配分は無作為ではなく、被験者の自己選択によって行われており、介入群の被験者は一年間の受講以前から適性をもつ集団であった可能性があるからである。

そこで、介入群と対照群が事前ほどの程度等質ではなかったかを調べる必要が出てくるから、被験者各人のファイルから、以前実施した各種適性検査の結果を取り出してくる。各種検査は幸いにして、今回の経営管理能力検査に対して一定の相関があったので、以前の検査が全体として今回の検査に対してさらに高い相関をもつように、各検査ごとに重みづけをして事前データ全体の合成得点を計算する。この合成得点が、正規の事前測定^(一)の代用（ O_A ）として使用されることになる。

(2) 別人のデータを事前測定の代用とする形式^(二)

前項(1)は、同じ被験者から事後測定とは別種のデータを入手して、これを事前測定の代りにする場合である。これに対して本項の手順は、事後測定と同種のデータを別の被験者から入手して、これを事前測定の代りにする場合である。次の図の中で事前測定と介入・事後測定の間にある縦の点線は、事前と事後が別の被験者であることを示している。

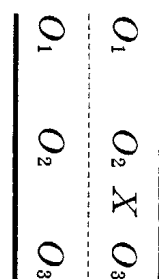
0	X	0
0		0

次の例は、原動芝刈機についての広報活動が芝刈機の安全使用に及ぼす効果を評価した研究である⁽²⁾。二つの町が選ばれ、一つの町は介入地域 (treatment site) として、広報活動と個人指導が行われた。もう一つの町は対照地域 (control site) として、介入は何も行われなかった。さらに、各々の町の内部で、被験者が事前測定群と事後測定群に分けられた。

その結果、事後測定における介入群の安全使用についての知識は対照群を上回っていた。しかしながら、広報と指導が知識差の原因であるという結論には、次の点で疑問がある。すなわち、事前測定の被験者と事後測定の被験者は、人口学的特徴 (性・年齢・学歴など) の面で、一つの地域については非常によく似ているから、同じ母集団に属するといえる。これに対して、もう一つの地域の事前と事後の被験者は等質ではなかった。したがって、前者の事前代用データはその町の実験以前の状態を表しているのに対して、後者のデータはもう一つの町の事前状態を表しているとはいえない。この代用データは全体として、二つの町が実験以前にどの程度非等質であったかについての適切な情報を与えていない。したがって、事後測定における地域差の原因が介入にあるのか、すでに事前にあった地域差にあるのかを知ることができない。ただし、被験者を高学歴と低学歴に分けると、高学歴層では事前と事後の被験者の人口学的特徴が一致しなかったが、低学歴層では事前と事後が一致していた。したがって、低学歴層については、別人のデータを事前測定の代用とすることができ。

(3) 事前測定を何度も行う形式⁽³⁾

前二項の手順が標準型に準じる形式に留まるのに対して、この手順は標準型をさらに強化した形式である。次の図にあるように、これが標準型と異なる点は、事前測定をもう一回付け加えたことにある。



二回以上の事前測定を行う手順には、次のような利点がある。前実験的形式において注意すべき妨害要因は、成熟や歴史のような基本的な要因である〔一〕を参照〕。これに対して、標準的な非等価対照群型準実験では、選抜×成熟の相互作用の方が主要な妨害要因となる〔二〕を参照〕。つまり、介入群と対照群の間に内発的変化率の差がありうるということが結果の解釈を妨害するのである。この点で、二回以上の事前測定を行うことは、各群の内発的変化の傾向を知る上で有益な手順である。ただし、この要因の介在を徹底的に点検しようとするれば、二回程度の事前測定よりも、長い時系列の事前データ〔三を参照〕の方がさらに有効であろう。

一回だけの事前測定の得点によって群を配置した場合には、その群は次の測定時に、母集団本来の水準に回帰する傾向がある〔一〕の(3)、〔二〕の(4)を参照〕。もし、二回以上の事前測定を行うならば、この回帰の要因が条件統制され、その群本来の水準を知ることができる。

事前測定を二回以上行う手順にはこのような利点があるが、これが実行されることはむしろ少ない。なぜなら、準実験は実務領域で行われることが多いが、そこでは事前測定と対照群の意義が十分理解されないからである。第一に、緊急を要する処置の導入を故意に遅らせて、その前に事前測定を行うことができること自体が異例であって、まして二回以上これを反復するのは不可能に近い。第二に、研究費用の支出責任者は、介入とその後の測定に

ついでに支出は認めるが、介入とは直接関係のない事前検査や対照群の費用を負担しようとする傾向がある。⁽¹⁾

(四) 対照基準を介入群内部に設定する単一群型

単一群による事前事後測定形式(一)の(3)を参照は、事前事後の測定を行う点で標準的な非等価対照群型(三)を参照と同格であるが、対照群を欠く単一群型である点で標準型より不完全な方法である。何らかの理由があつて対照群を設定できない場合には、対照群に準じる機能をもつもの(対照基準)を介入群の内部に設定することが望ましい。前実験的な形式はこれによつて、準実験に近い役割を果たすことが可能になる。したがつて、次に述べる手順は前節(三)と同様に、標準型の発展形式ではなく、むしろ標準型に準じる形式である。

(1) 非等価従属変数型——介入群内部における別種の測定を対照基準とする形式

これは、単一の介入群に対して事前事後の測定を行う方法であるから、被験者の構成については単一群事前事後測定の前実験的形式(一)の(3)を参照と全く同じである。しかし、同一の被験者に対して二種類の測定を行い、介入を行う部分に対する測定を介入「群」として、無介入の部分の測定を対照「群」とする点において、前実験的形式とは異なる特徴をもつ。非等価従属変数型は、外見上は前実験的でありながら、機能的には標準的な非等価対照群型の準実験に類似した方法である。ここでいう介入「群」と対照「群」は、互いに異なる被験者群ではなく、同一の被験者群内部における異なる測定尺度ないしは従属変数を表す。

O_{1A}	X	O_{2A}
O_{1B}		O_{2B}

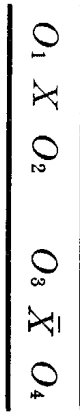
この図において、 O_A は介入データの測定値を表し、 O_B は対照データの測定値を表しているが、両者は共に同じ被験者から得たものである。なお、 O_A と O_B は異なっているが、さりとて全く無関係であってはならない (related but different)。極端な例をあげれば、新しい産業機械の導入が生産性 O_A には効果があったが、労働者の髪型 O_B には影響がなかったという調査結果が出たとしても、髪型という従属変数は有効な対照データとはなりえない^(二)。

この型の適用例として、騒音が産業生産性に及ぼす影響に関する実験がある^(三四)。結論は、騒音によって作業の一时的な失敗は増加するが、全体的な作業量には変化がないということである。実験は、フィルムの縁に穿孔をする作業を対象にして、実際の作業現場で実施されている。この工場は、従業員に支払う給料計算の都合上、毎日の穿孔作業量と穿孔に失敗したフィルム数の規則的な記録があったので、実験の面でも都合の場面であった。従属変数については、穿孔に失敗したフィルム数が作業の「一时的失敗」を表し、機械作動期間中の作業量が「全体的作業量」を表す、と操作的に定義されている。また、作業記録 (archival data) がよく整備されていたので、実験的介入 (騒音を出すこと) の前後と、騒音を停止する前後のデータを入手することができた。なお、この研究は、騒音を出した作業室とは別に、従来通りの条件の作業室のデータも得ていた^(三)ので、全体としては標準的な非等価対照群型の実験であった。ここに示した非等価従属変数型の測定はその一部である^(一)。

この方法は、次のような問題にも適用することができる。新しい数学教育法が幾何の学力に及ぼす効果を評価するために、幾何の学力を介入データとして、同じ生徒の分数操作能力を対照データとした分析を行うことができる。これを標準的な方法と比較すれば、次のような利点がある。つまり、同じ被験者を対象にして実験が行われるから、第一に、その個人全体の成熟要因を条件統制することが容易になることと、第二に、検査効果が介入と対照の両データに均等に及ぶから、その差別的効果を心配する必要がないということである。

(2) 処置の撤去を伴う単一群型^(二五)

これは、単一の介入群内部で処置を導入(X)する前後に測定を行うだけでなく、その後に、処置を中止ないしは撤去(\bar{X})して、その前後に測定を行う手順を追加した形式である。次の図は、 O_1 の後で処置が導入され、 O_3 の後で処置が撤去されることを示す。したがって、 O_2 と O_3 は介入時の測定であるが、処置導入前の O_1 と処置撤去後の O_4 は、いずれも無介入時の測定である。



この手順を図2-6に沿って説明すれば、次のようになる。 O_1 の後に処置を導入すると、図のように O_2 が増加して、 O_2 から O_3 へわずかに減少したとする。

問題はその後であって、 O_3 の後で処置を撤去すると、 O_4 の水準がどの程度になるかである。もし、図の(a)のように O_2 から O_3 への傾向の延長線に O_4 があるとするれば、処置の有無が影響力をもたないのであるから、処置以外の要因が作用している可能性がある。これに対して、図の(b)のように O_4 が介入前の O_1 の水準に戻ったとすれば、介入が有効であったことが示される。この形式は、処置の有効性をその導入と撤去によって二重に確認する手順であって、二種類の事前事後測定付き単一群型「(一)の(3)を参照」を結合させたものである。 X と \bar{X} の間に、単一では複数の観察点(O_2 と O_3)を設けるのは、 \bar{X} 以前の傾向性(増加・減少・安定・不安定)を点検するためであると思われる。

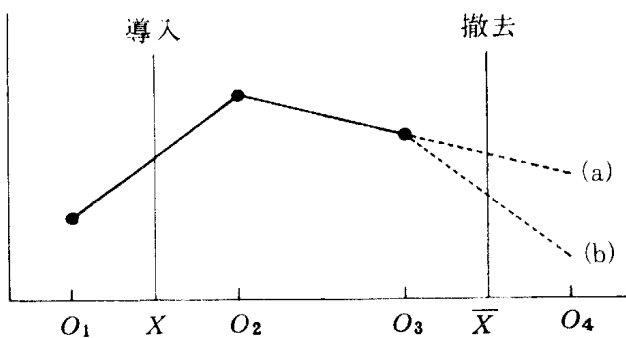


図2-6 処置の撤去を伴う型の結果

これに似た形式によって、役割変化に伴う態度変容を分析した研究がある。^(二六) すなわち、現場主任 (foreman) に昇任したが、その後地位を失って再び労働者に降任となった一群の人に対して、現場主任になる前 (O_1)、現場主任になった後 (O_2)、再び労働者に戻った後 (O_3) の三つの時点で態度測定が行われた。しかし、処置の導入 (昇任) と撤去 (降任) の間の測定が二回以上ではなく一回だけであるから、その後の O_3 の態度変容が降任によって直接的に引き起されたのか、それとも現場主任に在任中にすでに変容傾向があつて、その延長線に O_3 があるのかを区別することができない。

(3) 処置の導入と撤去を反復する単一群型^(二七)

処置は次の図のように、導入と撤去が反復される。一回目の観察 O_1 は無介入の状態で行われ、 O_2 は介入時に、 O_3 は無介入時に、 O_4 は介入時に実施される。

$O_1 \quad X \quad O_2 \quad \bar{X} \quad O_3 \quad X \quad O_4$

介入が有効である場合には、無介入から介入へ移行する O_1 から O_2 、 O_3 から O_4 への測定値の変化は同じ方向に生じると予想される。この手順は、スキナー派の動物行動分析や、労働組織と生産性の関連を人間関係の観点から分析したホーソン研究において使われてきた。^(二八) なお、動物行動分析の方法論は、現代の単一事例実験の基礎になっている。

ホーソン研究の中に^(二八)、休憩が生産性に及ぼす効果を女性労働者を対象にして検討するために、同じ休憩時間を二回の異なる時点で導入した実験がある。この手順の問題点は次の通りである。 O_1 は金曜日の午後に記録され、 O_2 は次週火曜日の午前に、 O_3 は金曜日の午後に、 O_4 はその次の週の火曜日の午前に記録されている。無介入時 (休憩な

し)の O_1 と O_3 は共に金曜日に実施され、介入時の O_2 と O_4 は火曜日に実施されたのであるから、一週間の内部的周期変動(cyclical maturation)の要因の影響と、介入効果を区別することができない。

処置の導入と撤去を反復するのは、無介入時の状態を基準にして介入の有効性を評価するためであるが、これを定期的に反復すると、測定値に周期性が生じる恐れがある。介入効果を周期変動から分離するための手段としては、介入を頻繁に行う場合に、介入時期を定期的に設定せずに、全体系列の中に無作為に配置する方法がある。その方法としては、準実験においては「等価時間標本型(equivalent time samples design)⁽¹⁾」、単一事例実験においては「無作為配置型(randomization design)」と呼ばれるものがある。⁽¹⁹⁾

(五) その他の形式

(1) 逆介入対照群型(逆方向の介入を行う群を対照群とする形式)⁽¹⁰⁾

この形式は、事前測定を行い、介入群とは別の対照群を設定することによって、介入群における処置導入効果が事前測定や対照群と比較対照して確認する点で、標準的な非等価対照群型と変りがない。ただし、通常の対照群が何の介入もない集団であるのに対して、この型の対照群は、介入群とは反対の方向の介入が行われる集団である。次の図式における X^+ が一定方向での介入であるのに対して、 X^- はこれとは逆の方向での介入であることを表す。

$$\overline{O_1 \quad X^+ \quad O_2}$$

$$\overline{O_1 \quad X^- \quad O_2}$$

次の研究は、民主的・参加的な意思決定方式(X^+)と中央統制的な方式(X^-)が生産性と職務満足感に及ぼす効

果を分析したものである。^(二) 仮説は、民主的方式が生産性と満足度を増加させるのに対して、統制的方式はこれを減少させるということである。これを実証するために、組織を四部門に分け、その中の二部門に対して、それぞれ民主的方式と統制的方式を導入した。

なお、この研究は正規の実験ではなく、現代の観点からは準実験の一種であるといえる。理由は次の通りである。第一に、部門の選抜が無作為に行われたかどうか不明である。第二に、職務満足感(従属変数)については、事前測定の段階で両群の間に差がある傾向があった。第三に、意思決定(独立変数)の主体が誰であるかについての認識において、事前測定の段階で両群の従業員間に差がみられた。

結果は、おおむね仮説の通りであった。この研究に対しては、準実験的な観点から次のような長所と短所を指摘することができる。^(一)

① 選抜×成熟の問題　この研究には、通常の準実験で最もよく問題になる選抜×成熟の要因〔(二)を参照〕の影響を心配する必要がない。その理由は次のようである。第一に、この実験の被験者は自己の判断で特定の作業部門に所属したのではないし、各作業部門が部門独自の決定によって民主的方式や統制的方式を選択したというのではない。群の配置は無作為ではないが、準実験にありがちな被験者の自己選択によるものでもない。したがって、被験者の選抜に伴って生じる群間非等質性の問題は、通常の準実験よりも少ない。

第二に、この要因は、二群の内部的変化が同じ方向に向かって異なる率で生じることを意味するのが普通である。しかし、この実験は、互いに逆方向の介入効果を比較するように作られている。その結果は、民主的方式の群はよい方向へ、統制的方式の群は悪い方向へ変化したことを示している。選抜×成熟の要因によってこれを説明できるとすれば、両群が反対の方向に変化する傾向を内部にもつ集団である場合だけであるが、これはほとんどあり

えないことである。したがって、この研究のように逆方向の介入を比較する形式は、選抜×成熟の妨害要因については、通常は無介入対照群型よりも強力な方法である。⁽¹⁾

② 概念的妥当性の問題　概念的妥当性 (construct validity)、すなわち、それが何を測っているかという問題に関して、この研究のような逆介入対照群型は、通常は無介入対照群型よりも強力な方法である。もし、この研究が通常の準実験のように、民主的介入群と無介入対照群を比較する形式で行われ、民主群の得点が無介入群よりも向上したとする。しかし、その原因が民主的意思決定方式の導入にあるのか、それとも、重要な研究の対象に指定されたこと自体が被験者に影響を及ぼした(ホーン効果)ということであるのかが不明である。この研究のように、無介入群の代りに逆介入群を対照群にする方法では、介入をするという点では民主的介入群と同一の条件を設定しているから、ホーン効果が両群に均等に作用するからである。

③ ホーン効果ないしは「偽薬」効果の問題　この研究には、次のような問題がある。この中で、民主的方式と統制的方式の両群共に、事前よりも事後の生産性が高かったと報告されている。職務満足感とは異なって、生産性は仮説のような方向への変化を示していない。無介入対照群が設定されていないから、正確なことは不明であるとしても、前述のホーン効果が両群の生産性上昇の原因となった可能性がある。前述の概念的妥当性についての逆方向介入群型の利点は、両群の変化が反対の方向に現実に起るという前提の下に成り立つものである。しかし、生産性では両群が同じ方向に変化した。この場合の逆介入対照群型は、事前事後測定付き単一群型(1)の(3)を参照)と同様に、対照群を欠く方法にとどまる。すなわち、変化は確かに起ったが、しかしそれは民主的方式とは別のいかなる方式でも起る可能性があるということである。この実験に必要なことは、民主的方式と統制的方式による介入を行う二群の他に、外見的には介入しているが実質的には以前と変わらない偽薬的対照群(Placebo control)

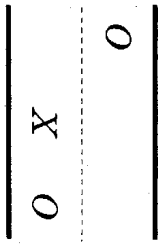
と、従業員に対する対応が事前も事後も変わらない無介入対照群を設定することである。

(2) コホート型^(二二)

コホート (Cohort) は、学校のような公的な組織や、家族のような私的な集団で一緒に生活する集団を指すものとする。^(二一) コホート型は、介入群の被験者と同じコホートに属する者を対照群の被験者とすることによって、介入効果を評価する方法である。

この例として、次の研究がある。^(二三) テレビ番組「セサミ・ストリート」が幼稚園児の知的発達に及ぼす効果を知るために、番組を見た幼児に対して年度の終りに発達検査を行い、これを介入群とした。しかし、同じ時期に番組を見なかった幼児、すなわち対照群の検査得点が欠けていた。

この幼稚園は、出身階層において著しく異質な集団構成であったから、仮に形式的にも対照群を設定できたとしても、これが介入群と等質の集団であるという保証はなかった。ところが幸いにして、介入群の幼児の兄や姉がこの幼稚園で同じ年齢の時に同じ発達検査を受けており、その当時はこの番組は始まっていなかった。そこで、兄や姉を無介入対照群として、テレビ番組の有効性を評価することになった。これを図式化すれば、次のようになる。点線の上段の○は同年齢時の兄や姉の得点、下段の○は番組を見た幼児の得点である。



この型は、兄や姉を非等価群とみなした時には、事後測定のみ非等価群型「(一)の(2)を参照」に似た方法である。番組視聴の有無ということ以外の点では幼児とその兄や姉が等質であるとみなした時には、事前事後測定付き

単一群型〔(一)の(3)を参照〕に似た方法である。コホート型の利点は、被験者が独立変数（介入の有無）以外の多くの背景変数において共通の特徴をもっているために、選抜の妨害要因を条件統制することが容易になる点にある。この手順は、ある小学校である年度の四年生に対して実施した新しい教育法の有効性を前年度の四年生と比較して評価する方式よりも、選抜の面では整備されている。群の構成のしかたによっては、正規の無作為配置型実験に匹敵する力をもつ方法となる可能性がある。^(二二)

(3) 手順の不備が避けがたい場合

アメリカの連邦交通安全本部は各州に対して、オートバイ利用者にヘルメット着用を義務づける法律を制定するように要請した。しかし、法律制定の状況は州によって異なっていた。ある時点をとれば、法律を制定した州もあれば、制定していない州もあつた。^(二四)そこで、法律を制定した八州を介入群として、法律を制定していない州の中で介入群と類似の特徴をもつ七州を対照群とすることによって、ヘルメット着用が交通安全に及ぼす効果を評価する調査が行われた。ここで、法律の有無が独立変数、オートバイによる死亡事故数が従属変数である。

結果は次の通りである。介入群におけるオートバイ登録数一万台当りの死亡事故数は、法律施行の前年には一〇件強であつたが、施行後は約七件に減少した。これに対して、対照群の死亡事故は、介入群における法律施行前後の期間を通じて約一〇件にとどまっていた。

なお、法律の有無とは別に、実際のヘルメット着用率を都市の過密道路で調査すると、次の結果が得られた。法律のある二都市では事実上すべての人がヘルメットを着用していた。これに対して、法律のない二都市のうちで、一つの都市では六〇%強の人がヘルメットを着用していたが、もう一つの都市でヘルメットを着用していたのは二五%にすぎなかつた。^(二五)

法律がない場合にもヘルメット着用を市民に勧める州があったし、法律があるとしてもこれを厳格に執行しない州があることも考えられる。ヘルメット着用を義務づける法律の有無と、実際のヘルメット着用は同じではない。この研究は標準的な非等価対照群型の調査であるが、ヘルメット着用の有無という独立変数の設定において一貫性が欠けていたと考えられよう。

- (一) Cook & Campbell (1979) ・前掲〔序論の注(一)〕。
- (二) The one-group posttest-only design.
- (三) The posttest-only design with nonequivalent group.
- (四) The one-group pretest-posttest design.
- (五) McNemar, Q. A critical examination of the University of Iowa studies of environmental influences upon the I.Q. *Psychological Bulletin*, 37, 63-92, 1940. [Cook & Campbell (1977) 参照]。
- (六) The untreated control group design with pretest and posttest.
- (七) Weber, Cook, & Campbell (1971) 「学校統合が公立学校生徒の学力自己評価に及ぼす効果」についてのアメリカ中西部心理学会における発表 [Cook & Campbell (1979) 参照]。
- (八) Cook, et al. (1975) 「ヤニック・ストーリー再訪——評価研究における事例分析」(ラッセル・セージ財団) [Cook & Campbell (1979) 参照]。
- (九) The untreated control group design with proxy pretest measure.
- (一〇) The untreated control group design with separate pretest and posttest samples.
- (一一) Kerpelman, et al. (1978) 「戸外の原動機に関する広報活動と教育計画の有効性評価」(合衆国消費者安全委員会に対する報告書) [Cook & Campbell (1979) 参照]。
- (一二) The untreated control group design with pretest measures at more than one time interval.
- (一三) The nonequivalent dependent variables design.
- (一四) Broadbent, D.E. & Little, F.A. Effects of noise reduction in a work situation. *Occupational Psychology*,

- 34, 133-140, 1960. [Cook & Campbell (1979) 参照]°
- (115) The removed-treatment design with pretest and posttest.
- (116) Lieberman, S. The effects of changes in roles on the attitudes of role occupants. *Human Relations*, 9, 385-402, 1956. [Cook & Campbell (1979) 参照]°
- (117) The repeated-treatment design.
- (118) Roethlisberger & Dickinso (1939) 「経営管理の労働性」 [Cook & Campbell (1979) 参照]°
- (119) Edginton, E.S. Statistics and single case analysis. In Hersen, M., Mislcr, R.M., & Monti, P.M. (Eds.) *Progress in behavior modification*, Vol. 16, pp. 83-119. Academic, 1984.
- (120) The reversed-treatment nonequivalent control group design with pretest and posttest.
- (121) Morse, N.C. & Reimer, E. The experimental change of a major organizational variable. *Journal of Abnormal and Social Psychology*, 51, 120-129, 1956. [Cook & Campbell (1979) 参照]°
- (122) Cohort designs in formal and informal institutions with cyclical turnover.
- (123) Minton, J.H. The impact of Sesame Street on reading readiness of kindergarten children. *Sociology of Education*, 48, 141-151, 1975. [Cook & Campbell (1979) 参照]°
- (124) 法律を制定したのは、一九六七年以前には三州、一九六七、六九年に三七州であり、この調査の時点ではカリフォルニア州とイリノイ州を除くすべての州であった。ただし、ユタ州では時速三五マイル以上の場合にのみヘルメットを着用しなければならない。
- (125) Posavac, E.J. & Carey, R.G. *Program evaluation: Methods and case studies*. Prentice-Hall, 1985.

三 中断時系列型

時系列 (time-series) とは、時間の経過と順序に合せて観察される事象の組をいう⁽¹⁾。観察は、特定個人に対する反復的な観察のように同一の対象に対して行われる場合もあるし、ある小学校の三年生の体格についての定期的検査のように、同一ではないが類似の対象に対して行われる場合もある⁽²⁾。

中断時系列 (interrupted time-series) とは、一定の対象に関する連続的な時系列データが、ある時期にその対象に対して介入を行うことによって、介入前と介入後に質的に中断 (interrupt) されていることを表す。中断時系列型準実験の目的は、介入が行われた後の測定値を介入前の状態と比較することによって、介入が対象に対して及ぼす効果を明らかにすることである。ここで、介入前の測定値は、介入効果を評価する際の比較対照基準の役割をもつという点で、無作為配置型の正規の実験や非等価対照群型の準実験における対照群と同様の機能を備えている。ただし、無作為配置型や非等価対照群型の対照群は介入群とは別の集団であるが、中断時系列型において対照基準の役割をもつ介入前のデータは、介入後のデータと同一の集団に属している。したがって、中断時系列型の基本的特徴は、測定が単発的ではなく連続的であることと、対象が二群の構成ではなく基本的には単一の群で構成されることにあるといえよう。

(一) 単純中断時系列⁽³⁾

これは、中断時系列型の最も基本的な形式である。次の図のように、同じ対象に対して一〇回の測定が行われる

とすれば、前半の五回が介入以前の測定であり、後半の五回が介入開始後の測定である。介入後の測定値の変化は、介入前の水準および傾向と比べることによって評価される。

$O_1 \quad O_2 \quad O_3 \quad O_4 \quad O_5 \quad X \quad O_6 \quad O_7 \quad O_8 \quad O_9 \quad O_{10}$

(1) 先駆的な適用例

① 速度違反に対する厳罰が交通事故の低減に及ぼす効果〔一の(一)の(3)を参照〕 一九五五年にアメリカの州で交通事故による死亡が急増した。知事の指示によって速度違反の取締りを強化すると、その翌年には事故死は前年より一四%減少した。そこで知事は、事故減少の原因が緊急対策にあると考えた。知事の発想は、一九五五年と五六年だけの単発的な事前事後測定の前実験的形式〔二の(一)の(3)を参照〕に基づき考察であるが、図3-1のように介入前後の状態を数年間にわたって観察するならば、中断時系列型のデータとして検討することが可能となる。^(四)

② 労働時間短縮が生産性に及ぼす効果 英国産業疲労委員会が一九二〇年代に実施した古典的な生産性調査は、現代の中断時系列型の先駆的な例である。^(五)これによれば、一九一九年八月から、一日の労働時間が一〇時間から八時間に短縮されたが、その結果として、単位時間当りの生産量が増加したといわれる。結果は、図3-2に示した。たしかに労働時間短縮前の平均生産量と短縮後の生産量を比較すれば、短縮後の方が短縮前よりも増加している。しかし、現代の準実験的な観点からデータを解釈すれば、この結論には次のような問題点がある。^(三)

第一は、成熟の要因である。生産性については、事前事後の平均値の比較によってだけでなく、時間系列上の傾向性という点にも着目しなければならない。生産性の向上は、時間短縮が実施される以前から持続的に生じており、時間短縮後の増加はこの延長線上にある。したがって、時間短縮後の生産性向上は、労働時間短縮が原因なの

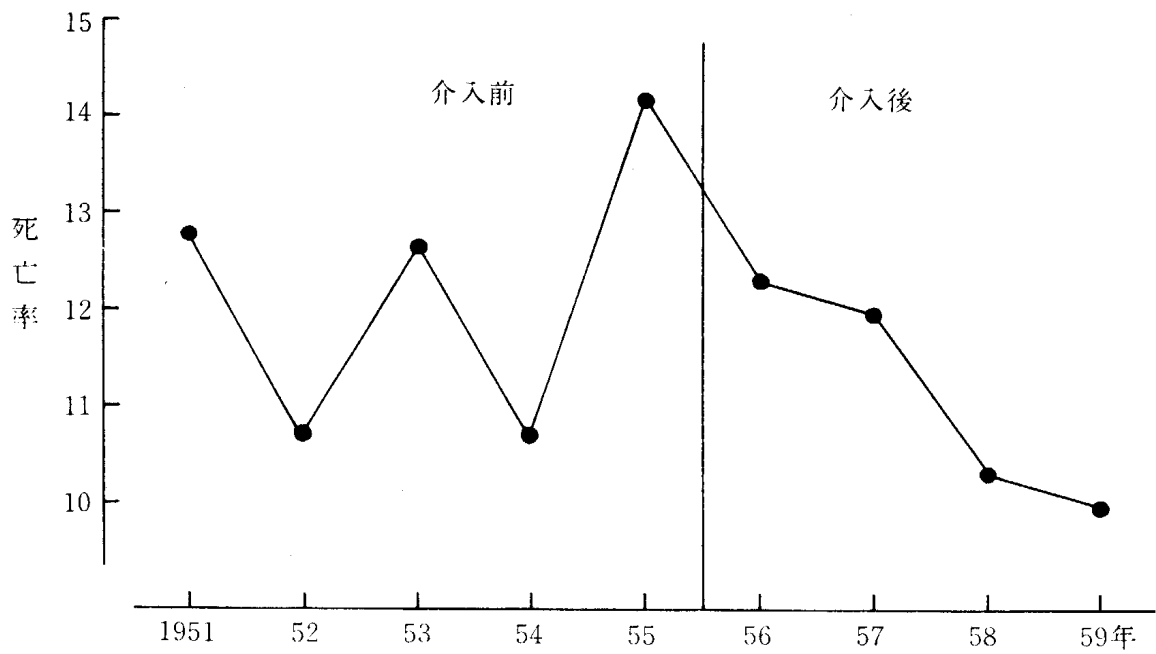


図3-1 交通事故の死亡率 [Campbell (1969) による]

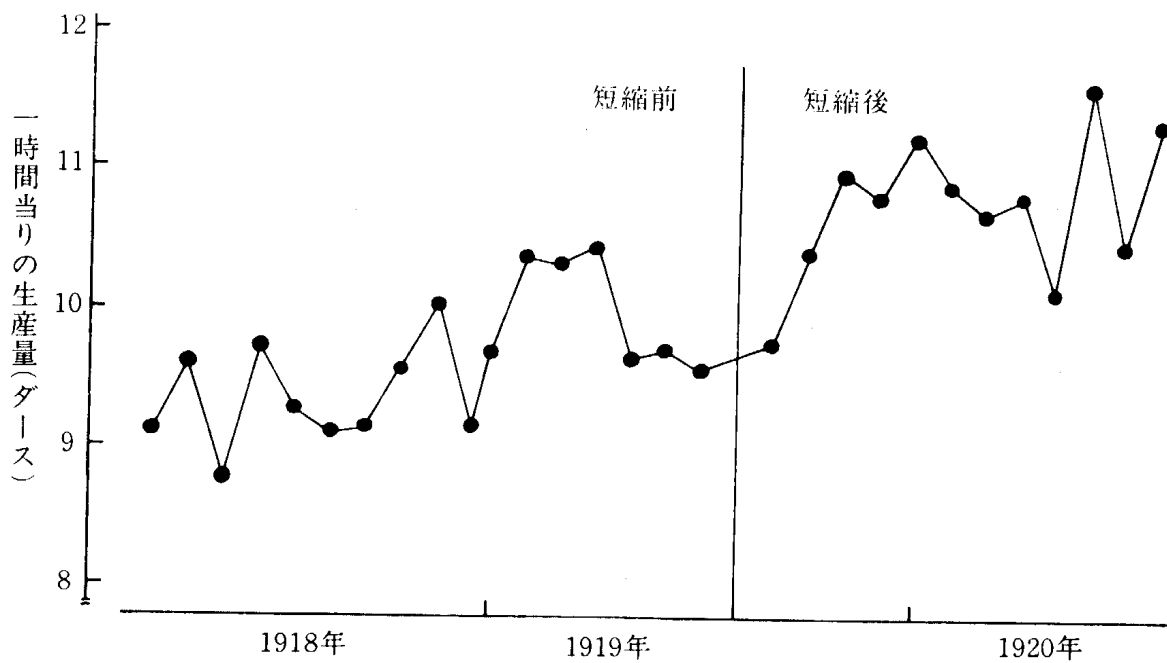


図3-2 労働時間短縮による生産性向上 [Farmer (1924) による]

ではなく、内部に以前からあった何らかの成熟要因によって生じた可能性がある。

第二は、周期変動 (seasonal trend) である。一九一八年と一九年の両年において、夏になると生産性が低下する傾向があった。労働時間の短縮が実施されたのは夏であるが、時間短縮がない場合にも、この季節が過ぎると、生産性は回復した可能性がある。ただし、短縮後の一九二〇年には夏のスランプは起っていないから、この可能性も確かなものではない。いずれにせよ、周期変動が介入効果を不明瞭にしているのは確かである。この点を明確にするためには、より長い時系列が必要である。

(2) 参加的運営が夜警の欠勤率低減に及ぼす効果^(木)

夜警の欠勤を減少させるために、管理運営に従業員を参加させる方を新しく導入した。新方式の有効性を評価するために、夜警従事者を三群に分け、この中の一群に参加的運営方式を適用した。従属変数は勤務率であり、規定勤務時間に対しての実際に勤務した時間の百分率で表される。結果は、図3-3に示した。この研究に基づいて、中斷時系列型のデータに対する解釈の主要な観点について述べる。

① 回帰の問題 介入直前の週における勤務率は七〇%近くまで

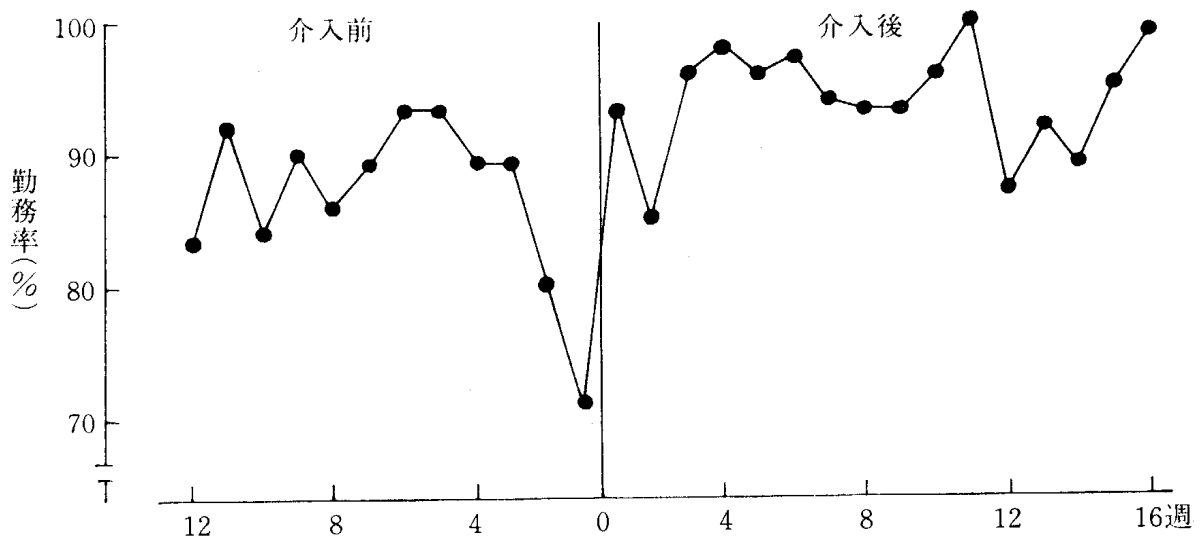


図3-3 参加的運営が勤務率に及ぼす効果 [Lawler & Hackman(1969)による]

低下していたが、介入後の勤務率は九〇%強の水準まで急上昇している。ただし、介入前の全体の平均水準は九〇%弱であり、直前の週の七〇%はむしろ例外的であった。したがって、介入後の勤務率上昇は、新方式の導入が原因となつて起つたということの他に、事前の平均水準への回帰にすぎないと考えることもできる。

反復的に測定が行われる中断時系列型は、事前の水準と傾向を比較的容易に知ることができるから、回帰や成熟の要因の影響を点検するのに適している。この点において、中断時系列型は、単発的に測定が行われる非等価対照群型よりも整備された方法である。

② 歴史の要因 中断時系列型は、実験期間中に生じうる別の要因すなわち歴史の要因の影響を除外する手順を欠いている。この点において、無介入対照群の設定によって歴史の要因を点検することができる非等価対照群型より不完全な方法である。ただし、ここで言及した研究については、測定が週単位で行われており、前項(1)のような年単位や月単位ではないから、その週に何があつたかを知ることが困難ではない。この点で歴史要因の点検は、次善の形ではあるが、可能であるといえよう。

③ 記録条件の要因 非等価対照群型は、比較的短期間に実施することができる。これに対して、中断時系列型は、場合によっては数十年にわたる連続的データを対象にする方法である。この間に対象を記録する条件が全く変化しないということはむしろ珍しい。記録条件は、中断時系列型における主要な妨害要因の一つである。

その例として、次のような研究がある。^(七) ドイツで一九〇〇年に離婚に関する法律が改正された。法律改正が国民生活に及ぼした影響を調べるために、一八八一年から一九一四年までの三四年間にわたつて、人口一〇万人当りの離婚件数を一年ごとに計算したところ、法律改正後二年間は離婚率が以前より低下していた。問題は、それが統計的に有意な変化であるかどうかということだけでなく、^(八) 離婚がもつ意味 (definition of constructs) が法律の

改正によつて変えられたということにある。^(三、四)この場合には、改正前と改正後の離婚率は異質な係数となっている恐れがあるから、両者の比較は困難であろう。

次は、芝刈機の安全使用運動の有効性を評価した研究である〔二の(三)の(2)を参照〕。この中で、病院に記録された件数を従属変数の一つとして使う試みがあった。その病院では、以前から事故の原因が記録されていたからである。調査者は、芝刈機による事故件数がこの運動によつて減少することを予想していたが、件数は逆に増加して、介入開始から介入終了までの一年間の件数は、介入前二年間の件数を上回るほどであった。その原因は、病院で事故を記録する担当者がこの安全運動を知っており、自分がこれに関わっていることに過敏になったことにある。そこで、病院に報告された三年間の事故を再分析すると、この間に有意な変化はなく、安全運動が事故を増加させていないことが示された。^(二)

④ 選抜の問題　非等価対照群型では、対照群と介入群の被験者が等質ではないことは常に大問題である〔二の(二)を参照〕。中断時系列型は同一の集団に対して適用される方法であるから、基本的には問題がない。しかし、この型でも介入が行われる前(対照期間)とその後の介入期では、被験者の構成が変ることがある。たとえば、ある集団に対して実験的介入が実施されることになった時に、一定の被験者が実験対象になることに同意せずに実験から脱落したとする。脱落者は、集団全体から無作為に抽出した標本ではなく、一定の特徴を共有する人達である。この場合には、介入前と介入後の被験者は非等質な集団となる。これを是正する方策としては、被験者の性や年齢などの背景変数を調べて、被験者の構成が介入前後にどの程度変化したかを知るという手順がある。また、介入開始時に脱落した被験者を、介入前のデータからも削除するという方法もありえよう。

⑤ 周期変動の問題　時系列データは、回帰のような確率変動だけでなく、季節のような周期変動からも影響

を受ける。前項(1)に述べた労働生産性にも季節変動の傾向がみられた。本項の夜警欠勤率においても季節変動の可能性がないということはできない。つまり、この実験が開始されたのは一二月の寒い季節であり、参加的運営という処置が導入されたのはその一二週間後の三月から四月にかけての快適な季節であった。もし、介入を始めたのが春ではなく冬である場合に、結果がどうかは不明である。なお、周期変動の要因をいかに処理するかについては、統計学的な時系列分析で種々検討されている。^(九)

(3) 刑務所内運営方針の変更が服役者の行動に及ぼす効果^(一〇)

① 問題 一九六九年七月一日にアメリカ・テネシー州において、都市部の刑務所で行動上の問題をもつ服役者を、山間部の刑務所に移すという新しい方針が定められ、七月二八日に実施された。行動上の問題は厳密には定義されていないが、違反をくりかえすことや、傷害を与える攻撃などがこれに属する。関係者の認識によれば、山間部の刑務所は都市部と比べて魅力がなく、山間部に移されることは、服役者に対して罰として作用すると予想されていた。

新方針の導入は、刑務所全体に一齐に行われた。違反行動は、一九六九年一月から七一年六月までの二年半の間、月ごとに記録されていた。この調査は、単純中断時系列型の準実験として分析することができる。問題は、一九六九年七月の新方針導入の前後に服役者の行動が変化したのか、変化があった場合にその原因が新方針の導入にあるのか、ということである。

② 方法 調査対象は、実験期間中(一九六九年一月〜七一年六月)に都市部の刑務所に服役した全員である。その人数は月平均一九〇六人であって、すべて男性である。施設の条件については、都市部の刑務所が監視・医療・その他の安全・生活の面で中程度の条件をもつものに対して、山間部の方は、遠隔地にあること、面会者が少

ないこと、職業訓練の機会が不足していることなどの不利な条件をもっていた。

従属変数は次の通りである。二年半の調査期間中における服役者全員のすべての違反行動が記録された。これに基づいて、全服役者を千人として、一日に起った違反行動の月別平均件数を計算した。図3-4に、これを示した。

③ 結果 服役者千人当りの一日平均違反件数は、実施前には約二件であったのに対して、実施直後の半年間は明らかに減少して約一件になったが、その後は再び増加している。つぎに、違反者の人数を、違反をくりかえす回数別(一回・二回・三回以上)に集計すると、実施直後の半年間はどの場合も減少したが、それ以降については、一回だけの違反者が実施前(半年間で三〇〇人弱)よりも多くなっている(四〇〇人強)。

④ 解釈 以上の結果を準実験の観点から考察すれば、次のようになる。

① 歴史 新方針の導入時に、新方針以外の条件に変化はなかったか。実施前に違反行動に対して与えていた罰は、

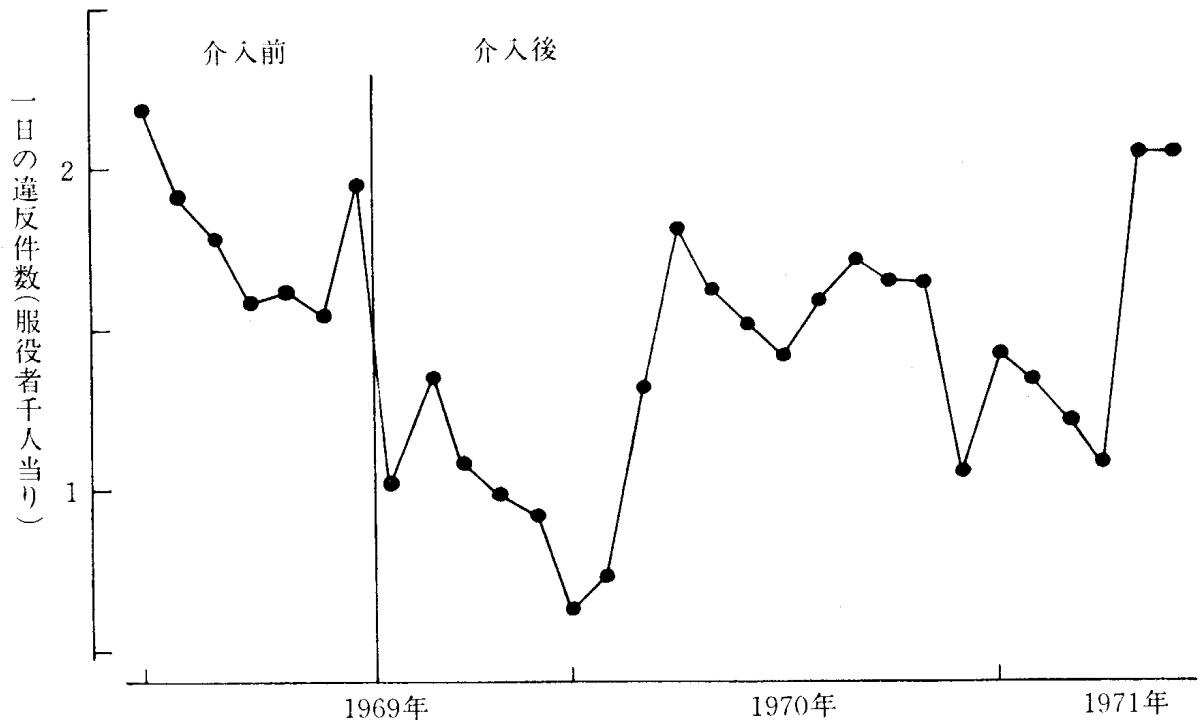


図3-4 刑務所の方針変更が違反行動に及ぼす効果 [Schnelle & Lee(1974)による]

味つけをしない食事を与えると同時に独房に入れること、自由時間を削減することであった。新方針を導入した後には、独房と自由時間の罰は従来通りに継続したが、味のない食事は中止している。この食事の中止の直接的効果を調べるができない場合には、それが従来もっていた効果についてのデータがあるとよい。しかし総じて、新方針導入の他に条件の大きな変化はなかったと考えられる。

④成熟　実施後半年間における違反の減少の原因が、実施前にあった内部的傾向にあるとはいえない。なぜなら、半年間の事前データに限ってみれば、違反の持続的な減少傾向はなかったからである。ただし、それよりさらに前の傾向は、データがないので不明である。

⑤記録条件　これが、本調査における主要な妨害要因である。違反の認定は、州矯正省の手引に従って行われた。しかし、これに看守がどの程度忠実であったか、所内の懲戒委員会の決定に一貫性があったか、ということが問題になる。第一に、違反をくりかえす者が山間部に移されることが多かったのであるが、違反がないにもかかわらず、山間部に移された者が全移送件数の八%を占めていた。第二に、違反の種類と罰の種類に対応に一貫性がなかった。たとえば、三ヶ月間に起った一〇一件の違反の中に六件の麻薬所持があったが、これに対する罰は一種類ではなかった。第三に、違反の種類によって、認定基準が明確でない場合がある。たとえば、格闘・同性愛・窃盗の認定基準は明確である。一方、看守に対する不服従は、発生頻度が最も高い違反であるが、その認定を客観的に行うのは難しい。観察の信頼性検定は、この研究では実施できない。これに加えて、違反をみつけて記録することが同時に服役者に影響を及ぼす (obtrusive) という検査効果の問題〔一の(二)の(4)を参照〕もある。

⑥回帰　実施直後の減少が回帰によって生じたとは考えられない。たしかに、実施直前の七月は介入前で二番目に違反が多い月であったから、その後は本来の水準に回帰して減少したという可能性はないとはいえない。しか

し、事前の傾向が不安定型ではないことと、介入時に起った変化が全期間中で最大の変化であったことを考慮すれば、回帰が違反の減少であるとはいえない。

⊕選抜 方針の実施前には、違反行動をとった服役者は都市部の刑務所にとどまっていた。これに対して、実施後になると違反者は山間部に移される可能性がある。服役者全員を対象にして、実施前と実施後の単位人数当りの違反件数を比較すれば、再び違反を犯す恐れのある者が山間部に移されている実施後の方が違反が少ないのは当然である。つまり、選抜の要因によって事前と事後の集団が非等質になる。これを防ぐために、実施後に山間部に移された服役者は、実施前のデータからも削除した。また、新しく入所する犯罪者の種類は、調査期間を通じて変化がなかった。服役者の種類が季節的に変動することもなかった。したがって、選抜の要因は結果に影響を及ぼしていないと考えられる。⁽¹⁰⁾

(二) 非等価対照群時系列付き中断時系列

中断時系列型は、成熟の要因を点検できる点でよく整備された方法であるが、歴史の要因を点検する点では非等価対照群型の方が勝っている〔(二)の(2)を参照〕。次の図に示したのは、非等価無介入対照群時系列付き中断時系列といわれる方法である。⁽¹¹⁾これは、単純中断時系列データに、介入を行わない別の集団の時系列データを並行させた方法であるから、中断時系列型と非等価対照群型の結合形式であるといえよう。

O_1	O_2	O_3	O_4	O_5	X	O_6	O_7	O_8	O_9	O_{10}
O_1	O_2	O_3	O_4	O_5	O_6	O_7	O_8	O_9	O_{10}	

(1) 交通違反の処分が事故低減に及ぼす効果「一」の(一)の(3)、三の(一)の(1)を参照」

前節(一)における処分の有効性評価は、当該の州単独のデータに基づいて行われた。これに対して、図3-5に示したように、当該の州の時系列データに加えて、この州に隣接した地域で、新方針が導入されていない四州の年間交通事故死数の平均値を並行させると、この型の準実験が作られる。当該の州を介入群、隣接の無介入の州を非等価対照群として、対照群との比較によって介入群の変化評が価される。非等価対照群型と異なる点は、一九五五年末の介入開始の前後の単発的測定ではなく、一九五一年から五九年までの九年間の連続的測定が行われることである。

この形式に基づいて介入効果を評価すれば、次のようになる。第一に、九年間の死者は両群共にわずかな減少傾向を示しており、介入群における一九五五年以降の減少は全体的傾向の一部にすぎない可能性がある。しかし、対照群の減少傾向と比べれば、介入群の減少率の方が明らかに大きく、この差を説明できるのは処分導入の有無以外にはないと考えるこ

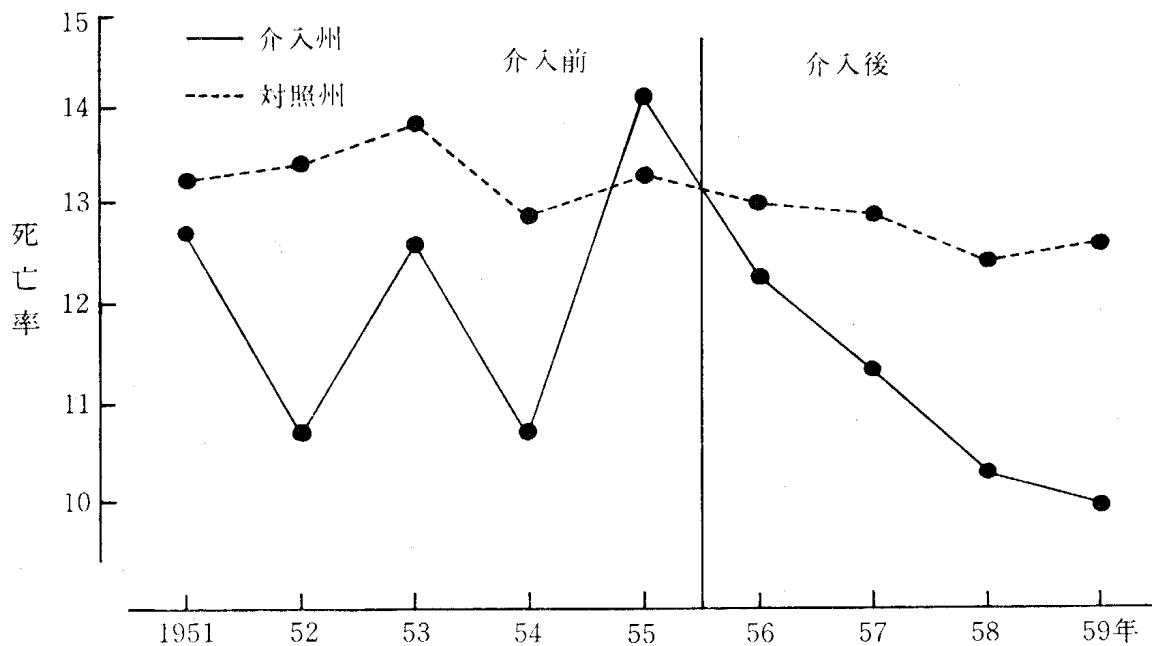


図3-5 交通事故の死亡率 [Campbell (1969) による]

ともできる。第二に、介入群の事前のデータは変動の大きい不安定型であるから、介入直後の減少は回帰によって生じた可能性がある。しかし、傾向性が事前には不安定型であったが、事後には減少型へと変化している。処分の導入がデータの傾向性を変化させたと考えられることもできる。しかしながら、同様の傾向性の変化は対照群にもわずかにあることを考慮に入れると、処分導入とは別の歴史要因が両群に共通に作用した可能性もある。

なお、対照群の値は隣接四州を合成したデータであるが、介入の州に最もよく似た一州をこの中から選び、合成的な対照群ではなく自然の非等価対照群を設定する方法もあったのではないかと思われる。これによるならば、歴史や成熟の要因の点検が容易になるからである。

(2) 参加的運営が欠勤率低減に及ぼす効果〔(一)の(2)を参照〕

図3-6は、前節(一)の図3-3に示した介入群単独のデータに、対照群の時系列データを並行させたものである。この対照群は、介入群と同じくボーナスは支給されるが、管理運営上の意思決定には参加できない集団である。対照群の勤務率が全期間を通じて八〇％の水準を中心に変動しているのに対して、介入群の平均水準は、事

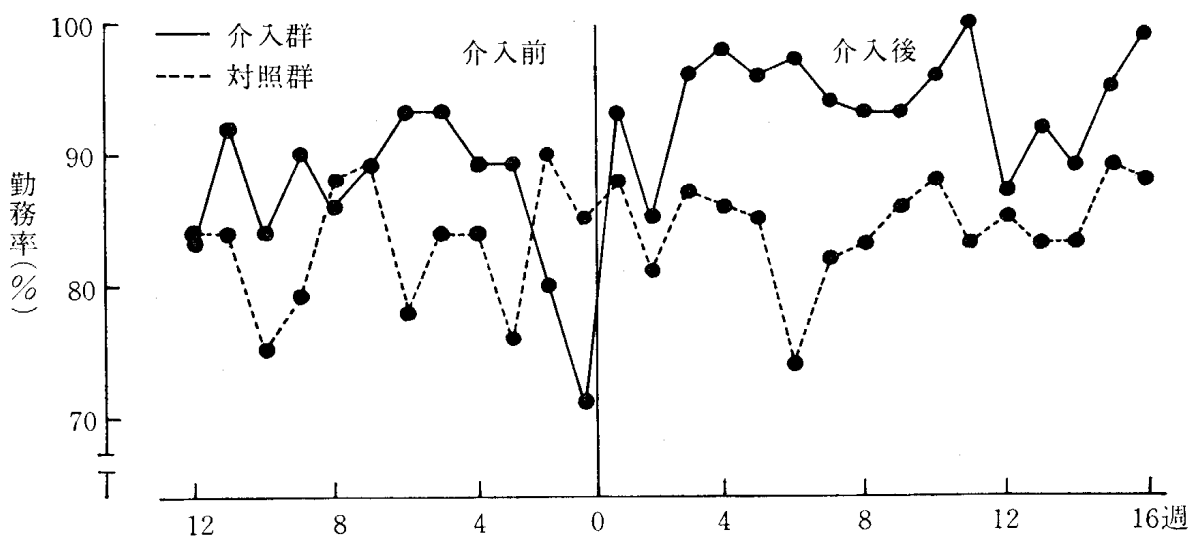


図3-6 参加的運営が勤務率に及ぼす効果〔Lawler & Hackman(1969)による〕

前から事後に向かつて上昇して九〇％台に達している。これは、介入が有効であったことを示唆している。

対照群と介入群の測定は同時に行われたので、歴史の要因は統制されている。選抜×歴史の相互作用（地方史の要因）は、両群が非常に異質な場合にはありうるが、この調査のように等質度が高い場合には影響は少ないであろう。欠勤の記録法は両群に共通である。記録法は従業員の行動に影響を及ぼさない性質（unobtrusive）のものであって、刑務所で服役者の違反行動を記録する場合〔一〕の〔三〕を参照〕とは異なる。仮に記録が従業員の行動に影響がある（obtrusive）としても、それは両群に同程度に及ぶ。季節による周期変動の要因は、両群に共通に作用している。

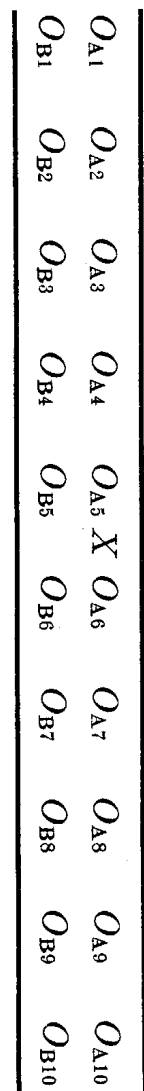
このように介入の有無以外のすべての条件を介入群と均等にした非等価対照群時系列を設定することによって、妨害要因の影響を除去することが、単純中断時系列型よりもはるかに容易になるのである。また、反復的測定を行う時系列型は、群ごとに差別的に生じる回帰の要因の影響〔二〕の〔一〕の〔四〕を参照〕を点検できるといって、単発的測定を行う非等価対照群型よりも整備された方法である。

（三） 非等価従属変数付き中断時系列^{(二)(三)}

中断時系列型の主要な問題の一つは、歴史の要因にいかに対処するかである。前節〔二〕における非等価対照群時系列は、このために設定されている。しかし、何らかの理由で非等価群を設定できない場合は、介入によって影響を受けることが非常に少ない変数を調査対象内部から採り出して、対照群に準じる機能をこの変数に与えるという手順をとることができる。つまり、同一集団の一定側面（従属変数）に対する介入の効果は、無介入の他の側面の状態と比較対照することによって評価される〔二〕の〔四〕の〔一〕を参照〕。この方法は、単一の集団に対して適用される点

で単純中絶時系列型〔一〕を参照〕と同じであり、実質的に対照群に近い機能を備える点で非等価対照群時系列付き中絶時系列型〔二〕を参照〕と似ている。

次の図において、Aは介入が行われる部分、Bは同じ群内で無介入の部分であることを表す。



(1) 飲酒検知器の導入が交通事故低減に及ぼす効果^(一三)

イギリスで一九六七年に道路安全法が制定された。これに伴って、飲酒運転の取締りが強化されることになり、飲酒検知器 (breathalyser) が導入された。研究の目的は、検知器の導入が交通事故による死亡と重傷に及ぼす効果を知ることであった。しかし、法律は全国一斉に施行されたので、検知器を導入しない対照群のデータがなかった。これを補うために、次のような手順をとることになった。

イギリス人が酒を飲むのは家庭よりも酒場であることが多いので、^(一四) 飲酒運転の取締りが強化されると、酒場の開店時間帯やその中でも特に週末の夜に、事故数が減少するであろう。これに対して、酒場が閉店している時間帯の交通事故に対しては、飲酒検知器の影響は少ないであろうと予想される。

そこで、酒場営業中の週末の夜の交通事故数を介入時系列、閉店中の事故数を対照時系列とすることによって、飲酒検知器導入の効果を評価するという手順が採用された。図3-7は、週末夜の事故数、閉店中の事故数、および全時間帯の事故数を、法律制定の前後五年間にわたって月ごとに記したものである。

法律が施行された直後には、閉店中と全時間帯の事故数がわずかに減少したにすぎないのに対して、週末の事故

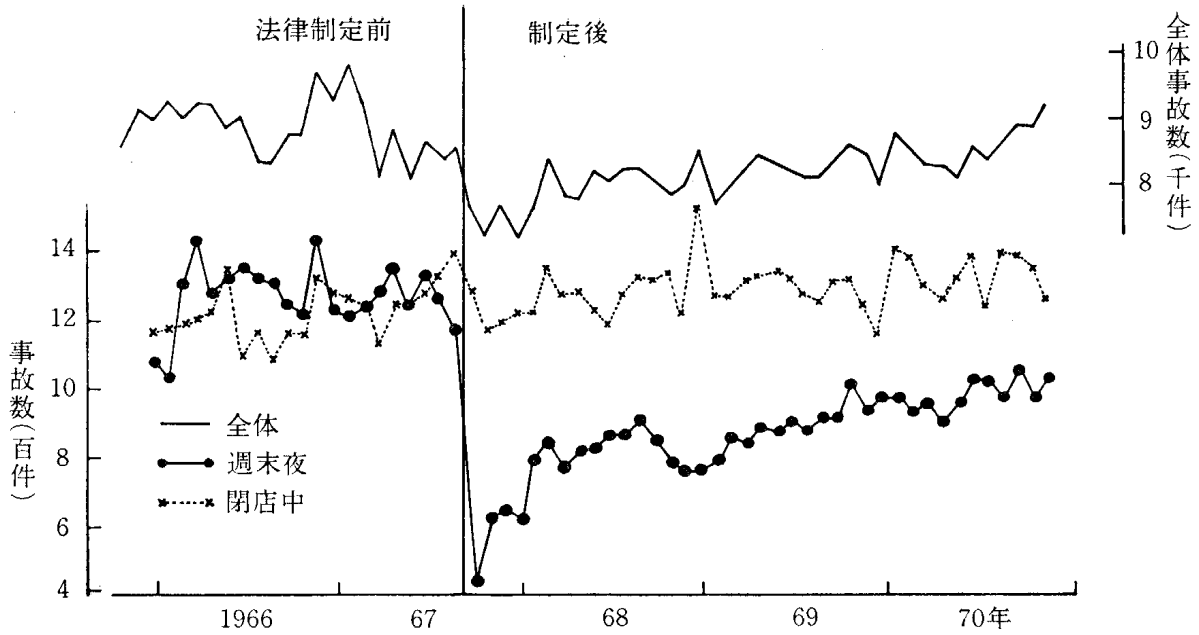


図3-7 飲酒運転に関する法律が交通事故に及ぼす効果 [Ross, et al. (1970)による]

は激減した。しかし、これによって、新方式が事故低減に直接的効果をもったとは必ずしもいえない。その理由は次の通りである。第一に、新方式が予期せざる副次効果（たとえば、災害保険の率の変更など）をもたらし、これが間接的に事故発生に影響を及ぼした可能性がある。第二に、週末の事故は法律施行直後に激減したが、その後は徐々に増加する傾向がある。

(2) 結果の再分析

飲酒運転に関する法律の施行が市民の運転行動に対して具体的にいかなる影響を及ぼしたかについての再分析が行われている^(二四)。分析の着眼点は次のようである。法律施行によって事故はどのように減少したのか。事故数の減少は、全体的な運転量の減少によって生じたのか、全体的な飲酒量の減少によるのか、それとも飲酒運転者が以前より注意深く運転するようになったためなのか。

第一に、英国道路研究所の走行距離調査に基づいて単位走行距離当りの事故発生件数を計算すると、これも法律施行後に減少していた。したがって、事故の減少は走行量が減少した結果として生じたというよりは、事故そのものが減少したと考える

ことができる。第二に、法律施行前後の酒の販売量を調査すると、この間に販売量の変化はみとめられなかった。したがって、飲酒量が減少した結果として事故が減少したとはいえない。第三に、法律施行前後一〇ヶ月間の調査では、飲酒後は歩いて帰宅する人が増加したことがわかった。第四に、事故死の遺体の血中アルコール濃度は、法律施行後に低下しており、深酒が少なくなったことを間接的に示している。第五は、法律施行直後に激減した事故が、なぜその後は増加傾向を示したのかである。これは、裁判所が飲酒運転に対する処罰を緩めていったためであると説明されている。^(一四)

なお、イギリスと同様の法律を制定した北欧諸国についても、この調査が行われた。しかし、厳格な法律の制定によっても、飲酒運転が減少することはなかった。^(一五)

(四) 撤去手順をもつ中断時系列

(1) 処置の撤去を行う形式^(一六)

この形式は、参加的運営が勤務率に及ぼす効果に関する調査〔(一)の(2)、(二)の(2)を参照〕の追跡研究で採用されている。^(一七) 次の図で、 X は処置の導入を、 \bar{X} は処置の撤去すなわち継続していた介入を中止することを示す。

$O_1 \quad O_2 \quad O_3 \quad O_4 \quad O_5 X O_6 \quad O_7 \quad O_8 \quad O_9 \bar{X} O_{10} \quad O_{11} \quad O_{12} \quad O_{13}$

これは、二つの単純中断時系列を組み合わせた形式である。第一は O_1 から O_9 までであり、これによって処置導入の効果が評価される。第二は O_6 から O_{13} までであり、これによって処置撤去の効果が評価される。もし、 O_5 から O_6 にかけて一定方向への変化があり、 O_9 から O_{10} にかけてこれとは逆の方向に変化があったとすれば、処置の有効性は二重に

確認される。この形式には、少なくとも単純時系列と同程度の意義があることは当然である。問題は、処置撤去の手順を設定することによって、いかなる解釈上の利点が生じるか、ということである。

まず、これによって歴史の要因が統制される。すなわち、処置導入時の一定方向への変化と処置撤去時の逆方向への変化を、歴史の要因によって説明できるのは、次の場合だけである。一つは、一つの歴史要因が処置導入時には一定方向に作用し、処置撤去時には逆方向に作用する場合である。もう一つは、二つの歴史要因のうち、一方が処置導入時に一定方向に作用し、他方が処置撤去時に逆方向に作用する場合である。しかし、いずれの場合も、それが生じる可能性は非常に小さい。

つぎに、選抜の要因が作用するのは、たとえば処置導入時には年少の被験者が脱落し、処置撤去時には女性の被験者が脱落するというように、対象が系統的に非等質になっていく場合だけであるが、これが起ることは非常に少ない。いずれにせよ、撤去型中断時系列において、単純中断時系列よりも多くの妨害要因が作用することはない。

むしろ、撤去型時系列の問題は、次のような点にある。第一は、実務領域の課題を解決するために導入した対策を、研究という目的だけのために中止する時に生じる道義的な問題である。第二は、処置の中止によって連続的な時系列が分断された結果、統計的な分析が困難になる場合があるということである。

(2) 処置の導入と撤去を反復する形式⁽¹⁾

次の図のように処置の導入 X と撤去 \bar{X} を反復することによって、従属変数が処置の導入に伴って一定方向へ変化し、処置の撤去に伴って逆方向へ変化することが反復的に示されたとすれば、処置が有効であったと評価される。

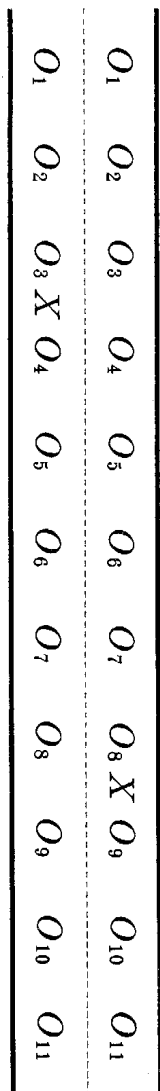
$O_1 \quad O_2 X O_3 \quad O_4 \bar{X} O_5 \quad O_6 X O_7 \quad O_8 \bar{X} O_9 \quad O_{10} X O_{11} \quad O_{12} \bar{X} O_{13} \quad O_{14}$

これは、処置の導入と撤去を周期的に行う方法であるが、これに基づいて変法を工夫することもできる。たとえば、導入と撤去を無作為に配置する方法も可能である〔二の(四)の(3)を参照〕。さらには、 X と X の代りに、 X_1 と X_2 という二種類の処置を設定すれば、一つの時系列の中で二つの処置の有効性評価を行うことが可能になる。場合によつては、 $X_1 X_2$ の相互作用効果を評価する方法を考へることができる。

しかし、このように短期間に複雑な変数操作を行う方法には、次のような欠点がある。第一に、変数操作が短期間に切り換えられるために、方法の適用は、効果の立上りと「立下り」の速い処置に限定される。第二に、これを実施する場合には高度の条件統制が必要であるから、閉鎖性の強い学校や刑務所では実施できることがあるとしても、フィールドでの実施は事実上不可能である。むしろ、この方法を最もよく使っているのは、精神衛生領域である。この中でも行動修正の研究者が最もよくこれを使うと指摘されている。その方法とは、単一事例実験の撤去型と相互作用型であると考えられる。撤去型では、処置の完全な導入と撤去が反復される。相互作用型では、処置の部分的な導入と撤去が反復される^(一九)。

(五) 介入時期を変えた非等価群中断時系列

これは、二つの非等価群を設定し、両群に対して同じ処置を異なる時期に導入する形式である。図式化すれば、次のようになる。



処置の導入時期は、上段の群が八回目の測定の後、下段の群が三回目の測定の後である。O₃までに期間を限定すると、上段は無介入群であり、下段は介入群であるから、標準的な対照時系列付き中断時系列型〔(二)を参照〕と同じである。つまり、上段を対照時系列として、下段に対する介入の有効性が評価されるのである。標準的な方法と異なるのは、対照群に対しても、時期は遅れるが介入を行うことである。

この時系列型準実験は、同じ処置を一方の群から他方の群へ切り換えて導入し、両群において同じ効果を再現するという意味で、スイッチ的再現型 (switching replications) の中断時系列と呼ばれている。⁽¹¹⁰⁾ なお、これは単一事例実験の多重基線型に酷似する形式である。⁽¹¹¹⁾

(1) テレビ放送の開始が住民の図書館利用に及ぼす効果

テレビ放送は、アメリカでは一九四〇年代後半に開始されていた。イリノイ州で人口増加率の高い富裕な都市地域では、この時期からテレビを視聴することができた。しかし、連邦コミュニケーション委員会は一九五一年に、テレビ局開設の免許証を発行することを停止した。この凍結が解除されたのは一九五三年であり、同州で人口が増加しない貧困な村落地域では、この年までテレビを視聴することができなかった。⁽¹¹²⁾

この時間差を利用して、テレビが読書に与える影響を調べた研究がある。⁽¹¹³⁾ テレビは読書の代用になると考えられるから、放送が開始されると図書館利用率が減少する、という仮説の下で調査が行われた。独立変数はテレビ放送の有無である。一九四八年順に放送が開始された五五の地域を初期介入群とし、一九五三年に放送が開始された五五の地域を後期介入群とした。従属変数は、住民一人当りの年間図書貸出部数である。データは、一九四五年から六〇年までの一六年間にわたるものである。

結果は、図3—8に示した。まず、放送開始の前後を比較すると、初期介入群では、放送が開始された一九四七

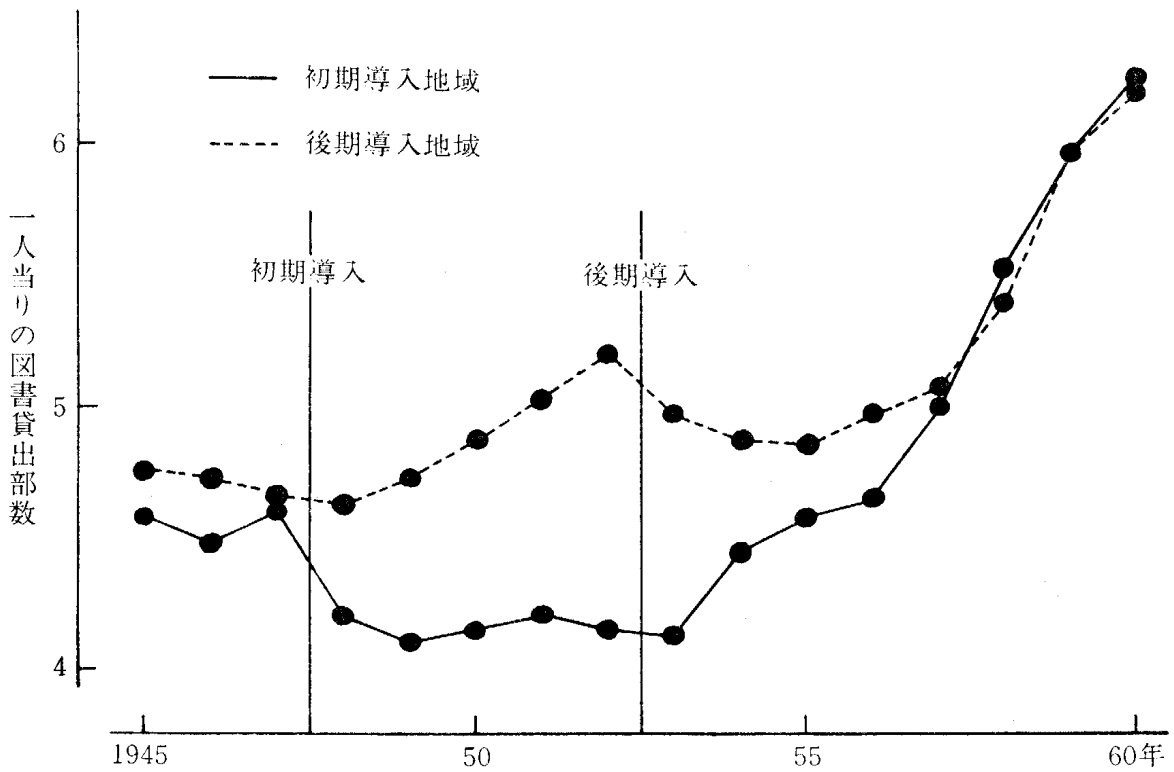


図3-8 テレビ放送の開始が図書館利用に及ぼす効果 [Parker, et al.(1971) による]

四八年に図書貸出部数が激減した。後期介入群では、一九五三年の放送開始までは貸出部数は増加傾向を示していたが、その後は減少傾向へと変化している。つぎに、群間を比較すると、次のようになる。一九四八年から五二年までの期間は、初期介入群では放送が開始され、後期介入群は未開始であるから、後期群は初期群に対して対照群の役割を果す。この間に、前期群は明らかに貸出部数が減少するが、対照群である後期群は逆に増加傾向を示す。以上のことから、テレビ放送の開始が原因となって図書館の利用率が減少したという結論を得ることが可能であろう。

ただし、この結論は異論の余地がないというほどではない。たとえば、一九四八～五二年の間において、都市地域では、安く購入できる文庫本が普及しつつあり、図書館を利用せずとも読書ができるようになっていたが、村落地域では文庫本は普及していない、という選抜×歴史の要因が作用して、両群の間に差が生じたという可能性がある。これを検討するためには、各

地域での文庫本の販売数を調べるといふ方法がある。

また、読書のいかなる部分に対してテレビが影響を及ぼすかを知るために、図書館の本をフィクションとノンフィクションに分けて、テレビがフィクションの代用となる媒体であるという仮定の下で、フィクションの図書貸出部数とテレビ普及の関係を分析した例もある。^(三三)さらに、フィクションの貸出部数を介入群の従属変数、ノンフィクションの貸出部数を非等価従属変数〔(三)を参照〕とすることによって、ノンフィクションをコントロール・データにして、テレビ普及に伴うフィクション貸出部数の変化を評価することも可能であろう。^(三二)

さらに、テレビの普及によって図書館の利用が確かに減少したとしても、その影響は両地域共に数年間続くが、その後はむしろ上昇傾向を示し、テレビ放送が開始される前の利用率を上回っていることに注目しなければならぬ。これは、テレビと読書が一時的には競合する関係であるとしても、長期的には両立する可能性があることを示すのではないかと思われる。

(2) ゴミ収集についての要請が市民の行動に及ぼす効果^(三三)

清掃作業員がごみを収集する際の安全を確保するためには、市民自身のごみのまとめ方を改善する必要がある。

このために、市民のごみのまとめ方の実態を観察し、まとめ方が悪い場合には、後述のような措置をとった。研究の目的は、一定の処置が市民のごみのまとめ方に及ぼす効果を、多重基線型（スイッチ型）によって評価することである。実験は、アメリカ・カンザス州ローレンス市で行われた。対象は一八三軒の住宅であり、これを主要道の西側の地区の九四軒と東側の地区八九軒に分け、西側を早期介入群、東側を後期介入群とした。

① 方法 清掃作業員が収集を始める前に、実験の担当者が対象地区を一定の順序通りに自動車で巡回し、各住宅の前で停車して、ごみの置き方を点検し、規則違反の回数と種類を記録した。^(三四)実験の全体期間は八週間であ

る。この中でごみ収集日は、一週平均約二日で、全体では一五日であった。担当者の点検は、収集日に必ず行われ、清掃車が来る前に行われた。

実験期間は、観察だけが行われる基線相（事前測定期間）と、観察と指導が行われる介入相（事後測定期間）に分けられた。西地区は、最初の五日間が基線相であり、六日目から最後の一五日目までの一〇日間が介入相である。これに対して東地区は、最初の一〇日間が基線相、最後の五日間が介入相である。したがって、六日目から一〇日目までの五日間では、西地区は介入相、東地区は基線相であるから、東側は西側に対してコントロール・データを提供する非等価対照時系列の役割を果すのである。

基線相の観察は、市委員会の許可と指示が出た直後に開始された。実験対象に特定の地区が選ばれたことや、置き方の不適切なごみは収集されない場合があるということは、地元の新聞に市委員会関連記事として、基線観察の開始前に発表されていた。ただし、基線相の期間は、置き方の不適切なごみも従来通りに収集された。

介入の具体的内容は、指示・強化・情報から成る。指示とは、基線相が終る日に対象の各世帯に対して、事務市長と清掃局長の署名による通知状を郵送することである。通知状には、ごみの置き方や違反の種類、違反を少なくすることが清掃業務の安全と効率を高める、という趣旨のことが書かれている。違反があった場合にごみ箱につけられる札のコピーが同封されている。強化とは、違反のごみを収集せずに残しておくことである。収集されないことと、残されたごみを隣人が見ることがその世帯に対する負の強化となる。情報とは、トラックの後に乗った作業員が、違反の内容を記した札をごみ箱の取っ手につけることである。

八週間（収集日に換算して一五日間）の実験が終了した後で、市から対象の一八三世帯に対して、協力を感謝し、計画が成功したことを記した手紙が郵送された。清掃員に対しては、業務を実験以前の状態に戻すように指示

が出された。さらにその四週間後(収集日に換算して二三日目)に、追跡調査が行われた。

② 結果 図3-9は、両地区における収集日ごとの違反件数を表す。介入前の一日当たり平均違反件数は、西地区で五九件、東地区で八〇件であった。介入後になると、西地区は一〇件、東地区は二三件に減少しており、実験終了の四週間後の追跡調査時にもこの水準が維持されている。

図3-10は、少なくとも一種類の違反をした世帯の数を表す。西地区では、介入前には一日平均四〇%の世帯が違反をしていたが、介入後になると一日平均一三%に減少した。東地区は、介入前には四五%であったが、介入後は二三%に減少し、追跡調査でもこの水準が維持されている。

一八三軒の対象全体のうち一五六軒(八五%)は、全実験期間中に少なくとも一回は違反をしている。このうち八八%の世帯は介入後に違反が減少し、六%は介入後も変化がなく、残りの六%は介入後には逆に違反が増加した。全体の一八三軒のうち七四件(四七%)は、ごみが収集されずに残された経験がある。このうち、四五軒はごみの放置が一日で終わったが、一七軒

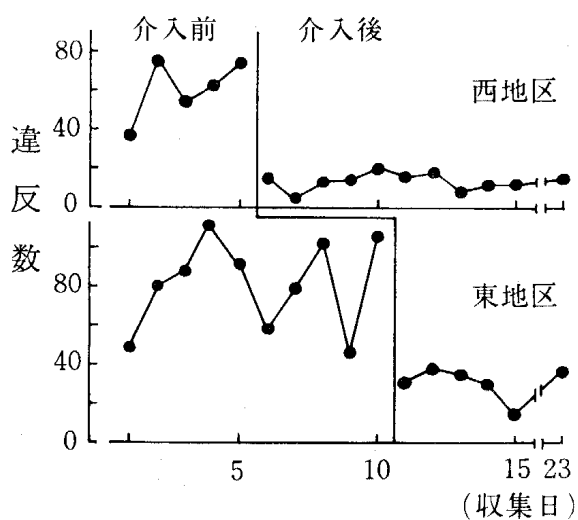


図3-9 ごみの置き方の違反の数 [Posavac & Carey (1985) による]。

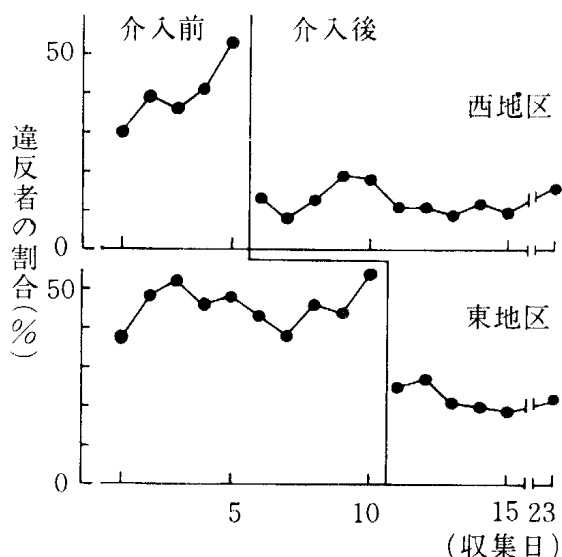


図3-10 ごみの置き方の違反者の割合 [Posavac & Carey (1985) による]。

は二日、七軒は三日、四軒は四日、残りの一軒はそれ以上の日数を要した。

③ 考察 多重基線型（スイッチ型準実験）の特徴は、同一群内部で基線（事前の水準）をコントロール・データとして介入後の変化（事後の水準）を評価するだけでなく、一方の群が介入相に入っている時期に、他方の群を基線相に留めておくことによって、後者をコントロール・データにして前者における介入の効果を評価するといふ点にある。

この実験では、六日目から一〇日目までの期間が、西地区は介入相、東地区は基線相であった。この期間に、西地区の違反は減少し、東地区はそれ以前の五日間の平均水準を維持していた。この期間に生じた群間の差は、同時期に介入した外部的な歴史要因が原因ではないことが示される。^(二二)

ただし、東地区の基線の型は、西地区で介入が始まる六日目から、それまでの増加型から不安定型に変化している（図3—9）。つまり、東地区の基線相における前半の五日間と後半の五日間を比較すれば、基線の平均水準は変わらないが、基線の傾向は変化したといえる。多重基線型は両群が相互に独立であるという前提の下に成り立つ方法であるが、この実験では、西地区に起った状況の変化が東地区に影響を及ぼした可能性があると考えられる。^(一九)

(3) 罰金制度導入が地域の水節約に及ぼす効果〔一〕の(3)、(4)を参照〕

一九七〇年代のアメリカ・サンフランシスコ湾岸地域で、降水量が極度に減少したため、湾岸の各自治体は水節約の対策を講じる必要が生じた。ある自治体は、一九七七年四月から水の超過使用者に対して罰金を課することに^(二一)なった。別の自治体は同年七月から同じ対策を実施した。もう一つの自治体は罰金制度を導入しなかった。その数ヶ月後に雨が降って水不足が解消したので、一九七八年一月に罰金制度は廃止された。この間の水使用量は平年の半分近くまで減少した。問題は、この減少の原因が罰金制度にあるかといえるかということであった。^(二二)

この研究は、次のような特徴をもつ。第一に、一九七五年一月から七八年一二月までの四年間にわたる毎月の地域別水使用量を従属変数とする時系列データである。第二に、罰金による介入時期を変えた二つの非等価地域を設定したという点で、本節で述べてきたスイッチ型中断時系列（多重基線型）である。第三に、罰金のない無介入の地域のデータを並行させた点で、非等価対照時系列付きの方法である〔二〕を参照〕。第四に、罰金が廃止された後の一年半のデータを備えている点で、撤去手順をもつ方法である〔四〕を参照〕。

(5) スイッチ的再現型の利点

スイッチ型は、事前の無介入時系列と、介入時系列と同時的に進行する非等価群無介入時系列を備えている点で、無介入時系列付きの標準型〔二〕を参照〕と少なくとも同程度の説明力をもつ方法である。さらに、無介入時系列の代りに、介入を遅らせた時系列を設定することによって、スイッチ型には次のような利点が生じる。

① 介入効果があらわれるまでの期間の確認 一定の介入が効果をもつにはどの程度の時間が必要であるか、つまりその介入が速効性なのか遅効性であるかを知るためには、単一の時系列では十分ではなく、二群以上の時系列が必要となる。たとえば、一つの群に対して介入が開始されて数日後に効果があらわれたとする。この介入から効果までの数日間の期間は、介入そのものの性質によってだけでなく、それ以外の要因によっても規定されている。介入そのものの効果が生じるまでの期間を知るためには、別の群の時系列の中でこの期間を再確認する必要がある。無介入時系列付きの標準型は、単一群だけに介入を行う方法であるから、効果期間を再確認する手順を欠いている。これに対してスイッチ型は、対照時系列の役割をもつ後期介入時系列において効果期間を再確認できる可能性をもっている。

ただし、スイッチ型でも次のような場合には、介入から効果までの期間を再確認することができない。第一は、

二つの群が著しく非等質な場合である。この場合には、介入から効果までの期間は群によって差があると考えられる。第二は、一方の群に対する介入時期と他方の群の介入時期が時間的に著しく離れている場合である。この場合には、長い間に作用する外部要因を統制することができない。

② 対照群でも介入を行うことの利点 介入が被験者にとって望ましいものであれば、実験に参加するすべての被験者には、その恩恵を受ける資格があるといえる。ところが、無介入対照群付きの標準型では、その恩恵を受けるのは介入群だけであって、対照群の被験者は除外されている。これに対して、スイッチ型の対照群には、時期が遅れるにしても介入が行われる。スイッチ型は、介入群と対照群に対して平等に近い形で対応するという点で、無介入対照群型に勝る方法である。^(二) なお、この点においての無介入対照群型とスイッチ型の関係は、単一事例実験における撤去型と、撤去手順のない多重基線型の関係によく似ている。^(二九)

(六) 中断時系列型の問題点

中断時系列型には、これまで述べてきたものの他に次のような問題がある。

第一は、独立変数に関わる問題である。この型における介入は、正規の実験における変数操作のように一気に行われるのではなく、むしろ対象全体の中に徐々に浸透していく性質をもつことが多い。

第二は、従属変数に関わる問題である。介入の効果は即時的であることもあるが、遅れてあらわれることも多い。どの程度遅れるかを事前に知ることは難しく、実験の対象によっても実験が行われる時期によっても異なる。

第三は、観察の回数に関わる問題である。中断時系列型の観察回数が、時系列という名の通りに非等価対照群型より多いのは当然であるが、統計的分析のために通常必要とされる五〇回よりは少ないのが普通である。

第四は、役所等の文書記録を分析の対象にすることが多いということである。この場合に、誰が記録を保管しているかを知るのは容易ではなく、仮にこれを知り得たとしても彼らは記録を公開するのを好まない傾向がある。また、公開された記録には欠けた部分があったり、説明なしに定義が変更されていることもある。^(三)

このような問題について、次に詳しく説明してみよう。

(1) 介入は一斉に開始されるが、それが浸透するのは漸次的である

中断時系列型における介入は、感覚の実験における立上りの悪い刺激に似た特徴をもつ。つまり、介入自体は一定時期に一斉に行われたとしても、それが対象全体に到達するまでに一定の時間を要することが少なくない(gradual rather than abrupt)。また、介入自体の特徴の他に、介入の対象や時期によって対象に到達するまでの所要時間が異なる場合もある。

① テレビの普及過程　テレビ放送の開始が市民の日常生活に及ぼす影響については、図書館利用との関連ですでに言及したが〔五〕の(1)を参照〕、犯罪との関連についても考察されている。^(二六)この中に、テレビ放送が開始された後で、市民がテレビを購入する過程に関する次のようなデータがある。

図3—11によれば、一九四八年頃から放送が開始された地域では、全世帯のうちの七五％がテレビを持つようになるのは一九五四年であった。一九五三年に放送が開始された地域では、七五％に達するのは一九五七年であった。いずれの地域でも、テレビが全体に拡散するのは、数年間にわたる累積的な現象(diffusion ogive)であった。^(三)ただし、五〇％程度の普及を目安にすると、これに達するのに要した時間は、一九四八年導入地域では四年間、一九五三年導入地域はわずか一年であった。これは、同じ刺激であっても、その導入時期によって立上りの速度が異なる場合である。

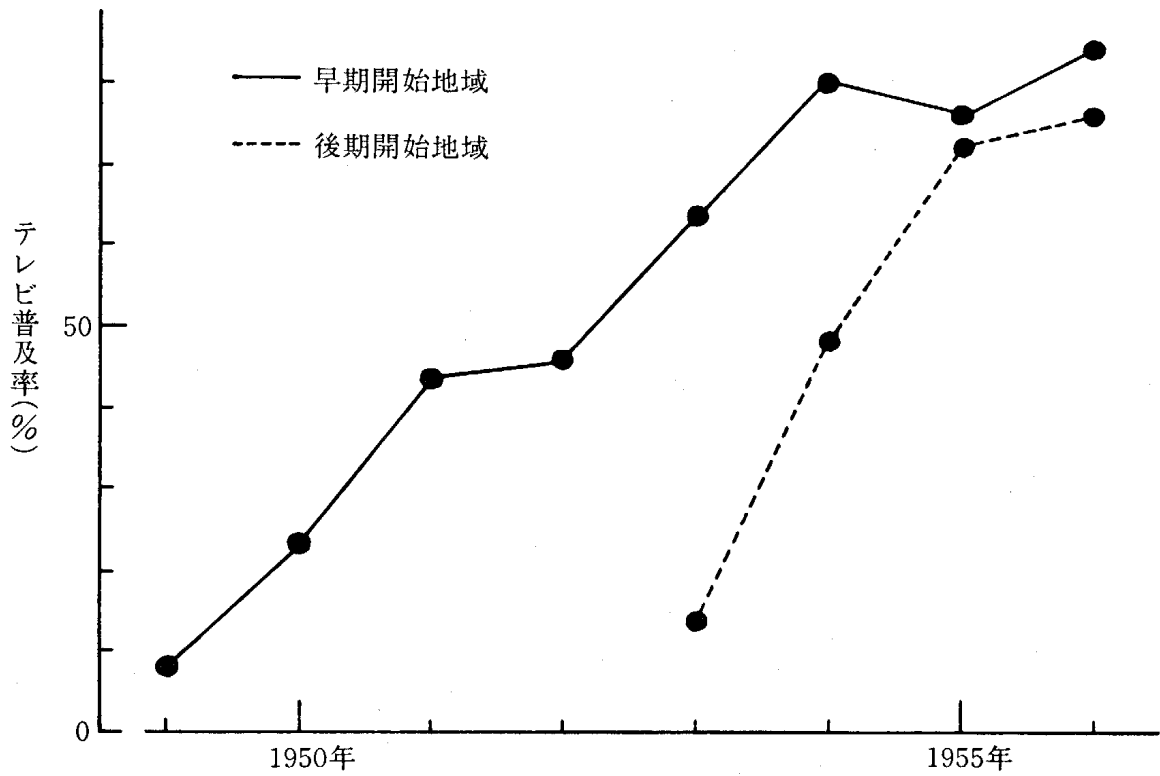


図 3-11 テレビの普及過程 [Cook & Campbell (1979) による]

② 飲酒運転規制の場合 法律が制定ないしは改正されたとしても、これを執行するのが迅速でなかったり、何らかの理由で国民の間に浸透するのが遅れることがある。つまり、法律上の変化は一定時間に一斉に生じるが、それに伴う社会的な変化は国民の間にゆっくりと浸透していくのである。ただし、飲酒検知器導入による飲酒運転の規制〔(三)の(1)を参照〕はその例外であって、国民は新聞等の報道を通じて検知器とその導入時期をよく知っていた。^(三)

③ 問題に対する視点 介入が対象に達するの
 が、全か無かの悉無的なステップ関数(step function)の形で起る場合には、介入と効果の関係を知るの最も容易であるが、現実にはむしろ漸次的・累積的なオージブ関数(ogive function)の形で起ることが多い。介入が対象全体に行われたとしてもその浸透は部分的であるという現実を変えることができないとすれば、これを前提にして方法を変える必要がある。統計学的な時系列分析の中には、これに対処する方法が開

発されているが、これとは別にこの問題を単純に考えれば、次のようになる。つまり、介入の浸透率が一〇〇%ではなく二〇%にすぎないとしても、介入が浸透部分に対しては即時的な効果をもつならば、その二〇%については介入と効果の関係を直接的に知ることができると考えられる。いずれにせよ、介入の有無だけで独立変数の設定が完成するわけではない。これに加えて、介入が対象に浸透していく過程 (diffusion process) のデータを集める手順を事前に準備しておくことは必要である。

(2) 介入の効果は即時的ではなく、遅れてあらわれる

前項(1)では、刺激そのものが対象に到達するまでの過程について述べた。本項の問題は、対象に到達した刺激が対象に対していつ影響を及ぼすかである。つまり、介入の効果は即時的であるとは限らず、ある程度の時間が経過した後で生じることが少なくない。問題は、介入と効果の間に時間差があるかどうかということよりも、どの程度の時間差であるかである。研究分野によっては従来の経験的蓄積によって効果までの時間を予測できることがあると思われるが、フィールドで起る問題はつねに新しいのであるから、事前にこれを予測するのは容易ではない。準実験においてこの問題に対応できるのは、スイッチ型時系列の方法だけであろう〔五を参照〕。これは、時間差を置いて二つ以上の群に同じ介入を行って、同じ効果を再現する方法である。両群において介入から効果までに要する時間が同じであったとすれば、これが介入の効果所要時間であると推定されるのである。

しかしながら、介入から効果までの時間は、対象集団の特徴や導入時期などの多くの妨害要因によって影響を受ける。数年間にわたるデータを扱う中断時系列型では、妨害要因の影響を取り除いて、介入そのものが効果を及ぼすまでの時間を確かめる方法はないと考えなければならない。

(3) 短縮型時系列の効用

中断時系列型は連続測定を重視する方法であるにもかかわらず、観察回数が統計的分析に必要な五〇回に達しないことがある。しかし、単発測定の新等価対照群型でも、一回の事前測定にもう一回追加して事前測定を二回行うことによって「二の(三)の(3)を参照」、妨害要因の点検が容易になる。短縮型の中断時系列 (abbreviated interrupted time series) は、統計学的な条件をみたしていないとしても、単発型測定よりは長所の多い方法である^(二七)。実例に従って、この点を次に説明する。

図3-12は、職業訓練計画への参加が労働者の所得に及ぼす効果を示す^(二七)。ここで、訓練群は、訓練計画に参加した労働者である。対照群は、アメリカ労働当局の年次資料に基づいて訓練計画に関係なく計算された労働者の平均所得である。調査期間は、一九六四年に訓練を開始する前後の一年間であるから、観察回数は一回であって、もちろん五〇回には達しない。従属変数は労働者の年間所得であり、社会保障機関の記録に基づいて計算した。

図に示すように、訓練開始前には対照群の方が所得が多かったが、訓練を開始すると訓練群の所得は急激に増加して、対照

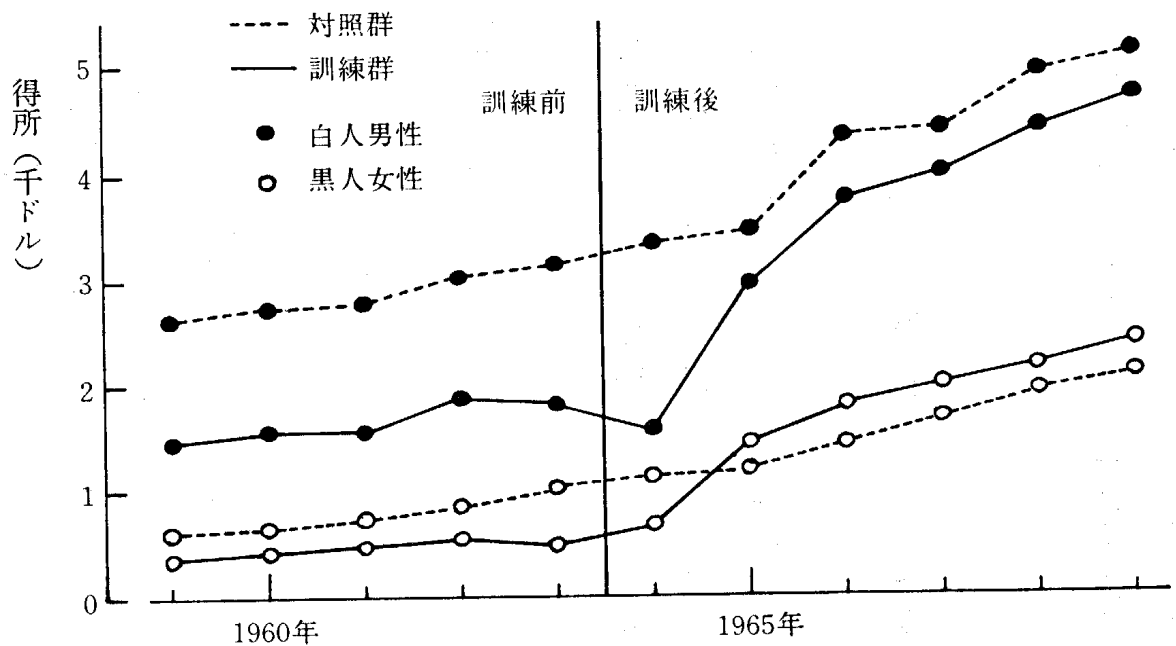


図3-12 職業訓練が労働者の所得に及ぼす効果 [Cook & Campbell (1979) による]

群に接近するか上回っている。これは、訓練が所得に対して因果的効果 (causal impact) を及ぼしたことを示している。ただし、標準的な非等価対照群型〔二の(二)を参照〕によって、一九六三年と六五年のそれぞれ一回だけの所得を事前事後の測定値として考えたとしても、訓練の因果的効果を推論することは不可能ではない。そこで、事前事後の二回の単発的測定と、一回の連続的測定を比較すると、連続的測定の方が妨害要因を排除するためには有効な方法であることを次に示してみる。

第一は、選抜×成熟の要因である。もし、結果が単発的測定によって示されていたとすれば、事後における両群の接近は、各群が事前にもっていた成熟率の差によって生じたという疑いが残る。これに対して、図のような連続的測定によるならば、訓練群と対照群が事前の内部的成熟傾向において同じであったことがわかるから、これによって選抜×成熟の要因が除外されるのである。

第二は、回帰の要因である。一九六四年に職業訓練課程に参加した人はその前の六三年には失業中であったから、その年の実所得は本来の潜在的所得より少ない。その人の所得は、訓練課程に参加しない場合でも、一時的な失業状態から脱すると増加して、自己本来の水準に戻る可能性がある。事前測定を一九六三年に一回だけ行う場合には、この回帰の要因を点検することは不可能である。

第三に、訓練が持続的な効果をもつのか、それとも一時的な効果にすぎないのかは、短縮型であるにしても連続的測定によってのみ知ることができるのである。

(4) 文書記録の限界と効用

時系列研究で使われるデータの多くは、公立や私立の施設の文書保管室から入手される。最近のアメリカでは、各種指標の収集と利用が活発になり、それは経済指標や社会指標だけでなく、主観的社会指標 (subjective social

indicators)にも及んでいる。したがって、時系列研究のデータは将来にわたって増加していくと予想される。しかし、この種の記録を収集して利用する際には、次のような困難がある。^(三)

① 記録の所在 記録がどこにあるかを知ること、利用できるものとできないものを区別することは容易ではない。連邦や州の資料であれば各政府の刊行物から知ることができるが、市町村や学校などの地方当局や私企業の資料を入手するのはさらに困難である。

② 結果中心的な記録 この種の記録は目的に中心をおいて行われるために、問題の過程 (process) よりは結果 (outcome) に重点があることが多い。当然のことながら、問題に関連があると推定される心理過程についての記録は非常に少ない。したがって、問題を説明する際に心理学的概念を好んで使う研究者にとっては、現行の文書記録は、それほど役立つものではないと考えられる。

③ 記録の不完全性 長い年月にわたって行われる記録は、機関の再編成、方針の変更、担当者の交替などによって、次のような欠点をもつことがある。第一は、記録法すなわち指標の操作的定義が途中で変化する場合であり、定義の変更が明記されていない時には事態はさらに悪くなる。第二は、一定期間の数値が記入されていない場合である。さらに悪いのは、実際に測定されていない数値が記録されている場合である。

以上は記録の不備であるが、その他に、たとえ記録が正確に行われてきたとしても、文書記録は基本的に硬直的 (inflexible) な性質をもつという問題がある。第一に、文書記録の多くは年ごとに集計されている。これに対して研究者が望むのは、妨害要因を除外するのに便利な月ごとや季節ごとのデータである。第二に、文書記録には調査対象者の個別的な人口学的特徴 (性・年齢など) が記されていないことがある。もし、この記録があったとすれば、調査対象を分割して非等価対照群を設定することができるから、コントロール・データを得ることが非常に

容易になる。第三に、問題の中心となる記録の他に、同じ調査対象者の各種記録を収集しておけば、これを非等価

従属変数としてコントロール・データを得ることが出来る。文書記録を利用する際には、問題とは直接の関係がない記録にも価値があるという点にも留意しなければならない。^(三)

文書記録にはこのような難点があるにもかかわらず、あるプロジェクトでは、二五年間にわたる時系列データを多方面（消費・余暇・政治活動・公衆衛生・犯罪など）にわたって集めることができた。^(二八) この中には中央（連邦）の記録が部分的に含まれているにしても、その他の大部分は、今までは研究の目的には使われることのなかった地方（州）の無名の文書記録であった。資料の公開は今後さらに進む傾向があるから、中央や地方の文書記録に基づく時系列研究は、過去や現在よりも将来において活発に行われるであろうと予想できる。^(三)

- (一) 芝 祐順・渡部 洋・石塚智一(編)「統計用語辞典」新曜社・一九八四年。
- (二) Cook & Campbell (1979)「前掲・序論の注(一)」。
- (三) Simple interrupted time series.
- (四) Campbell (1969)「前掲・序論の注(一)」。
- (五) Farmer (1924)「ガラス手工業における時間交替制の比較」(英国医学研究審議会産業疲労研究委員会報告書)「前掲・注(一)に於ける」。
- (六) Lawler, E. E., III, & Hackman, J. R. Impact of employee participation in the development of pay incentive plan: A field experiment. *Journal of Applied Psychology*, 53, 467-471, 1969. 「前掲・注(一)に於ける」。
- (七) Rheinstein, M. Divorce and the law in Germany: A review. *American Journal of Sociology*, 65, 489-498, 1959. 「前掲・注(四)に於ける」。
- (八) Glass, G., Tiao, G. C., Maguire, T. O. Analysis of data on the 1900 revision of German divorce laws as a time series quasi-experiment. *Law and Society Review*, 4, 539-562, 1971. 「前掲・注(一)に於ける」。

- (九) Gottman, J.M. *Time-series analysis: A comprehensive introduction for social scientists*. Cambridge University Press, 1981. マンデル「時系列入門——ボックス・ジェンキンスモデルの応用」多賀出版・一九八八年。
- (一〇) Schnelle, J.F. & Lee, J.F. A quasi-experimental retrospective evaluation of prison policy change. *Journal of Applied Behavior Analysis*, 7, 483-496, 1974.
- (一一) Interrupted time series with a nonequivalent no-treatment control group time series.
- (一二) Interrupted time series with nonequivalent dependent variables.
- (一三) Ross, H.L., Campbell, D.T., & Glass, G.V. Determining the social effects of a legal reform: The British "breathalyser" crackdown of 1967. *American Behavioral Scientist*, 13, 493-509, 1970. [Cook & Campbell (1979) 参照]。
- (一四) Ross, H.L. Law, science and accidents. The British Road Safety Act of 1967. *Journal of Legal Studies*, 9, 1-75, 1973. [Cook & Campbell (1979) 参照]。
- (一五) Ross, H.L. The Scandinavian myth: The effectiveness drinking-and-driving legislation in Sweden and Norway. *Journal of Legal Studies*, 4, 285-310, 1975. [Posavac & Carey (1985) 註(一六) 参照]。
- (一六) Interrupted time series with removed treatment.
- (一七) Scheflen, K.C., Lawler, E.E., III. & Hackman, J.R. Long-term impact of employee participation in the development of pay incentive plans: A field experiment revisited. *Journal of Applied Psychology*, 55, 182-186, 1971. [Cook & Campbell (1979) 参照]。
- (一八) Interrupted time series with multiple replications.
- (一九) 田中・前掲 [一の注(一一)]。
- (二〇) Interrupted time series with switching replications.
- (二一) スニッチ型を二つの集団に適用すれば、単一事例実験の個人間多重基線型(multiple baseline design across individuals) とは同値である。二つの地域に適用すれば、場面間多重基線型(multiple baseline design across settings) と全く同値である [田中・注(一九)]。

(一一一) Parker, E. B. The effect of television on public library circulation. *Public Opinion Quarterly*, 27, 578-589, 1963. [Cook & Campbell (1979) による]。

(一二二) Stokes, T. F. & Fawcett, S. B. Evaluating municipal policy: An analysis of a refuse packaging program. *Journal of Applied Behavior Analysis*, 10, 391-398, 1977. [Posavac & Carey (1985)・前掲11の注(一二五)による]。

(一二四)ごみの置き方で違反となるのは、①ごみ箱を収集地点から六フィート以上離れた所に置いた場合、②ごみが箱の上に九インチ以上出ている場合、③ポリ袋が破れているか口を結んでいない場合、④紙袋やボール箱などを使う場合、⑤刈り込んだ草木を小さく束ねていない場合(限度は幅が一八インチ、長さが六フィート)、⑥ごみをそのまま箱の中に置いた場合、である。なお、観察は担当者が一人で行った。観察の信頼性は、第三、五、九、一三日目に、担当者と車の運転者がそれぞれ単独に点検した結果の一致度によって確かめている。

(一二五) Agras, et al. (1980) [前掲・1の注(五)]。

(一二六) Hennigan, Del Rosario, Heath, Cook, Calder, & Wharton (1979) 「テレビの導入は合衆国の暴力および武器所持犯罪にいかなる影響を及ぼしたか」(国立科学財団に対する報告書) [Cook & Campbell (1979) による]。

(一二七) Aschenfelter, O. Estimating the effect of training programs on earnings. *The Review of Economics and Statistics*, LX, 47-59, 1978. [Cook & Campbell (1979) による]。

(一二八) Cook, Calder, & Wharton (1979) 「テレビの普及は、各種社会指標にいかなる影響を及ぼしたか」(国立科学財団に対する報告書) [Cook & Campbell (1979) による]。

結 語

単一事例実験が臨床研究の中で発展した方法であるのに対して、準実験は社会研究の中で体系化されてきた方法である。準実験を健康心理学の方法の中に組み入れる試みがあるにしても、^(一)まだ序論の域に留まると考えられる。したがって、研究の対象面に限定して見る場合には、単一事例実験と準実験は明確に区別しなければならない。

一方、研究の方法面における両法の類似性は明らかである。すなわち、標準的な実験計画とは別の方法を体系化するという目標設定ないしは問題意識、コントロール・データの整備、および対象の時間特性の重視は、二つの方法に共通する特徴である。本論では、方法の基本的位置づけとコントロールの問題を中心に考察を行ったが、時間特性については系統的には論じていない。

対象の時間特性に着目した方法は、心理学の中で中心的な位置を占めてきたとはいえない。しかし、その歴史は古く、伝記や日記、その他の個人的記録に基づく研究はこの系統に属するものであり、^(二)それに対応する統計法も考^(三)えられてきた。最近では、単一事例実験と中断時系列型準実験が対象の時間特性に最重点をおく方法であることは言うまでもないことであるし、これとは別の構想も出されている。^(四)そこで、時間特性の重視という観点から問題を再構成することを次の課題としたい。これと共に、統計法にも言及する必要がある。

(一) West (1985)・前掲「序論の注(三)の②」。

(二) オールポート「心理学における個人的記録の利用法」培風館・一九七〇年。

(三) Baldwin, A.L. The statistical analysis of the structure of a single personality. *Psychological Bulletin*,

37, 518-519, 1940.

- (四) Valsiner, J. Sequence-structure analysis: Study of serial order within unique sequences of psychological phenomena. In Valsiner, J. (Ed.) *The individual subject and scientific psychology*, Chapter 13 (pp. 347-389). Plenum, 1986.

〔付記〕 本論の主題である準実験については、東北心理学会第四二回大会（一九八八年九月一〇日・仙台）で課題講演の機会を与えられ、「単一事例実験——臨床領域における実験計画——」と題して発表した際に、単一事例実験の関連方法として部分的に言及した。また、岩手医科大学医学部衛生学公衆衛生学講座（主任・角田文男教授）において、フィールドの実験法として中断時系列型を中心に発表し、有意義な批評をいただいた（一九八九年六月九日）。関係の方々に対して、ここに記して感謝の意を表す。