RIETI Discussion Paper Series 19-J-003

RCTをめぐる3つの問題とその解法
―精度問題、ノンコンプライアンス、仲介変数による観察中断

山口 一男
経済産業研究所

Research Institute of Economy, Trade & Industry, IAA
独立行政法人経済産業研究所
https://www.rieti.go.jp/jp/

RIETI Discussion Paper Series 19-J-003
2019 年 1 月

RCT をめぐる 3 つの問題とその解法―精度問題、ノンコンプライアンス、仲介変数による観察中断 *

山口 一男（経済産業研究所）

要 旨

本稿では、RCT（randomized controlled trials）で政策評価などを行う際に直面する 3 つの問題に対しての標準的対応を解説する。その 1 つは、精度を高める（あるいは同じ精度で標本数を少なくすることで調査コストを削減する）標本抽出方法についてである。2 つ目はノンコンプライアンス問題である。調査による治療群と統制群の割り当てと、実際に「治療」（treatment）を受けた人と受けなかった人の対応が完全でなく、ランダム性が失われるときの因果推定の問題である。この問題に対して操作変数法による標準的な解法があるが、ここでも精度問題が関連し、本稿は一定条件がデータで満たされれば、より高い精度の推定が可能であることを示したブラックら（Black et al. 2015）の最新の手法をあわせて紹介する。3 つ目は、「死亡」など観察中断を意味する仲介変数が関連する場合(Truncation-by-death の問題といわれる)の治療効果の推定問題である。この問題の重要性と「完全な解決」の難しさは、近年因果推論での話題の 1 つである。本稿では、なぜ DID など通常の解決案が不十分と考えられているかについて、その背後にある考え方を紹介するとともに、この場合は統制群の平均治療効果（ATU）を推定する必要があり、やや強い仮定ではあるが、「無視できる割り当て」を仮定して推定する方法を紹介する。またこの問題の考え方の背後に、ノンコンプライアンス問題と共通の枠組みである、「主層化法（Principal Stratification）」と呼ばれる行動パターンに関する潜在クラスの考えがあり、その考えが解決法の評価にとって重要な役割を示すことをあわせて解説する。最後に、米国での初等教育政策のデータ例を用いて、ノンコンプライアンス問題についての精度の高い因果推定方法の利用に関し具体的に例示する。

キーワード：因果推論、RCT，ノンコンプライアンス、操作変数、Principal Stratification, Truncation-by-death
classification： C21, C26

RIETI ディスカッション・ペーパーは、専門論文の形式でまとめられた研究成果を公開し、活発な議論を喚起することを目的としています。論文に述べられている見解は執筆者個人の責任で発表するものであり、所属する組織及び（独）経済産業研究所としての見解を示すものではありません。

*本稿は、独立行政法人経済産業研究所（RIETI）におけるプロジェクト「日本におけるエビデンスに基づく政策の推進」の成果の一部である。

I.　RCT一何が問題か

実証的根拠に基づく政策立案（ＥＢＰＭ）で、最も頻繁に用いられるのがＲＣＴ（randomized controlled trial）である。ある政策が意図した効果をもたらしたかどうかを因果推定する際に問題になるのは、結果（outcome）について、政策の影響下にある人と、影響下にない人を比べて、その差が因果効果だとみなせるには下記の因果の定義に見合うという条件が必要だが、その条件が一番弱い仮定で満たされるのがＲＣＴの強みなのである。だが現実にはＲＣＴは実行できても、それが推定の精度を上げる上で必ずしも最適な方法でないことがある。また後述する「治療群（treatment group)」と「統制群（control group)」の割り当て上ではランダムでも、それが実際に「治療」を受けた人と受けなかった人との区別に完全に対応しない場合がある。いわゆるノンコンプライアンス問題である。ノンコンプライアンスはランダムには起こらないので、実際に治療を受けた人と受けなかった人の間では選択バイアスが生じる。しかし、RCT のノンコンプライアンス問題は、分析方法論上は、後述する操作変数法によるほぼ完全な解決策がある。１９９６年に JASA（Journal of American Statistical Association）にでたアングリストら（Angrist et al. 1996）の方法である。この方法は本稿でもその一端を紹介するが後にルービンが主層化法（principal stratification method）と呼ぶことになるルービンの因果モデル（RCM）の一つの核の考えになっていく。

しかし、操作変数法にはその二段階最小２乗法（two-stage least squares method）にともなう精度問題が起こる。後述するように操作変数法は、多くの場合非常に精度上効率の悪い方法だからである。本稿では簡単にできる、ノンコンプライアンスによる「歪み」の有無の検定に関するブラックらの方法（Black et al.2015）を紹介し、このテストに合格すれば、より精度の高い推定値が得られることを示す。

さらには RCT には、その主たる結果が後述する生存問題や就業問題であるとき、生存者や就業獲得者の間で治療効果が発生するか否かという重要問題がある。一般的には観察中断あるいは観察開始を示す仲介変数のある場合の仲介後の治療効果問題といわれる。特に前者は Truncation-by-death 問題といわれる。具体例については本稿で紹介するが、この問題はコンプライアンス問題と一見似ているが、実は全く質的に異なるノンランダムな選択バイアス問題が生じる。この問題には、ノンコンプライアンス問題のような「うまい解決策」はなく、後述する「無視できる割り当て（ignorabilty of treatment assignment)」の仮定を置けば選択バイアスを除去できるという標準的方法があるが、この強い仮定を好まない学者も多く、現在進行形でいろいろと別方法が考えられている。本稿では、この問題の解決策というよりは、問題の解説、因果推論上のこの問題の取り扱いの考え方と、それに見合う上記の標準的解決方法を示し、併せてその限界を指摘する。

なお以上の議論は、すべて既に他者により開発された統計的手法に依存し、本稿自体が新たな方法の開発を導入するものではないが、日本においての RCT の理解は、いまだ極めて初等的で、ノンコンプライアンスを無視するか、無視せずとも上記の操作変数法の応用

を超えて、この問題を議論している論文は見当たらない。そのため本稿は啓蒙的見地からも極めて有用と考えられる。以下（1）ノンコンプライアンス問題がない場合、（2）ノンコンプライアンス問題が発生する場合、（3）観察中断や観察開始を示す仲介変数がある場合、についてそれぞれの治療効果推定と精度問題を解説し、最後に（4）ノンコンプライアンス問題に関し、米国の初等教育政策実験のデータを用いて、精度の高い因果推定方法の応用例を示す。

なお、本稿は因果推論に関し広い利用者を想定して統計技術的には初等・中等レベルのものにとどめているが、因果推論の考えそのものに関しては本稿で議論する問題に関し、最新の考え方を併せて紹介している。

II.　ノンコンプライアンス問題がない場合の RCT

II-1　層化しない場合

一般に統計的因果推論では、反事実的（counterfactual）な因果の仮定をする。今各個人 i について治療（treatment）を受けた時の結果を Y_i^1、受けなかった時の結果を Y_i^0 で表すとするとその個人についての治療の因果効果は

$$Y_i^1 - Y_i^0 \qquad (1)$$

となる。だがこの値は計算できない。なぜなら各人は治療を受けるグループ（以下治療群）か治療を受けないグループ（以下統制群）かの一方にしか配置されないため、仮に人々が配置通りに治療を受けたとするなら、治療群なら Y_i^1 は観察できるが Y_i^0 は観察できず、逆に統制群なら Y_i^0 は観察できるが Y_i^1 は観察できないからである。式の（1）はこのように、それぞれのグループに対し、事実と反する状況下での結果を想定するので、反事実的定義と呼ぶのである。また、ここで「仮に」と断ったのは、後述するように、調査対象者のノンコンプライアンス問題というのがあって、治療群や統制群への配置と、実際の治療（treatment）の有無は必ずしも一致しないからである。

以下治療群・統制群への配置を示すダミー変数を Z で表すとする。$Z=1$ は治療群への配置、$Z=0$ は統制群への配置である。この時ＲＣＴは、ノンコンプライアンス問題が起こらない場合、この問題に簡単明瞭な回答を与える。もし人々が治療群に配置されるか統制群に配置されるかが、完全にランダムに決められるなら、それぞれが治療群と統制群を合わせた全体の標本が代表する母集団のランダム標本になる。そのため、それぞれの群の結果の平均値が、標本全体がそれぞれ治療を受けた場合と、受けなかった場合の結果の母集団の推定値となり、従って平均因果効果は、単純に治療群（$Z=1$）と統制群（$Z=0$）の結果の平均値の差

$$\overline{Y}_{Z=1} - \overline{Y}_{Z=0} \qquad (2)$$

で与えられる。

また、この平均治療効果の標準誤差は

$$\sqrt{V(\overline{Y}_{Z=1}-\overline{Y}_{Z=0})}=\sqrt{\frac{\hat{\sigma}^2_{Y|Z=1}}{n_{Z=1}}+\frac{\hat{\sigma}^2_{Y|Z=0}}{n_{Z=0}}} \qquad (3)$$

で与えられる。ここで $n_{Z=1}$、$n_{Z=0}$ はそれぞれ治療群と統制群の標本数で、$\hat{\sigma}^2_{Y|Z=1}$、$\hat{\sigma}^2_{Y|Z=0}$ はそれぞれ治療群と統制群における Y の分散の推定値である。また Y の分散が治療群と統制群で有意に変わらないときは、式（3）の右辺はさらに単純化されて

$$\hat{\sigma}_Y\sqrt{\left(\frac{1}{n_{Z=1}}+\frac{1}{n_{Z=0}}\right)} \qquad (4)$$

となる。これは Y を従属変数、Z を唯一の説明変数としたときの最小 2 乗法を用いた回帰分析における Z の係数の標準誤差と同じである。なお、数式の表記法であるが、「$\overline{Y}$」などのような上付きの「$\bar{}$」の記号は標本平均であることを、「$\hat{\sigma}_Y$」など上付きの「$\hat{}$」の記号は推定値であることを示す。また確率は P で表記し、「$E(Y)$」などの「E」は期待値、即ち母集団の平均値であることを示す。またウェイト ω 付きの期待値には $E_\omega(Y)$ の表記を用いている。

さて標本の最適配分であるが、全体の標本数 N が決められている状態では、Y の分散に治療群と統制群の差がなければ治療群、統制群の標本数が共に全体標本の半分となるときに標準誤差は最小で最も精度が高くなる。Y の分散が治療群と統制群で異なる場合は標本数を分散に比例させる場合が最適となるが、分散の正確な事前推定を要し、通常 Y の分散が治療群と統制群で大きく異なることはないので、同数の標本を採るのが、ほぼ最適とみてよい。

以下本節（II-1 節）では、$\sigma^2_{Y|Z=1}=\sigma^2_{Y|Z=0}=\sigma^2_Y$ を仮定する。通常、標本は治療群と統制群では一標本ごとにかかるコストが異なるので、全体の標本数ではなく予算が一定とした場合に精度を最大化する配分を考える。今一人当たりの標本にかかる費用が治療群と統制群でそれぞれ $c_{Z=1}$ と $c_{Z=0}$ で、標本に関する予算制約を $B=c_{Z=1}n_{Z=1}+c_{Z=0}n_{z=0}$ で表すと、最適配分はこの予算制約の下で、上記の標準誤差を最小化することになり、この解はラグランジェ法[1]を用いると、最適配分は以下の式を満たす。

---

[1] $L=\left(\frac{1}{n_{Z=1}}+\frac{1}{n_{Z=0}}\right)\sigma^2-\lambda(B-c_{Z=1}n_{Z=1}-c_{Z=0}n_{z=0})$ を最小にする λ、$n_{Z=1}$、$n_{Z=0}$ の値を求めることになる。

$$\frac{n_{Z=1}}{n_{Z=0}} = \sqrt{\frac{c_{Z=0}}{c_{Z=1}}} \qquad (5)$$

つまり、 **最適な配分において、治療群と統制群の標本数の比は、治療群と統制群の一標本当たりにかかる費用の比の平方根に反比例する**。例えば新薬の治療効果を測りたいが、新薬の治療費は今までの治療薬の費用の 9 倍である時、治療群と統制群の標本数を調査予算内で 1 対 3 とするのが最適となる。また少人数クラスの効果を測りたいが、少人数クラスの平均生徒数は 15 人、通常クラスは 30 人で、前者の一人当たりの費用が後者の費用の 2 倍の時、治療群と統制群の標本数を調査予算内で 1 対 $\sqrt{2}$ とするのが最適となる。それぞれの標本の絶対数は、この配分比を保つ標本数のうち、予算制限内で最大の数となる。

II-2　層化比例配分

RCT の精度は治療群、統制群のそれぞれについて単純ランダム抽出（simple random sampling）ではなく、層化比例抽出を用いることで、かなり精度を高められる可能性がある。これは後述するように、Yの分散を与えられた層について、層内分散と層間分散の和で表すとき、単純ランダム抽出での治療効果の推定値の標準誤差の 2 乗は Yの全体の分散に比例するのに対し、層化比例抽出の場合は標準誤差の 2 乗は層内分散の平均に比例し、かりに層間分散が 3 分の 1 であれば、標準誤差の 2 乗は 3 分の 2 となり、これは標本数が単純ランダム抽出の 3 分の 2（したがってコストも 3 分の 2）で、同じ精度を得ることを意味する。従って、層化比例配分は大幅に調査コストを削減する可能性がある。以下治療群、統制群別の標本数$n_{Z=1}$と $n_{Z=0}$ はすでに定まっていると仮定する。

今結果（outcome）に対し、性別や年齢など観察できる特性が影響すると仮定する。その中で特に影響の強いカテゴリー属性の組み合わせを「層」とよぶ。例えば年齢と性別が強く結果に影響する場合、層は年齢区分と性別の組み合わせである。今各層を s で表し、p_s を治療群と統制群を合わせた母集団における層 s の割合を示すとしよう。この時、治療群、統制群のそれぞれについて、治療群については各層 s から $n_{Z=1,s} = n_{Z=1} p_s$ の標本を、統制群については、各層 s から、$n_{Z=0,s} = n_{Z=0} p_s$ の標本を取り、各群で層の割合 p_s に比例させるのが層化比例抽出である。またこのとき、平均治療効果の推定値について以下の式を得る。

$$\overline{Y}_{Z=1} - \overline{Y}_{Z=0} = \frac{\sum_{s=1}^{S} n_{Z=1,s} \overline{Y}_{Z=1,s}}{n_{Z=1}} - \frac{\sum_{s=1}^{S} n_{Z=0,s} \overline{Y}_{Z=0,s}}{n_{Z=0}} = \sum_{s=1}^{N_s} p_s (\overline{Y}_{Z=1,s} - \overline{Y}_{Z=0,s}) \qquad (6)$$

つまり、治療効果は、層別の治療効果の加重平均となり、加重平均の重みは単純に層の割合である。さてこの推定値の標準誤差は

$$\sqrt{V\left(\sum_{s=1}^{N_s} p_s(\bar{Y}_{Z=1,s}-\bar{Y}_{Z=0,s})\right)}=\sqrt{\sum_{s=1}^{N_s} p_s^2\left(\frac{\sigma^2_{Y|Z=1,s}}{n_{Z=1,s}}+\frac{\sigma^2_{Y|Z=0,s}}{n_{Z=0,s}}\right)}$$
$$=\sqrt{\frac{1}{n_{Z=1}}\sum_{s=1}^{N_s} p_s\sigma^2_{Y|Z=1,s}+\frac{1}{n_{Z=0}}\sum_{s=1}^{N_s} p_s\sigma^2_{Y|Z=0,s}} \quad (7)$$

となり、式（３）比べると、治療群、統制群別に、式（３）における Y の全体の分散が、式（７）では Y の層内分散の加重平均に置き換わった形であり、Y の層間分散が取り除かれた分精度が上がることがわかる。

層化比例抽出は、必ずしも最適配分ではない。理論上はネイマンの最適配分（J.Neyman's optimal allocation）というものが存在する。しかしこの配分は各層内の Y の分散に関する正しい事前知識を必要とし、その知識が不正確なら比例配分より精度が落ちる可能性がある。また層内分散が層で変わらないならば、ネイマンの最適配分は比例配分と同等になる。したがって、正しい事前知識を要求しない比例配分が多くの場合望ましいと考えられる。なお、治療効果の分散の推定式は式（７）でなく、式（３）（Y の分散が治療群と統制群で同じなら式（４））をそのまま用いればよい。層化比例配分後は標本内の Y の分散が層内分散のみになるからである。

ただし、研究者が平均治療効果の推定ではなく、例えば男女間などグループ間の治療効果の差に主たる関心がある時は、比例配分でなく、男女の標本数は、それぞれ治療群内、統制群内で同数にするのが差の推定値の精度を高めるには良い。またこの場合男女合わせた標本の平均治療効果の推定には、男女の標本抽出率を考慮して、それぞれのグループの標本抽出率の逆数をウェイトとして加重平均を求める必要がある。

III.　ノンコンプライアンス問題がある場合の RCT

III-1　古典的操作変数法の論理と治療効果の推定値

ノンコンプライアンス問題というのは、せっかく人々を治療群と統制群にランダム配置しても、それに従わない人々がでてくることで生じる問題である。例えば新薬の治療効果の測定に関し、治療群に配置されても新薬を拒否する人が出てくるかもしれない。また逆に統制群に配置されたのに、別の医者を通じて新薬の治療を受けてしまう人が出てくるかもしれない。問題は、このようなノンコンプライアンスはランダムには起こらず、その結果実際に治療を受けた人と受けなかった人の区別は、もはやランダムな標本の配置とは言えなくなる点である。特に治療群の治療拒否者は多くのケースで存在する。例えば失業者への職業訓練プログラムでも、治療群に配置されてもプログラムを受講しない人が出る可能性はある。このような時に、どう治療効果を推定するかが問題だが、この問題解決は実は計量経済学的推定方法である操作変数法と深く関係している。ただ古典的操作変数法は因果効果

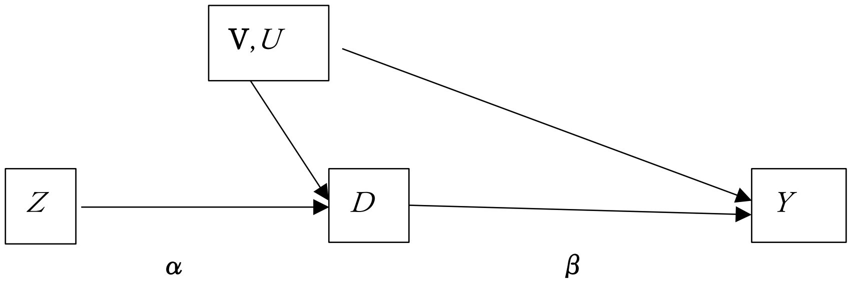
が一様であるという仮定を置くのだが、ルービンの因果モデル（ＲＣＭ）では因果効果は人々の間で多様で一定ではないと仮定する。このＲＣＭの仮定の下での操作変数法の解釈が、以下で解説する LATE（local average treatment effect、局所平均治療効果）であり、それはまた後述する主層化法（principal stratification）の考えのもとになるのである。

今治療群と統制群の RCT によるランダムな標本割り当てを Z、実際の治療の有無を D で表すとすると、以下の図 1 の因果構造があると仮定できる。この図において $\mathbf{V}$ と U はそれぞれ D の結果 Y に対する影響の観察できる交絡要因、観察できない交絡要因である。観察できない交絡要因があるため、D への選択バイアスを簡単に除去できないことが想定できることが問題の核である。

ここで RCT であることからくる重要な仮定が 2 つある。一つは Z はランダムな割り当てなので、$\mathbf{V}$ からも U からも独立という仮定である。この仮定は Z が真にランダムな割り当てなら成り立つ。図 1 では Z と（$\mathbf{V}, U$）を直接繋ぐ線がないことでそれを示している。二つ目の仮定は Z は D を通して結果 Y に間接的に影響するが、直接的な影響（Z から Y への直接の効果）はないという仮定である。これは、結果は実際に治療がされたかどうか（D）のみに依存し、治療群に配分されたか否か（Z）は、実際に治療が行われたかどうかを通じて間接的にのみ結果に影響し、直接結果には影響しないという仮定である。図 1 では Z から Y への直接的な影響を示す矢印がないことでこの仮定を示している。Z が D を通じてのみ影響するという仮定は、**除外条件**（exclusion criterion）と呼ばれ、操作変数法の応用には欠かせない条件であるが、Z がランダム割当であっても状況によりこの条件が成り立たない場合もあるので注意を要する。なお図 1 において $\boldsymbol{\alpha}$ と $\boldsymbol{\beta}$ は、それぞれ Z の D に対する因果効果、D の Y に対する因果効果を示している。

除外条件は通常は成り立つが、ランダムな割り当てが、個人単位でなく、グループ単位（たとえは学校のクラス単位）などで割り当てられると、自分は治療（例えば特別研究プログラム参加）を受けていなくても、クラスの多くが受けていると、クラスメートの影響を受けるというスピル・オーバー効果（波及効果）のせいで Z が直接 Y に影響する場合がある。この場合除外条件は成り立たない。

図 1.　ノンコンプライアンス問題がある場合の因果図式



さて、「古典的操作変数法」であるが、一般に Y が D に線形に依存し、D が Z に線形に依存すると仮定すると、Z の Y への影響はその二つの積となる（図の 1 では $\alpha\beta$ ）という性格を持つ。また仮定により Z はランダムなので、Z の Y への影響 $\alpha\beta$ は因果効果である。また Z の D への影響 α も因果効果である。従って、D の Y への因果効果 β の推定値は、$\alpha\beta$ の推定値を α の推定値で割った以下の式で与えられる。

$$\hat{\beta} = \frac{\widehat{\alpha\beta}}{\hat{\alpha}} = \frac{\bar{Y}_{Z=1} - \bar{Y}_{Z=0}}{\bar{D}_{Z=1} - \bar{D}_{Z=0}} \qquad (8)$$

これが治療効果（D の Y への因果的影響度）の推定値となる。またこの推定値の標準誤差は、二段階最小 2 乗法で得られ、詳細は省くが、既存の統計ソフトを用いた算出については後述の応用で示す。

III-2　操作変数の治療変数への影響および治療変数の結果への影響がどちらも一様でなく、2 つの影響が独立でないとき—LATE 法の解釈

実は、近年の因果推論の考えからみると、上記の「古典的操作変数法」には一つの大きな欠点がある。それは図 1 で暗黙に Z の D に対する影響も、D の Y に対する影響も一様だと仮定している点である。ルービンの因果分析モデルに始まる、反事実的定義に基づく因果効果の定義では、因果効果は各人で異なるため、平均因果効果を測定するという発想に基づいている。RCT における治療効果の推定式（1）も実は平均因果効果の推定式である。しかし図 1 の α, β を個人別の α_i, β_i に置き換えると、Z の Y に対する因果効果の平均 $\overline{\alpha\beta}$ は Z の D に対する因果効果の平均 $\overline{\alpha}$ と平均治療効果 $\overline{\beta}$ の積とは必ずしも一致しなくなってしまうのである。つまり、$\overline{\alpha\beta} = \overline{\alpha} \times \overline{\beta}$ が一般には成り立たないので、治療効果を $\overline{\alpha\beta} / \overline{\alpha}$ で求めるのは正しいといえなくなる。一般に $\overline{\alpha\beta} = \overline{\alpha} \times \overline{\beta}$ が成り立つのは、$\{\alpha_i\}$ と $\{\beta_i\}$ が独立の場合なのだが、ノンコンプライアンス問題ではこの独立性は極めて疑わしい。例えば医療の治療例を取ると、もし治療効果の高い人（β_i の大きい人）が統制群に配置されたのに

治療を受けてしまう（α_i が大きくなる）傾向があったり、治療効果が低い人（β_i の小さい人）が治療群に配置されたのに治療を受けない（α_i が小さくなる）傾向があったりすると、$\{\alpha_i\}$ と $\{\beta_i\}$ は正に相関し独立でなくなるからである。

この問題に一定の解決を与えたのが 1996 年に JASA（米国統計学会雑誌）に掲載されたアングリスト・インベンス・ルービン（Angrist, Imbens & Rubin　1996）の LATE(局所平均治療効果)に関する論文である。この論文で紹介された考え方は、下記の理由で統計学的には異例のものであるが、反事実的因果推論の創始者であるルービンが提唱したせいもあり、下記の潜在クラスの考えは、この論文を超えて、principal stratification（主たる層）の概念として因果推定のあり方に大きく影響を与えていくことになる。その一端は本稿でも紹介する。

LATE 法では、割り当て変数 Z と治療変数 D との関係に関し、理論的には下記の 4 つの異なる行動原理を持つ潜在クラスが存在し、実際にはそのうち 3 つのクラスのみ実在すると仮定する。

その 4 つの潜在クラスは、原著の用語を踏襲すると、それぞれ、「Always takers」、「Never takers」、「Compliers」、「Defiers」と呼ばれる。以下本稿では「A クラス」、「N クラス」、「C クラス」、「D クラス」と呼ぶ。

A クラスの人々は、治療群に配置されても、統制群に配置されても治療を受ける人々を意味する。逆に N クラスの人々は治療群に配置されても、統制群に配置されても治療を受けない人々を意味する。C クラスは治療群に配置されたら治療を受け、統制群に配置されたら治療を受けない、つまり配置に従う人々である。D クラスは、日本語では「天邪鬼」とでもいう人々で、治療群に配置されたら治療を受けず、統制群に配置されたら逆に治療を受ける人々である。

表 1 は各クラスに対し治療群（$Z=1$）と統制群（$Z=0$）への配置別の実際の治療の有無である D の値を示している、最後のコラムは各クラス別の Z の D への因果効果である、α_i の値である。A クラスと N クラスでは Z の値は D の値に影響を全く与えないので $\alpha_i = 0$ であり、C クラスでは $D = Z$ となるので、$\alpha_i = 1$、D クラスでは $D = 1 - Z$ となるので $\alpha_i = -1$ となる。

表 1． 4 つの潜在クラスの行動原理

	D の値		α の値
	$Z=0$	$Z=1$	
A クラス	1	1	0
N クラス	0	0	0
C クラス	0	1	1
D クラス	1	0	-1

さらに LATE 法では、理論上ではこれらの 4 つの行動原理は可能だが、実際には割り当てと常に反対の行動をとる D クラスの人びとは存在しないと仮定する。この仮定は $D(z)$ について、$D(1) \geq D(0)$ が満たされることを意味し、**単調性の仮定**（monotonicity assumption）と呼ばれる。

さて、LATE 法による結果を紹介する前に、筆者はこの方法の仮定が統計学的には異例と述べたがその理由について述べておこう。統計学では、通常変数を確率的に定義し、潜在的変数でもそれは通常は同じである。しかし LATE 法においては、潜在クラスの A、N、C、D の各々において、治療の有無を表す治療変数 D の値は確率的にではなく、１か０の定数に定まると仮定されている。またその結果、D クラスが存在しないという追加仮定の下では、表 1 が示すように α_i の値は個人別には 1 か 0 と 2 種類に限定されることになる。これは分析上は非常に便利な仮定であるが、人間の行動原理に即しているかどうかは疑いの余地は残る。通常の統計学では変数 Z が 2 値をとる変数 D に影響する時、Z が**個人レベル**で $D=1$ となる確率に影響を与えると考える。一般に人間行動を確率的に考えるからである。だが、LATE 法において集団レベルでは Z の D に対する影響は確率的だが。個人レベルでは非確率的で、そこが異例である。しかし非確率的潜在クラスの考えは、LATE 法での洞察に留まらず、因果分析に大きな影響を与えていくことになった。

さて以上の仮定の下に Z の Y に対する効果 $\alpha_i\beta_i$ の平均を計算すると、A クラスと N クラスは $\alpha_i\beta_i=0$、C クラスでは $\alpha_i\beta_i=\beta_i$ である。従って以下の式を得る。

$$\overline{Y}_{Z=1}-\overline{Y}_{Z=0}=\frac{1}{n}\sum_{i=1}^{n}\alpha_i\beta_i=\frac{1}{n}\sum\nolimits_{i|compliers}\beta_i=\frac{n_{compliers}\overline{\beta}_{compliers}}{n} \qquad （９）$$

また、Z の D に対する影響は

$$\overline{D}_{Z=1}-\overline{D}_{Z=0}=\frac{1}{n}\sum_{i=1}^{n}\alpha_i=\frac{1}{n}\sum\nolimits_{i|compliers}1=\frac{n_{compliers}}{n} \qquad （１０）$$

となる。従って、式（９）と（１０）から以下を得る。

$$\hat{\beta}_{compliers}=\frac{\overline{Y}_{Z=1}-\overline{Y}_{Z=0}}{\overline{D}_{Z=1}-\overline{D}_{Z=0}} \qquad (11)$$

式（８）と式（１１）を比べると平均治療効果の推定式は全く同じであることがわかる。つまり LATE 法は、操作変数法による推定値は潜在クラスである C クラスの平均治療効果であるという解釈上の注意を別にすれば、治療効果が個人別で異質であるという現代的の因果推論の仮定の下でも有効な治療効果推定法であるということである。ただノーベル経済学者で著名な計量経済学者のヘックマン（Heckman　1997）は（１１）式の推定では C クラスは潜在クラスであり個人は特定できず、従ってその母集団を特定できない点は、解釈

上の大きな制約であるとの意見を述べている。ただし、今調査において、治療群割り当て者のノンコンプライアンスは可能だが、統制群は割り当て者が治療を受けることは調査デザイン上不可能にできれば、A クラスは存在せず、治療群は C クラスのみになる。この場合は操作変数法の推定値は治療群における平均治療効果である ATT（average treatment effect for the treated）の推定値となる。

III- 3　ブラックら（Black et al. 2015）による治療変数への選択バイアスのテストと、テストの結果一定条件件が満たされるときのより精度の高い治療効果推定法

このようにノンコンプライアンス問題に関する操作変数法の応用は、因果推論の理論上でも妥当であるという結論に達したのだが、応用上は大きな問題がある。それはその推定値の精度が高くない点である。一般にもし実際の治療の有無 D に選択バイアスがない場合は、平均治療効果は単に $\overline{Y}_{D=1}-\overline{Y}_{D=0}$ で簡単に与えられるが、この場合の標準誤差と操作変数法の標準誤差の間に

$$SE(\hat{\beta}_{IV}) = SE(\hat{\beta}_{OLS})/\rho(Z,D) \qquad （１２）$$

の関係が得られる。これは操作変数法の治療効果（ β_{IV} ）の推定値の標準誤差が、D の状態への選択バイアスがない場合に、治療群と統制群の分散が同等と仮定した時の最小 2 乗法による $\overline{Y}_{D=1}-\overline{Y}_{D=0}$ の標準誤差の推定値（式（１２）では $SE(\hat{\beta}_{OLS})$ と表示）に比べ、Z と D の相関係数 $\rho(Z,D)$ に反比例して大きくなるという事実を示す。もし Z と D の相関が 0.5 なら、標準誤差は 2 倍となり、選択バイアスがない場合の標準誤差と同じ程度の精度をえるには標本数を 4 倍にしなければならず、従って標本にかかるコストも 4 倍にしなければ達成できないことになる。またこのことは実際には因果効果があっても、精度が低いために結果に有意差を得られなくなる可能性が非常に高いことを示す。

この問題に対し最近ブラックら（Black et al. 2015）は非常に簡明で有力な D の内生性のテストの方法を提案した。内生性とは、D と Y に共に影響する観察できない交絡変数（図 1 の U）が存在することを言う。この方法によれば、このテストに合格すれば（D の内生性を否定できれば）精度の悪い操作変数法でなく、$\overline{Y}_{D=1}-\overline{Y}_{D=0}$ もしくは、Y の重み付き平均差、$\hat{E}_\omega(Y|D=1)-\hat{E}_\omega(Y|D=0)$を治療効果の推定値として用いることができることになる。$\overline{Y}_{D=1}-\overline{Y}_{D=0}$ はもとより重みつき平均の差である$\hat{E}_\omega(Y|D=1)-\hat{E}_\omega(Y|D=0)$であっても、操作変数法の精度を上回る（因果効果の標準誤差が小さくなる）。

ブラックらの方法の特徴は、LATE 法同様、A、N、C の 3 つの潜在クラスを仮定し、D クラスは存在していないという単調性の仮定に基づいている点である。以下その考えを述べる。今 Z はランダムな割り当てで、A、N、C の 3 潜在クラスを仮定すると、$Z=1$ および $Z=0$ のそれぞれに対し A クラス、N クラス、C クラスの割合（それぞれ、P_A, P_N, P_C で示す）は $D=1$ と $D=0$ の別に表 2 のようになる。

表２　割り当て別の潜在クラスの構成比

	$D=1$	$D=0$	計
$Z=1$	P_A+P_C	P_N	1.0
$Z=0$	P_A	P_C+P_N	1.0

表 2 から C クラスの割合が $\overline{D}_{Z=1}-\overline{D}_{Z=0}$ で推定できることも容易にわかる。さて表２から、D = 1 のデータについて、Yは Dが定まれば Zには直接依存しないこと（除外条件が満たされること）を考慮すると、以下の一連の式が得られる。

$$E(Y\mid D=1,Z=1)=\frac{P_A}{P_A+P_C}E(Y\mid D=1,A)+\frac{P_C}{P_A+P_C}E(Y\mid D=1,C) \qquad (13)$$

$$E(Y\mid D=1,Z=0)=E(Y\mid D=1,A) \qquad (14)$$

ここで A、C はそれぞれ A クラス（Always takers）と C クラス（Compliers）を示す。従って、式（１３）と（１４）の差をとることにより、

$$E(Y\mid D=1,Z=1)-E(Y\mid D=1,Z=0)=\frac{P_C}{P_A+P_C}\left(E(Y\mid D=1,C)-E(Y\mid D=1,A)\right) \qquad (15)$$

を得る。今 D = 1 のデータに限って Yを従属変数 Zを唯一の説明変数として以下の線形回帰式を当てはめると Zの係数は式（１５）の右辺と同値となる。

$$E(Y\mid D=1)=\alpha_{CA}+\beta_{CA}Z \qquad （１６）$$

$$\beta_{CA}=\frac{P_C}{P_A+P_C}\left(E(Y\mid D=1,C)-E(Y\mid D=1,A)\right) \qquad （１７）$$

式（１７）は、回帰式（１６）の β_{CA} の推定値$\hat{\beta}_{CA}$が有意に 0 と異ならなければ、実際に治療を受けた人（D = 1）の中で結果 Yの平均は A クラスと C クラスは有意に異ならない、つまり $E(Y\mid D=1,C)=E(Y\mid D=1,A)$ の帰無仮説が棄却できないことを示す。ここで、重要な点は除外条件を仮定しているので、式（１６）で Zの Yへの影響が有意であれば、それが即ち D = 1 の状態への式（１５）で表せる選択バイアスがあるということになり、逆に有意でなければ選択バイアスがない（A クラスと C クラスの間に結果の有意差はない）ということになり、$\beta_{CA}=0$ 検定は、その選択バイアスの有無の検定になっているという点である。

同様に $D = 0$ のデータについて以下の一連の式が得られる。

$$E(Y \mid D=0, Z=1) = E(Y \mid D=0, N) \quad （１８）$$

$$E(Y \mid D=0, Z=0) = \frac{P_N}{P_N + P_C} E(Y \mid D=0, N) + \frac{P_C}{P_N + P_C} E(Y \mid D=0, C) \quad (19)$$

ここで N、C はそれぞれ N クラス（Never takers）と C クラス（Compliers）を示す。従って、式（１８）と式（１９）の差を取ることにより、

$$E(Y \mid D=0, Z=1) - E(Y \mid D=0, Z=0) = \frac{P_C}{P_N + P_C}\left(E(Y \mid D=0, N) - E(Y \mid D=0, C)\right) \quad (20)$$

を得る。今 $D = 0$ のデータに限って Y を従属変数、Z を唯一の説明変数として以下の線形回帰式を当てはめると Z の係数は式（２０）の右辺と同値となる。

$$E(Y \mid D=0) = \alpha_{CN} + \beta_{CN} Z \quad （２１）$$

$$\beta_{CN} = \frac{P_C}{P_A + P_C}\left(E(Y \mid D=1, N) - E(Y \mid D=1, C)\right) \quad （２２）$$

式（２２）は、回帰式（２１）の β_{CN} の推定値 $\hat{\beta}_{CN}$ が有意に 0 と異ならなければ、実際に治療を受けなかった人（$D =$ ０）の中で結果 Y の平均は C クラスと N クラスは有意に異ならないという結論を得る。

従って以下が成り立つ。

もし、式（１６）と式（２１）をデータに当てはめて、帰無仮説 $\beta_{CA} = 0$ も $\beta_{CN} = 0$ も共に棄却できないとき、平均治療効果は

$$\overline{Y}_{D=1} - \overline{Y}_{D=0} \quad （２３）$$

で与えられる。

またデータによっては治療群の治療拒否者は存在しうるが、統制群に割り当てられた人が治療を受けることは不可能という場合がある。この場合 A クラスは存在しえないので、式（１６）のテストは必要なく、式（２１）のテストで Z の係数が有意でなければ、推定式（２３）を用いることができることになる。

しかし、上記の条件は必ずしも満たされるとは限らず、より弱い条件で操作変数法より精度の高い推定値が得られることが望ましい。上記の式（１６）と（２１）は観察されない交絡要因だけでなく、観察される交絡要因もないという強い仮説の検定になっているか

らである。では、観察される交絡要因は存在するが、観察されない交絡要因はないという仮説を検定したい場合はどうすればよいか。

ここで「強く無視できる割り当て」が成り立つ場合の因果推定の考えを生かすことができる。「強く無視できる割り当て」というのは、ルービンの因果分析では、選択バイアスをもたらす交絡要因は観察できる変数 $\mathbf{V}$ で与えられるという、比較的強い仮定を意味する（星野 2009)。**しかし以下では「強く無視できる割り当て」を仮定ではなく、検証できる仮説として扱うことができ**、これが大きな違いである。

Z はランダムな変数だが、標本を D=1、あるいは D ＝ 0 を満たすものに限ると、上記で示したように、$D = 1$ に限ると A クラスと C クラスの混合割合が $Z = 1$ と $Z = 0$ で異なり、A クラスと C クラスの区別はランダムな割り当てではないため、Z= 1 と Z= 0 ではランダムな割り当てにならなくなってしまう。$D = 0$ に限る場合でも N クラスと C クラスの割合が Z= 1 と Z ＝ 0 で異なるという同様の問題が起こる。

今 $\mathbf{V}_{CA}$ を $D = 1$ における、A クラスと C クラスの区別と結果 Y への観察できる交絡要因とする。$\mathbf{V}_{CA}$ は、割り当て変数 Z が定まる以前に決まる変数（因果的に Z に影響をうけない変数）で、結果 Y に影響する変数のうちから $D = 1$ という制約の下で Z と相関する変数である。ブラック（Black et al 2015）は、このような交絡変数の制御について、式（１６）に $\mathbf{V}_{CA}$ を制御変数として回帰分析的に加えて行う方法を紹介している。つまり式（１６）と式（２１）を以下のようにそれぞれ制御変数 $\mathbf{V}$ で統制する形に拡張し、

$$E(Y \mid D=1)=\alpha_{CA}+\beta_{CA}Z+\boldsymbol{\gamma}_{CA}'\mathbf{V}_{CA} \qquad （１６R）$$

$$E(Y \mid D=0)=\alpha_{CN}+\beta_{CN}Z+\boldsymbol{\gamma}_{CN}'\mathbf{V}_{CN} \qquad （２１R）$$

そのうえで、もし帰無仮説 $\beta_{CA}=0$ も $\beta_{CN}=0$ も共に棄却できないときは、観察されない交絡要因はない、つまり治療変数の内生性が否定できるという結論を得る。

しかし、筆者は式（１６R)、（２１R）の利用は簡便であるという点で利点はあるが、交絡要因 $\mathbf{V}$ をより弱い仮定の下で制御するには IPT（Inverse Probability of Treatment）ウェイトを用いる方法がより優れていると考える（ただし、$\mathbf{V}_{CA}$ が単独あるいは少数のカテゴリー変数なら、ダミー変数を用いて主効果と有意な交互作用効果を回帰モデルで制御したほうが簡便である)。なぜなら内生性を否定したあと、平均治療効果の推定には $\mathbf{V}$ の分布を治療群と統制群で同じにする必要があるので、いずれにせよ IPT ウェイトなどを用いる必要性が生じるからである。

A クラスと C クラスの区別は直接観察できないが、$\mathbf{V}_{CA}$ が Z と独立になる状態を $D =$ 1 を満たすデータ上作り出せば、もし $\mathbf{V}_{CA}$ の制御による「強く無視できる割り当て」が成立するならば、結果 Y は A クラスと C クラスの区別とも独立になる。逆に $\mathbf{V}_{CA}$ を Z と独立にしても結果 Y が A クラスと C クラスの区別と独立にならなければ、「強く無視できる割り当て」は成立しない。

詳細は星野（2009）や山口（2017）における第3章のDFL法に関連した傾向スコアによるIPTウェイトの算出法を参照されたいが、$\mathbf{V}_{CA}$ が Z と独立になる状態を $D=1$ を満たすデータ上作りだすには以下のIPTウェイト

$$\omega_{D=1}(z)=\frac{P(\mathbf{v}_{CA}\mid D=1)}{P(\mathbf{v}_{CA}\mid D=1,Z)}=\frac{P(Z\mid D=1)}{P(Z\mid D=1,\mathbf{v}_{CA})} \quad (24)$$

を $Z=1$ と $Z=0$ のそれぞれの場合に $P(Z\mid D=1,\mathbf{v}_{CA})$ についてのロジスティック回帰分析の推定値を用いて算出し、そのウェイトをデータに掛ければ良い。そして以下の回帰式をウェイト付きの $D=1$ のデータに応用する。

$$E_{\omega_{D=1}}(Y\mid D=1)=\alpha_{CA}+\beta_{CA}Z \quad (25)$$

この時 β_{CA} の推定値が有意に0と異ならなければ、$\mathbf{V}_{CA}$ を制御した後では、$D=1$ における結果 Y の平均はAクラスとCクラスの区別に依存しないこと、つまり「強く無視できる割り当て」が実際に成り立つことを示す。

同様に今 $\mathbf{V}_{CN}$ を $D=0$ における、NクラスとCクラスの区別と結果Yへの観察できる交絡要因とする。$\mathbf{V}_{CN}$ は、割り当て変数 Z が定まる以前に決まる変数（因果的に Z に影響を受けない変数）で、結果 Y に影響する変数のうちから $D=0$ という制約の下で Z と相関する変数である。

NクラスとCクラスの区別は直接観察できないが、$\mathbf{V}_{CN}$ が Z と独立になる状態を $D=0$ を満たすデータ上作り出せば、もし $\mathbf{V}_{CN}$ の制御による「強く無視できる割り当て」が成立するならば結果 Y はNクラスとCクラスの区別とも独立になる。この場合のIPTウェイトである、

$$\omega_{D=0}(z)=\frac{P(\mathbf{v}_{CN}\mid D=0)}{P(\mathbf{v}_{CN}\mid D=0,Z)}=\frac{P(Z\mid D=0)}{P(Z\mid D=0,\mathbf{v}_{CN})} \quad (26)$$

を $Z=1$ と $Z=0$ のそれぞれの場合に $P(Z\mid D=0,\mathbf{v}_{CN})$ についてのロジスティック回帰分析の推定値を用いて算出し､以下の回帰式をウェイト付きの $D=0$ のデータに当てはめる。

$$E_{\omega_{D=0}}(Y)=\alpha_{CN}+\beta_{CN}Z \quad (27)$$

この時 β_{CN} の推定値が有意に0と異ならなければ、$\mathbf{V}_{CN}$ を制御した後では、$D=0$ における結果 Y の平均はNクラスとCクラスの区別に依存しないこと、つまり「強く無視できる割り当て」が実際に成り立つことを示す。

これらの結果から、もし、式（25）と（27）をデータに当てはめて、帰無仮説 $\beta_{CA}=0$ も $\beta_{CN}=0$ も共に棄却できないとき、平均治療効果は

$$\hat{E}_{\omega_{D=1}}(Y|D=1)-\hat{E}_{\omega_{D=0}}(Y|D=0) \qquad (28)$$

で与えられることがわかる。

また、この場合も調査上統制群の人々が治療を受けることは不可能で A クラスが存在しえない場合には、$D=1$ は C クラス（Compliers）のみになり、式（25）のテストは必要なく、式（27）のテストで Z の影響が有意でなければ、（28）の式より簡略で精度の高い

$$\overline{Y}_{D=1}-\hat{E}_{\omega_{D=0}}(Y|D=0) \qquad (29)$$

を用いて平均治療効果を推定できることになる。

IV. 観察中断や観察開始を示す仲介変数がある場合の治療効果の推定法と考え方

「観察中断を示す仲介変数問題」には以下がある。「Truncation-by-death」問題とも言われる。ザングとルービン（Zhang and Rubin 2003）の医療の例で説明しよう。今新たな癌治療法の効果を見たいとする。しかし、医療研究者はその治療法により癌治癒率が上がるか下がるかだけでなく、治癒後の健康状態の影響にも関心があるとする。強い薬の場合、治癒はするが、治癒しても健康を害してしまう恐れがあるからである。新たな治療法で治療を受けるか否かはＲＣＴで行ったと仮定する。すると治癒率への因果的影響はＲＣＴの仮定の下で式（１）を適用できる。だが、生存者の間で新たな治療法が健康を害したか否かの評価は簡単ではない。この場合、治癒できずに死んでしまった人々が治癒後の健康度の観察のできない「観察の中断された」人々で、当然病状の重い人が死ぬので、結果（健康度）と独立ではない。一般に新たな治療を受けても受けなくても生存できる人々は、治療群の生存者と統制群の生存者の双方に存在する。だが新たな治療法で治療を受けたから生存できたが、そうでなければ死んだであろう人々は、治療群の生存者のみに存在する。また後者の生存者はもともと前者の生存者より健康度が低い可能性が高い。つまり、例え当初の治療群と統制群への配置がランダムでも、治癒後の健康度を計測できる生存者の間では、もはや人々は治療群と統制群にランダムに配置されているとは言えなくなるのである。

「観察開始を示す仲介変数問題」にも同様の選択バイアス問題が生じる。ラロンド（Lalonde 1995）が問題にした労働市場分析での行政における失業者への職業訓練プログラムの効果を計る場合で例示しよう。職業訓練プログラムへの配置はランダムと仮定する。一般に行政者はプログラムによって就業する機会が増えるか否かとは別に、就業できた場合の所得も向上するかどうかに関心がある。職業訓練プログラムを受けたことで就業機会が増すが、就業できた場合の賃金が増さないのなら、就業訓練プログラムは、就業機会に対しシグナリング効果を持つが、人的資本育成効果はないことになるからである。逆に賃金も

向上するのなら、人的資本も向上したと考えることができる[2]。この場合の職業訓練プログラムへの配置にＲＣＴを用いれば就業率への因果的影響は式（１）で簡単に測定できる。だが就業できた場合の賃金への効果はそうではない。この場合就業できない場合が観察開始がない観察値で、当然就職後の賃金のデータはない。問題は、職業訓練を受けて就業した人と、訓練を受けずに就業できた人では、後者の方がもともと潜在的に賃金獲得能力が高い可能性があり、就業者に限ると、治療群と統制群の配置はランダムでなくなることである。

この問題はノンコンプライアンス問題とは全く異なる状況の問題となる。それはコンプライアンス問題では、割り当て変数からみたら「仲介変数」にもあたる治療変数 D を通して間接的にのみ結果に影響するという除外条件が成り立つので操作変数法を用いることができるが、本節の仲介問題では割り当て変数 Z の結果 Y への直接的影響そのものが問題で、Y の観察中断や観察開始の区別を表す仲介変数 M について、$M = 1$（生存問題の場合生存者、就業問題の場合就業獲得者）という条件の下では、もはや Z への割り当てがランダムでなくなるという点が問題だからである。もし操作変数法を用いるなら操作変数は別途見つけなければならないが、これらの問題ではそれが存外難しいのである。例えば癌の治癒には強く影響するが、治癒後の健康には全く影響しない変数は見つけるのが難しい。同様に、失業者の一定期間内での就業率には強く影響するが、就業後の賃金には影響しない変数というのも見つけるのは難しいであろう（個人が持つネットワークが前者にのみ影響しそうだが、ネットワークの豊かさと過去の賃金とには相関もあり、過去の賃金は将来の賃金と独立ではない)。

例示では結果変数が健康度や所得なので DID（difference-in-differences）が使えそうに思われるかもしれないが、これは適正な「事前」の Y が計測可能かという問題に加え、以下で説明する絶対に検証不可能な仮定を含むという問題がある。本稿では、仲介変数問題への「標準的」方法である、「無視できる割り当て」を仮定する方法を紹介するが、この問題の理解や上記の DID の問題の理解にも、潜在クラスを仮定する principal stratification の考えがここでも重要となるので、それを合わせて解説する。

まず生存問題が仲介変数となる癌治療の例で説明する。ここでも以下の 4 つの潜在クラスが理論上は可能だが実際には３つしかないと仮定するのである。４つのクラスはそれぞれフランガキスとルービン (Frangakis and Rubin 2002) に従い「Always survivors」、「Never survivors」、「Protected」、「Harmed」と名付けることにする。また以下これらを A クラス、N クラス、P クラス、H クラスとも呼ぶ。A クラスは新しい治療を受けても今までの治療を受けても癌の治癒する人々である。N クラスは新しい治療を受けても今までの治療を受け

[2] 職業訓練プログラムを受けることで、留保賃金が上がると、この解釈は正しくなくなる。また留保賃金の上昇は就業率を下げるので潜在クラスのうち、治療（この場合職業訓練）を受ければ就職できないが、治療を受けなければ逆に就職できる、という人（後述する潜在クラスの H クラス）はいないという単調性の仮定も成立しない可能性が生じる。

ても癌が治癒せず死ぬ人々である。P クラスが新しい治療では治癒するが、今までの治療では治癒せず死ぬ人々である。H クラスは逆に新しい治療では死ぬことになるが、今までの治療なら治癒する人々である。H クラスは存在しないという仮定の理由は H クラスの人が想定されるとき、この治療法の RCT による社会実験は倫理上認可されないと考えるからである。表 3 は、各潜在クラスについて各 Z の値に対応する M の値を記している。$M=1$ は治癒、$M=0$ は死亡を意味する。表 1 と同様、H クラスが存在しないという仮定は $M(z)$ に付き $M(1) \geq M(0)$ という単調性の仮定を示す。

表 3．仲介変数が観察中断の
場合の 4 つの潜在クラス

	M の値		$Z=0$ の場合の Y の観察可能性
	$Z=0$	$Z=1$	
A クラス	1	1	有り
N クラス	0	0	無し
P クラス	0	1	無し
H クラス	1	0	有り

表 3 の各潜在クラスの結果は表 1 の結果と同じである。表 3 の最後の列に $Z=0$ の場合の Y の観察可能性について（$M=1$ なら可能、$M=0$ なら不可能）について記しているが、これは以下の議論に関係するからである。

ここで重要な事実は、$M=1$（治癒者）という条件の満たすものには治療群（$Z=1$）では、A クラスと P クラスが共にいるのに対し、統制群（$Z=0$）では A クラスのみとなる点である（H クラスはいないと仮定しているので）。

さてこの問題に、DID が有効かという点だが、ここで治癒後で計測しているのは癌以外の観点から見た健康度である。従って、変化が意味を持つには治療群への割り当て以前に「癌以外の観点からの健康度」について実際には癌患者である被験者から信頼できる尺度が得られるかどうかという点が一つの問題となる。

もう一つのより理論的に重要な問題だが、これは DID の仮定に関係する。一般に DID は治療なし（$Z=0$）という条件の下で、治療群と統制群の 2 時点（以下 a と $b(>a)$ で表す）の変化、$Y_{t=b}(0)-Y_{t=a}(0)$ （ここで 0 は治療なしを意味する）は平均で変わらない、

$$E\left(Y_b(0)-Y_a(0) \mid Z=1\right)=E\left(Y_b(0)-Y_a(0) \mid Z=0\right) \qquad (30)$$

という条件が満たされることを仮定し、またこの時後述の治療群の治療効果である ATT

（Average Treatment Effect for the Treated）と一致するのだが[3]、治療群の一部である P クラスについては治療を受けない場合の時点bでの Yの値である反事実的$Y_b(0)$の値は、P クラスは仮定上治療を受けないと死亡するとされるので「存在しえない値」となり、式（３０）の仮定は（例え個人の潜在クラスを知りえたとしても）検証できない仮定となる。通常、分析が絶対に検証不能な仮定に基づくことは望ましくない。

労働市場の例でも同様の問題が起こる。まず時点aでの Y の値であるが、職業訓練前の所得は失業中だから一律に 0 とするのも潜在的賃金獲得能力が皆同じとする誤りを犯すことになるし、一方前職の賃金とするのも、継続就業者や転職者の場合と異なり、失業が介在しているので、同じ基準で賃金を測っているとは言えないという問題がある。比較可能なのは、時点a以前にも失業と再就職の経験がある場合の、その再就職時の賃金だが、そういう経験がない人（初めて失業した人）が除かれるという標本バイアスが生まれるのに加え、その「以前の失業後の再就職時」という時点が個人で異なるため、時間差の分布を治療群と統制群で標準化する（同一の分布を持つようにする）必要がある。

だがここでも DID の応用の問題は、やはり絶対に検証不可能な仮定を含む点なのである。DID は上記の式（３０）が成り立つという仮定の下で ATT と一致するが、ATT では治療群の人が、治療を受けなかった場合の反事実的 Yの値、$Y_b(0)$、について想定することになる。しかしここでも職業訓練を受ければ職が得られるが、職業訓練を受けなければ職が得られないという P クラスの場合反事実的$Y_b(0)$の値は存在しえない値となり、式（３０）は絶対に検証し得ない仮定となってしまうのである。

さて、この問題への一つの解として、後述する $M=1$ での A クラスと P クラスの決定要因に「無視できる割り当て」を仮定する方法を以下で紹介するのだが、その前に「平均治療効果」について、重要な区別を述べたい。一般に治療効果の異質性を仮定するルービンの因果モデルでは平均治療効果には「平均治療効果（ATE）」、「治療群の平均治療効果（ATT）」「統制群の平均治療効果（ATU、Average Treatment effect for the Untreated）」の区別がある。これは治療効果が個人個人で異なると仮定しているので、どの集団を母集団と考えるかによって平均が異なるからである。ATE は標本全体が代表する母集団での平均治療効果である。これは治療群の人々には「治療を受けなかったならば」という反事実的結果との差を、統制群の人には「治療を受けていたならば」という反事実的結果との差を共に想定することになる。ATT の場合は治療群の人々にのみ実際に観察された結果と、彼らがもし治療を受けていなかったならば、という反事実的結果との差の平均を推定することになる。逆に ATU の場合には統制群の人々に対し観察された結果と、もし彼らが治療を受けていたならば、という反事実的結果との差の平均を推定することになる。

通常の因果推定法ではほとんどが ATE か ATT の推定になっている。ノンコンプライ

[3] $ATT = E(Y_b(1) - Y_b(0) \mid Z = 1)$ と $DID = E(Y_b(1) - Y_a(0) \mid Z = 1) - E(Y_b(0) - Y_a(0) \mid Z = 0)$ は式（３０）が成り立つと一致する。ここで $Y(1)$は治療を受けた場合の結果。

アンス問題がない場合の平均治療効果の推定式（１）は Z がランダムなので ATE＝ATT＝ATU が成り立つ場合で、理想的な推定値である。一方式（１０）で示したように操作変数法の平均推定式は潜在クラス C クラス（Compliers）の平均治療効果というやや特殊なものになっているが、A クラスが存在しない時には ATT になることは既に述べた。一方一定条件が満たされるので D に内生性がないとする場合の式（２３）や式（２８）の推定値も ATE=ATT=ATU を満たす理想的な場合である。この ATE=ATT＝ATU は治療変数がランダムな場合にのみ成り立ち、一般には成り立たない。一方注 3 で示したように DID（Difference-in-differences）による治療効果の推定値は ATT であり、ATE ではない。

さて以上の概念レビューを行ったのは、VI 節での仲介変数 M の値が M= 1 となる場合の、IPT 法の推定値は ATU（統制群の人々がもし治療を受けていたならば、という反事実的状況を想定する場合）の推定値で、その稀な応用例となるからである。理由については技術的には以下でより詳しく述べるが、A クラスのみが M = 1 の状況で治療群にも統制群にも共通な人々で、統制群の生存者（あるいは就業獲得者）は全員仮定により A クラスなので、その人たちがもし治療を受けていたならば、という反事実的結果を推定することになり、それは ATU の推定に他ならないからである。つまり、前節 III でのノンコンプライアンス問題のある場合の平均治療効果は C クラスの平均治療効果であるのに対し、本節 IV での平均治療効果はAクラスの平均治療効果になるという、興味深い違いがあることになる。

なお、DID 同様、ATT を用いることができない（従って ATE を推定できない）理由は、治療群の人々がもし治療されなかった時という反事実的状況では、治療群に含まれる P クラスの Y が存在し得ない値になり、治療効果を定義できないためであり、その理由は DID の用いることのできない理由と同じである。なお、これらの論理は A クラス、N クラス、P クラスという 3 つの潜在クラスが存在するという仮定に依存しており、その仮定を外せば、DID に対する（したがって ATT に対する）「存在し得ない反事実的状況の Y の値に関する仮定」云々の批判も成り立たなくなる。しかしこの潜在クラスの仮定を絶対化はしないものの、以下で説明する ATU がその批判を免れるという点は、考慮に値すると筆者は考えている。

さて、「強く無視できる割り当て」の仮定であるが、一般には割り当てはランダムではないが、観察できる変数群 **V** の値にのみ依存するという仮定である。さて、A クラスと P クラスの両方を含む治療群の生存者と、A クラスのみを含む治療群の生存者の結果が比較可能になるためには、以下の条件が成り立てばよい（十分条件である)。

$$Y^1 \perp L(\mathrm{A\ vs.\ P})|\mathbf{V} \qquad (３１)$$

これは、治療を受けて治癒した場合の健康度（Y^1）は、制御変数 **V** の値が一定であれば、A クラスと P クラスの別（L(A vs. P)）には依存しない、という仮定である。この仮定が成り立てば **V** が一定なら、生存者（M=1）の治療群（A クラスと P クラスを含む）の Y の

平均の推定値は、生存者の治療群の A クラスのみの平均の推定値と同等となる。一方生存者の統制群の Y の平均値は A クラスのみの平均なので、$\mathbf{V}$ の値が一定なら観察された治療群の統制群の Y の平均の差が平均治療効果となる。またこのことから、$\mathbf{V}$ の分布が治療群と統制群で同じとなるような状況をデータ上作れば、統制群生存者（A クラス）の平均治療効果が測れることになる。

式（３１）は M=1 における、以下の「無視できる割り当て」の仮定の式（３２）と同等である。

$$Y^1 \perp Z|M=1,\mathbf{V} \qquad （３２）$$

これは通常の「強く無視できる割り当て」の仮定、$(Y^1, Y^0) \perp Z|M=1,\mathbf{V}$ 、よりはやや弱い仮定になっている。$Y^0 \perp Z|M=1,\mathbf{V}$ の仮定が省かれているからであるが、実際省かれた $Y^0 \perp Z|M=1,\mathbf{V}$ の仮定は治療を受けない場合の生存後の健康度（Y^0）が P クラスでは得られないので、DID 同様絶対に検証不可能な仮定となる。一方式（３２）の仮定はそうではない。統制群の生存者（Z= 0、M= 1）人々がもし治療を受けていたならば、という反事実的状況は、統制群の生存者が A クラスで、治療の下でも生存者と仮定されているので、この反事実的状況で Y^1 に「存在し得ない値」は含まれないからである。

一般に統制群生存者における治療効果（ATU）は

$$E(Y^1 - Y^0 \mid M=1, Z=0) = E(Y^1 \mid M=1, Z=0) - E(Y^0 \mid M=1, Z=0)$$

となるが、$E(Y^0 \mid M=1, Z=0)$ の推定値は単に標本平均なので、反事実的 $E(Y^1 \mid M=1, Z=0)$ の推定値を求めればよい。それは式（３２）の無視できる割り当ての仮定から以下の式で与えられる。

$$\begin{aligned}
E(Y^1 \mid M=1, Z=0) &= \int_{\mathbf{V}} E(Y^1 \mid M=1, Z=0, \mathbf{v}) P(\mathbf{v} \mid M=1, Z=0) d\mathbf{v} \\
&= \int_{\mathbf{V}} E(Y^1 \mid M=1, Z=1, \mathbf{v}) P(\mathbf{v} \mid M=1, Z=0) d\mathbf{v} \ \text{(from (32))} \\
&= \int_{\mathbf{V}} E(Y^1 \mid M=1, Z=1, \mathbf{v}) P(\mathbf{v} \mid M=1, Z=1) \omega(\mathbf{v}) \\
&= E_{\omega}(Y^1 \mid M=1, Z=1) \qquad (33)
\end{aligned}$$

ここで E_{ω} はウェイト付き平均を意味する。$E_{\omega}(Y^1 \mid M=1, Z=1)$ は治療群生存者の Y のウェイト付き平均で、これが「統制群に配置された人々が治療を受けていたならば」という反事実的状況での Y の平均の推定値となる。

また式（３３）のウェイト $\omega(\mathbf{v})$ は以下の式で与えられるので $P(Z=1 \mid M=1, \mathbf{v})$ に関す

るロジスティック回帰分析などで推定できる。

$$\omega(\mathbf{v})=\frac{P(\mathbf{v}\mid M=1,Z=0)}{P(\mathbf{v}\mid M=1,Z=1)}=\frac{P(Z=1\mid M=1,\mathbf{v})}{P(Z=0\mid M=1,\mathbf{v})}\frac{P(Z=0\mid M=1)}{P(Z=1\mid M=1)} \quad (34)$$

従って、統制群における治療効果は

$$E_{\omega(\mathbf{v})}(Y\mid M=1,Z=0)-E(Y^0\mid M=1,Z=0) \quad (35)$$

で与えられ、この値は治療群の Yのウェイト付き平均と統制群の Yの単純平均の差である。

なおこの仲介変数が、観察中断や観察開始を意味する場合の治療効果の推定には、上記の無視できる割り当ての代わりにより弱い仮定で治療効果の有無を検証しようとする数々の試みがあり、それらの多くも上記の単調性を含む潜在クラスの仮定に基づいている。それらの方法論には先に述べた Frangakis and Rubin (2002)に加え、Imai (2008)、Ding et al. (2011), Zhang and Rubin (2013), Yang and Small (2016)などがあるが、いまだに決定版というものはなく、本稿でそれらの方法の紹介や議論は省くが、以下それらの中から最も簡単な方法のみ紹介する。

「弱い仮定」の一つは、A クラスと P クラスの間に結果 Yに平均の差があるという仮定である。実はこの仮定は下記で説明するように、就業後の治療効果推定には有用だが、生存後の治療効果推定には有用でない。

今職業訓練を受けても受けなくても就職できた人々（A クラス）と職業訓練を受けたから就職できた人（P クラス）の間では、職業訓練を受けた後の就職後の平均賃金は A クラスの方が P クラスと比べ等しいかより大きいと仮定できるとしよう。すると

$$E(Y(1)\mid A)\geq \overline{Y}_{M=1,Z=1} \quad (36)$$

が成り立つ。式の左辺は観察されていない A クラスの職業訓練を受けた場合の就職後の平均賃金で、右辺は観察される治療群（$Z=1$）の就業獲得者（$M = 1$）の平均賃金である。前者が後者に比べ等しいか大きくなるのは、前者が A クラスのみなのに、後者は A クラスと P クラスの混合で、P クラスの賃金が A クラスより低いと仮定しているからである。一方、

$$E(Y(0)\mid A)=\overline{Y}_{M=1,Z=0} \quad (37)$$

が成り立つ。つまり職業訓練を受けないで就職したときの A クラスの平均賃金は、観察された統制群の就業獲得者の平均賃金に等しい。従ってもし $\overline{Y}_{M=1.Z=1}>\overline{Y}_{M=1,Z=0}$ が成り立てば、$E(Y(1)\mid A)>E(Y(0)\mid A)$ が成り立つ。つまり、A クラスの人々の間で職業訓練は就職後の賃金を上昇させたと結論できる。ただし $\overline{Y}_{M=1.Z=1}>\overline{Y}_{M=1,Z=0}$ はこの結論を得る十分条件であっても必要条件ではないので、$\overline{Y}_{M=1.Z=1}>\overline{Y}_{M=1,Z=0}$ が成り立たなくても、職業訓練効果が有

意でないとは結論できない。

　　しかしこの方法と論理は生存者の健康への治療効果には適応できない。生存者の治療効果問題でも新しい治療でのみ治癒する人々（P クラス）は治療法にかかわらず治癒する人々（A クラス）より健康度が劣るとすると、上記の式（３５）と式（３６）は共に成り立つ。しかしここでは強い治療が治癒患者の健康に負の影響を与えるか否かを検証したいので、仮説は $E(Y(1)\,|\,A) < E(Y(0)\,|\,A)$ が成り立つか否かである。しかし、この仮説は上記の就業後の賃金に関する仮説とは不等号の向きが反対であり、$\overline{Y}_{M=1.Z=1} < \overline{Y}_{M=1,Z=0}$ が成り立っても、検証されないのである。ただし逆に $\overline{Y}_{M=1.Z=1} > \overline{Y}_{M=1,Z=0}$ が成り立てば、新治療法は治癒後の健康をむしろ改善するという、仮説と矛盾する結果を得るが、当初の関心が新治療法は健康を害するか否かということであれば、このような結果が生じるケースはまれであろう。

　　一般に A クラスと P クラスの Y の平均の差を仮定する方法は、A 対 P の区別の仲介変数 M への影響と結果変数 Y への方向が同じ場合（例えば職業訓練の就業と賃金への影響は共に正）には有効であるが、方向が逆の場合（例えば強い治療の治癒度への影響は正で、治癒後の健康度への影響は負）には有効でないという性質を持っている。

　　以上のように、この問題には様々な考え方があるが、因果推論上、潜在クラスを考えることが、ノンコンプライアンス問題のみでなく、仲介変数によりランダム性が損なわれる場会の推定法にも方法が妥当か否かの判断の一つの基準となっていることが重要である。

V. 応用例；ノンコンプライアンス問題のある場合

　　以下の応用例は米国における教育実験でノンコンプライアンス問題が生じた例で例示する。

　１９８５年に米国テネシー州全体で行われた任意参加の小学校で「少人数教室」が教育効果を高めるか否かについての社会実験的データを用いる。この実験の詳細については Finn and Achilles (1990)に説明がある。参加校はテネシー州公立小学校で小学就学前年に付属する幼稚園時に生徒は一クラス１３－１７人の少人数クラス（治療群）、一クラス２２－２５人の通常サイズのクラス（統制群）にランダムに配置された。この段階ではランダム割り当てに対するノンコンプライアンス問題は生じていない。また幼稚園終了時の英語力テストの結果には治療群と統制群で有意な差は全くなかった。さて、この実験に参加した生徒は小学一年時にも割り当てられた少人数教室と、通常サイズ教室にそれぞれ留まることが期待されたのだが、小学校時のクラス再編成などのさまざまな理由でこれは徹底せず、下記の表４で示すように、幼稚園時のランダムな配置（Z）と小学一年での少人数クラスへの配置（D）に差が生じてしまった。なお Z および D の値が１は少人数クラス、０は通常サイズのクラスを意味する。

　　表４．人種別、治療割り当て別、治療の有無別標本数

	全標本		黒人以外		黒人	
	$D=1$	$D=0$	$D=1$	$D=0$	$D=1$	$D=0$
$Z=1$	1,293	108	913	74	380	34
$Z=0$	248	2,867	168	2,016	80	851

この表から潜在クラスの単調性を仮定すると、人種区分別の A クラス、N クラス、C クラスは以下の表 5 の割合になる。C クラスの割合に人種による統計上の有意な差はない。

表 5．人種区分別の潜在クラスの割合（%）

	C クラス	A クラス	N クラス
全体	84.3	8.0	7.7
黒人以外	84.8	7.7	7.5
黒人	83.2	8.6	8.2

さて、小学校一年修了時の英語力のテスト結果であるが、操作変数法のよる推定値とその標準誤差を統計ソフトの SPSS の二段階最小 2 乗法を用いて [4]「全体」「黒人以外」「黒人」別ににについて推定すると以下の結果を得た。

表 6．操作変数法による治療効果の推定値

		係数	標準誤差	有意度（P）
全体	切片	545.20		
	治療効果	9.00	4.30	0.036
黒人以外	切片	559.44		
	治療効果	7.70	5.23	0.141
黒人	切片	511.83		
	治療効果	11.52	6.93	0.097

[4] SPSS のコマンドは以下である。ただし Z、D、Y の変数を含む SPSS データファイルが既に読み込まれていると仮定している。また以下のシンタックスを用いずインタラクティブにも二段最小 2 乗法の変数を指定できる。

```
TSET  NEWVAR=NONE.
2SLS Y WITH D
    /INSTRUMENTS Z
    /CONSTANT.
EXECUTE
```

表６で切片は、統制群のテストの平均値の推定値である。黒人の方が平均は低いことがわかるが、問題は治療効果である。表６の結果は、全体では少人数クラスの方が小学一年度末の英語力は有意に高くなっていることを示す。一方黒人と黒人以外の別にみると、少人数クラスの効果は共に有意ではない。しかし、人種別だと標本数が減るので、操作変数法の非効率のせいで有意にならなくなった可能性がある。

従って、ブラックらの方法に従って、Dの内生性のテストを人種別に見てみることにする。まず制御変数なしに、式（１６）を $D=1$ の標本に、式（２１）を $D=0$ の標本に当てはめ Zの影響の有意性を見ると以下のようにになった。これが操作変数法でなく、単純な最小２乗法による回帰分析の結果である。

表７．ブラックらの手法による Dの内生性のテスト結果

			係数	標準誤差	有意度（P）
黒人以外	$D=1$	切片	555.18		
		Zの効果	11.54	9.49	.224
	$D=0$	切片	560.44		
		Zの効果	4.16	13.86	.764
黒人	$D=1$	切片	525.73		
		Zの効果	-2.96	10.52	.779
	$D=0$	切片	511.60		
		Zの効果	6.46	18.09	.357

表７の結果は、黒人以外の場合も黒人の場合も、$D=1$ において Yの平均は A クラスと C クラスで有意に異ならず、また $D=0$ において Yの平均は C クラスと N クラスで有意に異ならないという結果を得た。従って、Dの内生性はどちらの場合もデータ上は支持されない結果となった。従って、Yの平均について実際に治療を受けた標本の平均と受けなかった標本の平均の差を検定すればよいことになる。式（３）と（４）で述べたが、治療群と統制群で Yの分散が異なるかどうかで標準誤差は変わるが、異なるか否かも含め SPSS の平均の差の検定の T-TEST プログラムでテストできる[5]。その結果は表８に示している。

表８の結果は、治療群と統制群では Yの分散が有意に異ならない（表８の最後のコラムの P値が 0.05 より大きい）ので、どちらを用いても良いことになる（他方分散が有意

[5] SPSS のシンタックスは以下である。

```
T-TEST GROUPS=D(1 0)
  /VARIABLES=Y
  /CRITERIA=CI(.95).
```

に異なる場合は、必ず分散が異なる場合を用いる方が正しい）が、いずれにせよ少人数クラスは黒人以外では英語力を高めないが、黒人では有意に高めるという結論を得る。表6の結果と比べると、黒人の場合有意差が出たのは、Dの内生性が否定されたことにより、より精度の高い（標準誤差の小さい）推定値を得たことによることがわかる。

表8．人種別治療効果

		治療効果（Dの効果）	標準誤差	有意度（P）	分散の比の F検定
黒人以外	分散同じ	4.34	4.34	0.316	$F = 0.040$
	分散異なる	4.34	4.28	0.311	$P = 0.841$
黒人	分散同じ	11.46	5.61	0.041	$F = 1.399$
	分散異なる	11.46	5.29	0.030	$P = 0.237$

VI.　結論

RCT（randomized controlled trials）は最も単純で有力な政策評価の社会実験方法であるが、調査が理想的状況で行われるとは必ずしも言えない。本稿では、特に RCT に関連する推定の精度を高める方法（したがって調査コストを低める方法）と、治療群・統制群への割り当てがランダムでも、ノンコンプライアンス問題や、仲介変数問題が生じて、ランダム割り当ての仮定が成り立たなくなる場合についての、因果推定の基本的な考え方と、有力な方法について解説した。なお、因果推論はいまだ発展しつつある方法である。今後さらなる技術的改善もあるであろうが、基本的な考え方については、既に学者間でかなりの合意があり。本稿は特にその「考え方」の部分に着目し、わが国の EBPM、特に RCT を用いる場合に、留意すべきこと、理解しておくべきことについて、現在の理解のレベルより一段階上のレベルとなることを期待して、解説を試みたものである。

引用文献

星野崇宏。　2009.　『調査観察データの統計科学』岩波書店。

山口一男。　2017.　『働き方の男女不平等　理論と実証分析』日本経済新聞出版社。

Angrist, J.D., G.W. Imbens, D.B. Rubin. 1996. “Identification of Causal Effects Using Instrumental Variables.” *Journal of the American Statistical Association* 91:444-455.

Black, D.A., J. Joo, R. Lalondo, J.A. Smith, and E. J. Taylor . 2015. “Simple Tests for Selection Bias: Leaning More from Instrumental Variables.”　IZP DP No. 9346. University of Chicago.

Ding, P., Z. Geng, W. Yan, and X.H. Zhou. 2011. “Identifiability and Estimation of Causal Effects by Principal Stratification with Outcomes Truncated by Death.” *Journal of the American Statistical Association* 106: 1578-1591.

Flinn, J.D. and C.M. Achilles. 1990. “Answers and Questions about Class Size:

A Statewide Experiment." *American Educational Research Journal* 27: 557-577.
Frangakis C.E. and D.B. Rubin. 2002. "Principal Stratification in Causal Inference." *Biometrics* 58(1): 21-29.
Heckman, J.J. 1997. "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations." *The Journal of Human Resources* 32(3): 442-462.
Imai, K. 2008. "Sharp Bounds on the Causal Effects in Randomized Experiments with Truncation-by-death." *Statistics and Probability Letters* 78(2): 144-149.
Lalonde, R.J. 1995. "The Promise of Public Sector-Sponsored Training Programs." *The Journal of Economic Perspectives* 9(2): 149-168.
Yang, F. and Small D.S. 2016. "Using Post-outcome Measurement Information in Censoring-by-death Problem." *Journal of the Royal Statistical Society: Series B* 78(1): 299-318.
Zhang, J.L. and D.B. Rubin. 2003. "Estimation of Causal Effects via Principal Stratification When some Outcomes are Truncated by Death." *Journal of Educational and Behavioral Statistics* 28(4): 353-368.