

コントロールや無作為化は なぜ必要なのか

高木廣文，林 邦彦

コントロールや無作為化はなぜ必要なのか

高木廣文 TAKAGI Hirofumi
新潟大学医学部保健学科

林 邦彦 HAYASHI Kunihiro
群馬大学医学部保健学科

はじめに

本連載ではこれまで、科学的研究デザインとしてコホート研究^{1, 2)}、ケース・コントロール研究³⁾、介入研究⁴⁾などについて説明してきた。また、EBMやEBNの枠組みでは、無作為化臨床試験のように実験的にデザインされ、かつ対象に割り当てる処理を無作為に行う研究が最も科学的とされることなどもそのつど紹介してきた。

それでは、なぜ実験的研究や無作為化を行う研究デザインが、最も科学的といえるのだろうか。これは当たり前の疑問だと思うのだが、これまで、この点に関しては説明していなかった。この疑問に答え

るためには、何のために我々は研究を行うのかを、その基本に戻って考える必要がある。

今回は、どのようにすれば治療効果や疾患発生に関する因果推論が行えるのかを説明する。そして、因果推論のためには、どのような研究デザインが必要とされるかを、いつもよりも詳しく述べたいと思う。

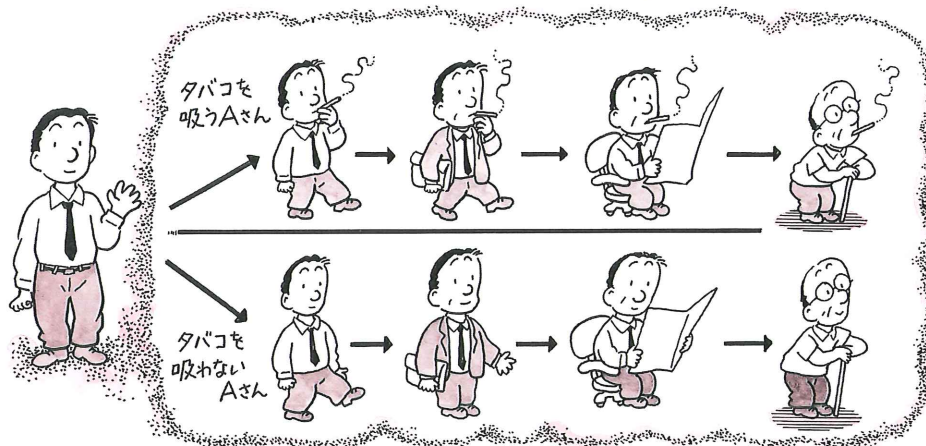
研究デザインと因果推論

喫煙が肺癌のリスク因子であることは、現在、広く知られている。しかし、喫煙していても肺癌になる人もいれば、そうでない人もいる。

カウンター・ファクチャル・モデル

タバコ以外は、まったく同じに生活を送ったとして、その後に起きたことを見てみると……

もし、タバコを吸うAさんと吸わないAさんが同時に存在するとしたら……



それでは、今、A さんに対する喫煙の影響、特に肺癌の発症の原因になるかどうかを調べるためには、どのようにすればよいのだろうか。

まず、単純な発想として、A さんにタバコを吸わせて、長期にわたって追跡調査することが考えられるだろう。A さんが肺癌にならないければ、少なくとも A さんにとって喫煙は肺癌の原因ではなかったことになる。それでは、実際に A さんが肺癌になった場合、喫煙が原因で肺癌になったといえるのだろうか。このような場合であっても、喫煙が A さんの肺癌の原因であるという仮説に対して、いくつかの反論が考えられる。たとえば、

- (1) A さんはもともと遺伝的に肺癌になりやすかった。
- (2) 肺癌の原因は A さんの喫煙習慣ではなく、自動車の排気ガスによるものである。

など、そのほかにも適当な反論はいくつでも可能である。

このような反論に対して、どのようにすれば A さんに対する喫煙と肺癌の関係を正しく評価できるだろうか。

最も単純な方法は、喫煙しない A さんを長期間追跡し、肺癌を発症するかどうかを調べ、その結果を

喫煙する A さんの結果と比べればよいだろう。このような観察研究が可能ならば、A さんが肺癌を発症するかもしれないかに関しては、以下の4通りの状況が考えられる。

- (1) 喫煙する A さんも、喫煙しない A さんも肺癌になる。
- (2) 喫煙する A さんは肺癌になり、喫煙しない A さんは肺癌にならない。
- (3) 喫煙する A さんは肺癌にならず、喫煙しない A さんは肺癌になる。
- (4) 喫煙する A さんも、喫煙しない A さんも肺癌にならない。

研究対象である A さんに関しては、2つの状況で喫煙以外はすべて完全に一致しているので、結果として起きたことは喫煙が原因であると断定できるだろう。

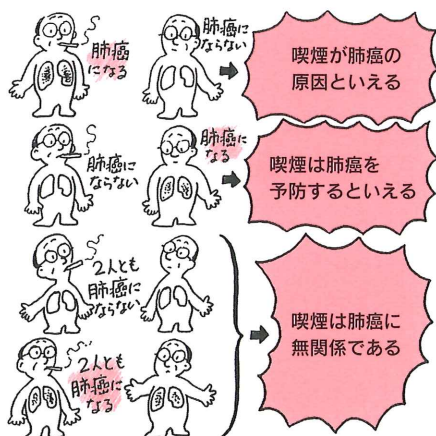
- (1) の場合は、どちらにしろ A さんは肺癌になったので、喫煙は何の影響も与えていないことになる。
- (2) の場合は、A さんの肺癌は喫煙によるものである。
- (3) の場合は、喫煙は A さんの肺癌を予防することになる。
- (4) の場合は、(1) と同様に、喫煙は A さんの肺癌には関係がないことになる。

上記のような研究を行うことができれば、A さんに対する喫煙と肺癌の関係を正確に調べることが可能である。しかし、現実には「タバコを吸う A さん」と「タバコを吸わない A さん」を同時に観察することは不可能である。このため、このような因果推論モデルは「反事実的モデル」「仮定法的モデル」「カウンター・ファクチャル・モデル (counter factual model)」などとよばれている^{5, 6)}。

コントロールの意味

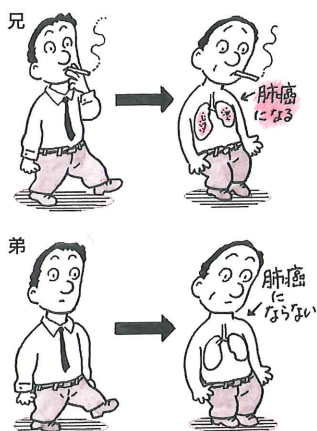
同一人物に関して、相反する処理状態を同時に観

違いは喫煙の有無のみなので、A さんに起きたことの原因としてタバコの影響だけを考えることができるはず。



一卵性双生児の場合

タバコを吸う兄と吸わない弟。兄だけ肺癌になったとしたら……。



一卵性双生児の場合、遺伝的にはまったく同一なので、(2人が同じ生活をしていれば)喫煙が肺癌の原因の可能性が高いと考えられる。(しかし、現実的には兄弟2人がまったく同じ生活を送るのは不可能であり、喫煙の影響だけを調べることには限界がある。)

測できないのであるから、どのような処理であれ、その処理の正確な因果推論は、理論的には不可能である。しかし、それでは新薬の開発、新しい治療法や術式などの評価ができず、現実問題として困ってしまうだろう。それでは、完全に正確な方法ではないにしろ、ほぼ満足できるような妥当な次善の方法はないだろうか。

先ほどの例のように、Aさんの喫煙と肺癌発症の関係を調べるための次善の方法として、実際にはどのようにすればよいのだろうか。

遺伝的な側面から考えてみると、Aさんに双生児の弟(兄でもよい)がいて、運よくタバコを吸わなければ、喫煙の影響を調べられるのではなかろうか。一卵性の双生児ならば遺伝的にはまったく同一なので、肺癌になるかならないかは、ほかの環境要因や生活習慣の相違によるものといえるだろう。また、Aさんが二卵性双生児ならば、同様の研究によって、遺伝的に完全には一致しないが、遺伝的に兄弟の近さをもつ場合について、ほかの要因の影響がわかるだろう。

このように、「双生児研究」は遺伝的影響と環境要因の影響を検討するのに、きわめて有用である。この理由は、すでに明らかなように、処理を受けた者

と受けない者の比較を行うことについて、その処理の効果に関する因果推論が妥当性をもつからである。

カウンター・ファクチャル・モデルのように、完全に同一人が処理の曝露者であり、かつ非曝露者となることはあり得ない。したがって、現実問題として我々に必要なのは、妥当性のある比較すべきコントロールということになるだろう。

ここまでの議論で明かなように、Aさんに対する喫煙の影響を調べるためには、理論的には、喫煙以外は完全に同一なAさんが必要である。実際には、完全に条件を満たすようなコントロールは存在しないだろう。仮に、この世のどこかにいたとしても、その人がどこにいるのか我々が知らない場合がほとんどだろう。そこで、我々ができることは、少なくとも年齢、性、職業などがよく似たコントロールを比較すべき者として選ぶくらいのことである。

標本は多いほどよい訳は

Aさんの喫煙と肺癌の因果関係を調べるためには、正しくはカウンター・ファクチャル・モデルによる必要があった。それでは、Bさんの喫煙と肺癌の関係を調べるにはどうすればよいだろうか。この答えも、同様に、正しくはカウンター・ファクチャル・モデルによる必要がある。

もしも、SFでお馴染みの多次元世界があり、一方の世界のAさんは喫煙者で、もう一方のAさんは非喫煙者というカウンター・ファクチャル・モデルによる調査研究が行えたでしょう。その結果、喫煙者のAさんは肺癌になったが、非喫煙者のAさんは肺癌にならなかった。Aさんの喫煙は肺癌の原因であることがわかった。ところが、喫煙していたBさんについての観察結果では、Bさんは心疾患で死亡

してしまったというものであった。すなわち、Bさんの場合、喫煙は心疾患の原因であり、肺癌の原因ではなかったことになる。

このように、個人にとっては遺伝的にも環境的にも違いがあるため、ある要因と疾患の因果関係は個別的なものである。また、心疾患で死亡してしまえば、当然、肺癌になることはない。当たり前のことであるが、このように、ヒトの生死に関しては競合的な疾患が多数、存在している。したがって、Aさんという1例のみを調べても、それはAさん個人にしか当てはまらないことになる。すなわち、このようにケース・スタディは一般性をもっていないことになる。ケース・スタディがEBNの枠組みであり重視されないのは、このような理由に基づくといっ

てよいだろう。それでは、どのようにすれば喫煙と肺癌の関係をより一般的に普遍化して研究することができるのだろうか。

Aさんは喫煙により肺癌になった、Bさんは心疾患になった、Cさんは特別の病気にならなかった、Dさんは肺癌になった……のように、多くの人について喫煙の影響を調べることが考えられる。個人については、喫煙の結果は一つしかあり得ない。それが2人の場合には対立する結果になるかもしれないし、同一の結果が得られるかもしれない。3人、4人……とその人数が増えるほど、個人の遺伝的な変動やほかの環境要因からの影響がならされていき、多くの人々からなる集団になったときには、その集団がもつ一般的な傾向、この場合ならば、喫煙による肺癌発症の頻度を明らかにすることができるだろう。この意味からすると、対象は可能なかぎり多いほうがよいことになるだろう。

さらに、喫煙の影響を正しく調べるためには、カウンター・ファクチャル・モデルでは、それぞれの人が喫煙をしなかった場合についても、同様に観察する必要がある。現実には、同一対象が同時に処理

を受けたり受けなかったりするという、二律背反の状況を観察することはできないので、適当なコントロールを設定することになる。

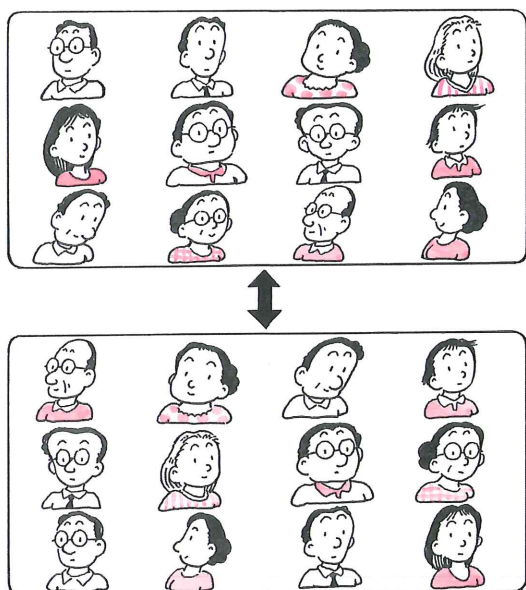
コントロールは、Aさんによく似たaさん、Bさんによく似たbさん……のように、とにかく喫煙以外はできるだけすべての特性が一致している必要がある。この操作は、ケース・コントロール研究でのマッチングに相当する。しかし、性、年齢を一致させることはできても、ほかの特性を、個々人について一致させるのはほとんど不可能に近い。ケース・コントロール研究では、ケースである各患者に対して、どのようにコントロールを設定するかが大きな問題になる。それでは、介入研究では、どのようにすれば処理群と未処理群の各対象がよく似たものになるようにできるのだろうか。

無作為化の意味

さて、実際に各1,000人の喫煙者と非喫煙者の集団を設定し、長期間の追跡研究（コホート研究）を行う場合を考えてみよう。このような場合、各対象について、ほぼ正しい因果推論を行うためには、そのコントロールが必要となる。すなわち、喫煙者のAさんに対しては、理想的には喫煙以外はその特性が完全に一致する非喫煙者aさんがマッチされている必要がある。

現実的な問題として、このようなマッチングは可能であろうか。数人ならば可能かもしれない。よく似た2人ずつのペアを探してきて、一方にある処理を行い、もう一方にはその処理を行わず、経過を観察することは可能かもしれない（ここでは、倫理的な問題に関しては考えていない）。ここで問題となるのは「よく似た」とは、どのような状況を示しているのだろうか、ということである。いろいろな解釈が可能かもしれないが、疾病の因果推論に焦点を当てて考えてみると、これは「疾病の発生」に差をもた

無作為化



一人ひとり異なるが、集団で見るとかなり似たものになる。

人は一人ひとりが異なる特性をもっているため、特性が一致した人同士のペアをつくるのは難しい。

しかし、ある程度の人を集めたグループを無作為に2つの集団に分けることにより、特性がかなり一致した集団同士のペアをつくることができる。

これを無作為割付けという。無作為割付けでは、既知の特性についてだけでなく、未知の特性についても一致させることができる。無作為割付けした集団を介入する群（介入群）としない群（対照群）に分けて試験を行ったとき、結果は介入の影響によるものとみなすことができる。

らさないように、リスク因子に関して同一条件を有している個体同士というふうに考えればよいだろう。

一般的にリスク因子となりうるのは、性、年齢はまず第一に考慮すべき要因であろう。そのほかの因子については、たとえば、飲酒状況、食物の好き嫌い、摂取状況、肥満度など、どの程度まで一致すべきかが、現実問題として重要になってくる。ペアで一致すべき特性が増えれば増えるほど、そのようなペアを探すのは困難になり、研究対象者数は十分に得られないであろう。

それでは、どのようにすれば喫煙者と非喫煙者のようなペアを、多数得ることができるだろうか。残念ながら、ペア単位、言い換えれば個々人の対象に

マッチするようなペアづくりでは、すでに述べたようにペア数を増やすのに限界がある。

カウンター・ファクチャル・モデルによる正確な因果推論をあきらめて、次善の策であるよく似た個人のペアの集団を観察するのもかなり困難であることがわかった。それでは、さらに次善の策として、よく似た集団同士を比較研究すればよいだろう。しかし、先ほどの議論と同様に、よく似た集団をつくるというのも、その集団の構成員を一人ひとり選定していたのでは、どのくらいの時間が必要かわからないだろう。

そこで、まず、よく似た人たちのグループをつくる。たとえば、健康な男性で50歳代の2,000人を集める、などである。当然、職業や飲酒状況などは個々人で相違があるだろう。そのように個人個人は異なっているとしても、その一つの集団を、よく似た2つの集団に分けることは可能である。

我々が、個人個人の年齢や飲酒状況、生活習慣などで分類することも可能であるが、そのようにすると、ほかの要因、たとえば遺伝的な要因や食習慣などに偏りが生じるかもしれない。

このような場合、我々は無作為化した選択方法を用いて、集団の構成員を2つに分けるのが無難な方法である。すなわち、コインを投げて、Aさんはコインの表なので第1グループ、Bさんは裏なので第2グループ、Cさんは……のように人為を排除した方法により、各個人の割り当てグループを決めていけばよいのである。実際には、コインやくじ引きではなく、コンピュータで乱数を発生させ、その値に従って割り当てを決めていくのである。

統計学の教科書では、これでめでたく比較すべき2グループが設定できることになっている。それでは、無作為にすると、分割した2グループの特性が本当に一致するのであろうか。残念ながら、実際には、年齢について考えてみても、無作為割付けによる2つのグループの平均年齢が完全に一致するわけではない。

特にグループの人数が50人とか100人とか、それほど大きくない場合には、無作為に行ったからといって、グループ間の特性の差がまったくないというわけではない。

それでは、何のための無作為割り当てかという問題が生じるかもしれない。実際、無作為割り当ての意味は、それを無限回繰り返した場合、各施行ごとの年齢の平均値などの全平均値（期待値）が2つのグループで一致するということを保証するためのものである。すなわち、個々の無作為化による施行に偏りがないことを保証するのではなく、同様の施行全体で平均として偏りがないことを保証するものである。したがって、無作為割り当てで2グループに分けたのだから、年齢に差はないはずであるとか、ある特性の差を検出するのに年齢調整などは必要ないという議論は、場合によっては間違いということになる。

実際には、標本数が大きくなるほど、偶然に2グループでの何かの特性の差が大きくなる確率は小さくなることがわかっている。また、現実問題として、この無作為化の方法以外のよい方法を、我々は今のところもち合わせていないのも事実である。

統計学の役割は

結局、我々が因果推論を行うための研究デザインを考えた場合、到達した結論は無作為割り当てによる比較研究ということになる。しかし、そのような無作為化の方法であっても、研究している処理以外の特性に関して、比較すべき2グループ間でその差をなくすことはできないのである。

しかし、無作為化の方法のよい点は、偶然により生じうる特性差に関して、その値がどの程度になるかを統計学的に評価することができる点である。さらに、多変量解析を用いれば、研究目的でない変数の影響を除いた推定を行うことも可能である。

研究対象の選定の段階で人為的に生じた差は、その影響の程度がどのくらいかわからないため、我々の主観によってその評価を下すしかない。しかし、偶然によって生じる差は、統計学というEBNの道具によって、客観的に評価することが可能なのである。

文献

- 1) 林邦彦, 高木廣文: EBNのための研究計画; 観察的研究 コホート研究 (1). EB NURSING 2001; 1 (2): 202-206.
- 2) 林邦彦, 高木廣文: EBNのための研究計画; 観察的研究 コホート研究 (2). EB NURSING 2001; 1 (3): 352-357.
- 3) 高木廣文, 林邦彦: ケース・コントロール研究. EB NURSING 2001; 1 (4): 474-479.
- 4) 林邦彦, 高木廣文: EBNのための研究計画; 介入研究の原理と方法. EB NURSING 2002; 2 (3): 364-370.
- 5) 椿広計, 藤田利治, 佐藤俊哉編: これからの臨床試験—医薬品の科学的評価—原理と方法. 朝倉書店; 1999. p.21-33.
- 6) Rothman KJ, Greenland S: Modern Epidemiology, 2nd ed. Philadelphia Lippincott-Raven Pub; 1998. p.49-50.