

研究デザインなくして臨床研究なし

高木 廣文

新潟大学医学部保健学科教授

1. 数を集めれば何とかなる？

科学的な研究を行うには、統計学を使った分析が必要である。統計学を使うためには、「統計をとる」という位だから、とにかく沢山の患者からデータを集めなければならないと考えていないだろうか。または、自分はもう長いこと臨床で心疾患の患者を診ているので、膨大な資料がある。このデータをまとめて、何とか研究論文の一つ、二つはできないものだろうかと考えたりしていないだろうか。

人間を対象とした科学研究で、データ処理には統計学を用いるというのは、一般的に了解されていることである。それでは何故、統計学の手法を用いれば科学的な研究と見なされるのだろうか。

TV番組の視聴率調査というのがある。今週の第一位は、瞬間視聴率35%の〇〇〇の番組でした、といったことが、TV番組で放送されたりしている。そのような番組を見ることはあっても、大部分の人達は視聴率を調べるための装置が付いたTVを見たことはないに違いない。凄く不思議な気がするが、実際には視聴率を調べる装置は、おそらく1万台のTVに1台未満程度しか付けられていないだろうと推測される。そんな少なくて視聴率が分かるのかというと、それで分かるというのが統計学の教える所である。

統計学を使う利点は、全体のほんの一部のデータから、全体について推論ができるという点にある。つまり、TVの視聴率について言えば、日本中の全家庭のTVに装置を付けなくても、数千、数百台のTVの情報で日本全体の視聴傾向が分かるのである。

ということは、常識とは逆に統計学を使うには、それほど人数を沢山集める必要はないのかと考えるかもしれない。ある意味ではその通りなのだが、そうでない場合もある。この辺りについては、何を調べたいかにもよるので、今後機会があればその都度説明していこう。

問題は、何でも数を頼りに集めればよいという訳ではない点である。TVの視聴率調査が、日本全体のTV台数に比べて遙かに少ない台数でよいのは、装置が付けられたTVが、それなりに日本の一般家庭で見られているであろう番組を代表している場合である。

例えば、すべての装置がNHKの職員の家庭にだけ付けられていたら、その結果得られた視聴率調査結果を信用するだろうか。できるだけ偏りなく、いろいろな職種や年齢層を代表するように、装置が設置されなければならないだろう。そのように対象者を選ぶための方法として、統計学では「無作為 random」に対象を選ぶことが推奨されている。

無作為とは作為のないこと、すなわち、知り合いだけとか、たまたまそこに居合わせた人のみとかを選ぶのではなく、その集団の全ての人に公平になるように選ぶ方法である。現実には、コンピュータで「乱数 random number」を発生させ、作為が入らないように対象を選ぶのが普通である。イメージとしては、くじ引きの要領で行う選出方法と思えばよいだろう。

2. どのように比較するのか

喫煙が心疾患の罹患に関係があるように思えるのだが、どのようにしたら喫煙が心疾患の原因であることを証明できるだろうか。

著者自身について、まず考えてみよう。私が現在、喫煙者である場合、そのままタバコを吸い続けて、心疾患を発症したら、それは喫煙のためとってよいのだろうか。そうではなく、大酒を毎日飲み続けているせいではないのだろうか。もしくは、我が家系は遺伝的に心疾患を発症するように運命として定められているのだろうか。本当のところは、これら全てが混ざり合った結果かもしれないのだが、どのようにすれば喫煙の影響のみを評価できるだろうか。

理論的には、喫煙する私と、喫煙しない私を追跡調査して、その結果を比較するのが最も確実である。そして、

- (1) 喫煙する私は心疾患を発症し、喫煙しない私は心疾患を発症しない。
- (2) 喫煙する、しないに係わらず、私は心疾患を発症しない。
- (3) 喫煙する、しないに係わらず、私は心疾患を発症する。
- (4) 喫煙する私は心疾患を発症しないが、喫煙しない私は心疾患を発症する。

の4通りの内、どれか1つが観察されるはずである。

(1)の場合であれば、喫煙は確かに私の心疾患の原因と言える。(2)と(3)の場合は、喫煙習慣の有無は心疾患と関係のないことが分かるだろう。(4)の場合は、逆に喫煙は私にとって心疾患を抑える働きがあることになる。

このように、同一対象について、喫煙の場合と非喫煙の場合が比較できれば、他の条件が完全に一致するので、文句なく因果関係が証明できることになる。

現実の世界では、このような二律背反の状況を同時に観測することはできないため、このような因果推論のモデルは、「counterfactual model 反事実的モデル、仮定法的モデル」と呼ばれている。

もしもSFでお馴染みの多重世界が実在し、このような状況が実際に観察できる場合はどうだろうか。私は(1)の場合であったが、Aさんは(2)、Bさんは(3)の場合かも知れない。人間には個体差が大きいし、病気は心疾患だけではないので、他の病気や事故で死んでしまうかも知れない。そう考えると、喫煙の影響を一般的に言うためには、ある程度の人数についての観察が必要だということが分かるだろう。その結果として、(1)の場合が最も多ければ、喫煙が心疾患の発生のリスクを高めるということが、私個人に対してではなく、集団に対して言えることになる。

実際には、多重世界を観測することはできない。我々はタバコを吸うか吸わないかのどちらかである。喫煙以外は全く条件が等しい比較対照などという虫の良い存在はあり得ない。それでは、どうすれば、比較ができるのだろうか。完全に同等な存在がいなければ、よく似た人間を捜して来るしかないだろう。Aさんの比較にはAさんによく似た人を、BさんにはBさんによく似た人という具合である。よく似たといっても、別に顔が似ている必要はないかも知れない。性別、年齢、職業、嗜好など、研究したい病気に関係がありそうな点を同じにしておけば、両者の比較が可能となるだろう。

このような比較対照になる人達は「コントロール control」と呼ばれている。コントロールは誰でもよいという訳ではないことが、上記の点から明らかであろう。

3. コホート研究とケース・コントロール研究

現実世界で、喫煙の心疾患発症への影響を調べるためには、喫煙者と非喫煙者のよく似たペアをある程度集める必要があった。しかし、よく考えてみると、この条件を実際にクリアするのはかなり大変である。そこで実際には、ある程度条件を予め絞ってから研究対象者を探すことになる。もともと女性の喫煙者は、男に比べて少ないので、対象を男性に限ってしまう。また、心疾患を発症するのは若年者には少ないので、50歳以上にする。あまり高齢者でも情報が収集しにくかったりするので、69歳以下にす

る、などである。このようにしても、まだ日本人全部から選ぶのは大変である。

そこで、居住地域も絞って、東京近郊のA市の男性で50～69歳の住民から、1万人を選んで研究対象とし、その後、10年間くらい対象者を追跡し、心疾患の発生状況や他疾患の状況などを観察するという調査研究を行うことがある。このような研究デザインは「コホート研究 cohort study」と呼ばれており、疫学研究の代表的な方法である。

コホート研究では、喫煙の影響を調べる場合、喫煙者と非喫煙者について、個々にペアを作るのではなく、喫煙者のグループと非喫煙者のグループに分けて、両者を比較するという方法を用いる。比較する2グループは、喫煙以外は理想的には完全に一致している必要があるが、少なくともよく似ているべきである。もしも、年齢や飲酒状況などにグループ間に相違がある場合、そのままでは比較ができないことになる。従って、そのような場合には、年齢の影響や飲酒の影響を取り除くような分析方法が必要である。そのための統計学的方法については、またの機会に触れる予定である。

コホート研究での対象者は、ある地域の全員を対象にすることもあるし、全体から一部の対象者を無作為に抽出する場合もある。統計学を使用するためには、この無作為抽出が重要である。しかし、この種の研究は膨大な資金を必要とするし、臨床研究向けではないだろう。

喫煙と心疾患の関係を調べるためには、上記のようなコホート研究が望ましいが、すでに心疾患を発症してしまった患者について、喫煙の影響を調べるにはどうしたらよいだろう。心疾患患者の中には喫煙者もいるし、非喫煙者もいるのだから、心疾患患者だけを調べても何の結論も得られないことは明らかだろう。そこで、心疾患患者と比較するために、それぞれの心疾患患者によく似た健常者を調べて、喫煙状況を比較すればよいだろう。

ある疾患をもつ患者は「ケース case」、ケースに対して比較対照となる健常者は「コントロール control」と呼ばれ、このような研究デザインは、「ケース・コントロール研究 case-control study」と呼ばれている。すでに述べたように、比較するペアであるケースとコントロールは、よく似ている必要がある。コントロールを選ぶときに、性、年齢、居住地、などをケースと一致させる手続きは「マッチング matching」と呼ばれている。

ところで、最初の話では、喫煙者と非喫煙者の比較であったが、ケース・コントロール研究では、患者と健常者の比較になっている。本来であれば、喫煙が原因であり、心疾患が結果となるべきである。時間軸に沿って観察していく中で、心疾患の発症などのイベントが発生し、それをデータとして収集するのが筋だろう。ところが、ケース・コントロール研究は、この方向が逆転しており、結果から原因を推測しなければならない。原因から結果を見るという意味で、コホート研究は「前向き prospective」であり、ケース・コントロール研究は「後ろ向き retrospective」であるといわれる。

勘違いしてはいけない点は、後ろ向き研究だからと言って、それがよくない研究デザインであるということではない。統計的な解析の効率はよいデザインである。ただし、コントロールの選定が難しく、とくに他疾患の患者をコントロールとした場合、疾患特有の偏りを、当該疾患の原因と間違えることがあり得るし、過去について調べるため、その情報の信憑性が問題になることも多い。

4. 無作為割付について

コホート研究もケース・コントロール研究も、どちらも基本的には研究対象者の自然な状態を観察して、データを収集し、結論を得るものであった。喫煙の影響を調べるのなら、より積極的に、よく似た2人のうち一方にタバコを吸わせ、もう一方にはタバコを吸わせないで、それぞれのペアのその後を観察すればよいと考えられるだろう。このような研究デザインは、実験的研究であり、よく似たペアを

作れば、かなり確実性の高い結論が得られるはずである。

しかし、喫煙のように、人間に悪影響を及ぼす可能性のあるものに関しては、倫理上から、この種の研究は禁止されている。許されているのは、よい影響があると考えられる場合だけである。従って、このような実験的な研究方法は新薬の臨床試験のための研究などでよく使われている。

高脂血症患者の心筋梗塞を防ぐために、血中コレステロール値を下げるA薬を開発したが、その効果を臨床試験で確かめたいとしよう。ある期間中、A薬のみを投与して、誰も心筋梗塞にならなかったとしても、A薬が効いたということにはならない。同時に併用した食事指導がよかったのかも知れなし、軽度な運動が効果的だったのかも知れない。

単純な比較研究として、すでに標準薬Sがあれば、A薬投与グループとS薬投与グループを比較して、A薬に何らかの優れた点があるのかを調べればよいだろう。何をもちて効果があったのかを調べるためには、研究のエンドポイントを明らかにする必要があるが、簡単のため、ここでは心筋梗塞の発生の有無を調べることにしよう。

このような研究では、研究目的や内容について説明を受け了解した（インフォームドコンセント）患者に対して、無作為に、A薬かS薬かを割り付けていくという方法をとる。この方法は、結局のところ、治療対象となっている患者グループを、構成がよく似た患者グループとして2つに無作為に分けていることになる。

実際の治療や研究では、患者がどちらのグループなのかということが分かると、患者自身への心理的影響や、医師の態度が変化することもあり得るので、医師も患者もどちらの薬を使用しているのかわからないようにする必要がある。その方法は、「二重マスク化 doubly masked」と呼ばれており、適当な研究管理者のみがその割付情報を保管するようすべきである。

5. 無作為化臨床試験が最高？

結局のところ、臨床研究で科学的に比較研究を行おうとすると、比較すべき治療方法や試験薬を無作為に患者に割り付けるという研究デザインをとる方法が、今のところ最もよいことになる。

ただし、勘違いしてはいけないのだが、人間を対象とする研究では、何でも無作為化できる訳ではない。人間に悪影響を及ぼすような要因について、実験的に無作為割付を行うことは、倫理的に許されない。従って、そのような場合には、ケース・コントロール研究が唯一の選択可能な研究デザインかも知れない。

EBNの枠組みでは、無作為化臨床試験が上位の研究デザインとして位置づけられており、ケース・コントロール研究は下位のレベルになっている。しかし、これはあくまでも、今回述べたような、科学的な因果推論の視点から言えることである。

とにかく、担当患者のデータが多数集まったから、何とかまとめられませんか、という研究は考え物であることだけは確かである。

参考文献

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。