

EBNのための研究計画

観察研究におけるバイアスを考える

林 邦彦・高木廣文

EBN のための研究計画

観察研究におけるバイアスを考える

林 邦彦 HAYASHI Kunihiro
群馬大学医学部保健学科

高木廣文 TAKAGI Hirofumi
新潟大学医学部保健学科



- データのばらつきを誤差 (error) とよび、誤差はバイアス (系統誤差) と偶然誤差の2つに大きく分けられる。
- 真の姿からの系統的なずれをバイアス (系統誤差) という。
- 偶然にばらつく部分を偶然誤差という。
- 知ろうとしている実質・真値からずれていないことを妥当性 (validity), または適切性 (relevancy) があるという。バイアスが小さな場合、妥当性が高いといえる。
- 同じものを測定すると同様な値が得られる再現性 (reproducibility), または精度 (precision) が高いことを信頼性 (reliability) があるという。偶然誤差が小さな場合、信頼性が高いといえる。
- 研究結果が、ほかの集団、状況でも真実である度合いを外的妥当性という。
- 外的妥当性を脅かすバイアスとして、想定している集団から観察対象集団をうまく抽出できないときに引き起こす「抽出バイアス」がある。
- 得られた観察の結果が研究対象集団のなかでいかに真の姿に近いかを内的妥当性という。
- 内的妥当性を脅かすバイアスに、「選択バイアス」「測定バイアス」「交絡」がある。
- 交絡のために、層化した結果と層化しない結果に矛盾が起こる典型例としてシンプソンのパラドクスが知られている。

因果性を推論する能力は、人が生きていくのに欠かせない特性のひとつである¹⁾

ケネス・ロスマン

バイアス(系統誤差)と偶然誤差

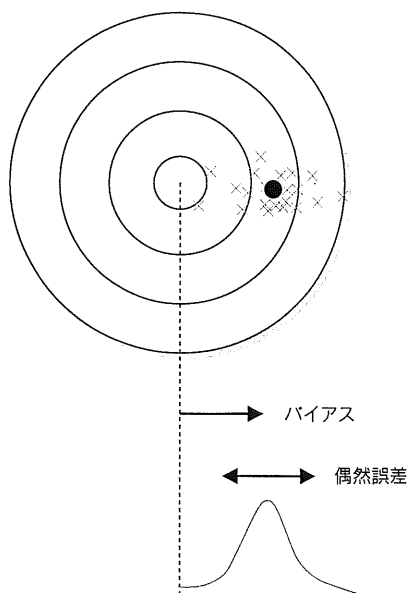
前号まで、特に、コホート研究とケース・コントロール研究とよばれる研究デザインを中心に、観察研究の特徴をみてきた。この2つの研究デザインの要点は、研究対象者には実験的な介入を一切せずに、観察と比較から因果関係を推論するところにある。

また、前号では、マンテルーヘンツェルの方法な

どデータ解析時における交絡因子の影響を除く方法も紹介した。交絡に限らず、真の姿からの系統的なずれをバイアス (系統誤差) とよぶ。バイアスについては、本誌他稿にもすでに臨床試験の事例を中心に解説されているが²⁾、本稿では観察研究でのバイアスを中心に、いま一度考えてみよう。

いま、矢、ダーツなどを、的に当てることを考えてみよう (①)。

丸い的の中心を狙って何本かの矢を射った。それらの矢の当たった跡が×で示されている。的の中心が狙いの真値で、各×が各試行での観察値、つまりデータである。高さについては的の中心から大きく



① 的の中心に狙ったのに

はずれていないが、横方向の位置については大きくはずれているものが多い。

各データが横軸方向で見て、どの位置にどれだけの個数あったか、つまり度数の分布の様子を的の下に書いてみた。的の真ん中を狙ったのにもかかわらず、個々の×の位置はばらついている。このばらつきを、誤差 (error) とよび、大きく2種類に分けられる。いずれの観察値もそろってどちらかの方向へ偏る系統的なずれの部分と、ある箇所 (この図では●あたり) を中心に、ある時は右にずれたり、ある時は左にずれたり、矢を射るたびに偶然にばらつく部分がある。前者をバイアス (系統誤差)、後者を偶然誤差とよぶ。この例の場合、左側から大変強い風が吹いていたなどのためか、系統的に右側にずれている。

この場合、的が見えているので、系統的にどの程度ずれているかはわかる。しかし、そもそも研究とか調査で何かを観察したり測定したりするのは、真値がわからないからデータをとるのであって、初め

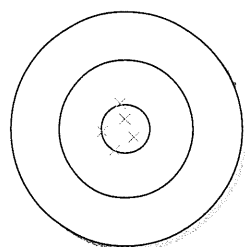
からの的の位置がわかっているのではない。①から見ると、的がない状態で×のみが観察測定され、それらの値からおそらく●あたりに真値があったのではないかと推測することになる。ということは、観察測定した後に、データから偶然誤差の大きさを推測することはできても、的がないのでデータからバイアスの大きさを直接推測するすべがない。そのため、得られたデータや結果からではなく、研究の計画デザインからどのようなバイアスが混入する可能性があるかを考えることになる。

妥当性と信頼性

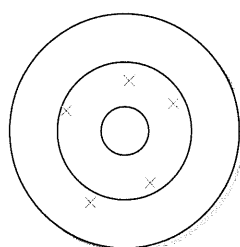
このように誤差には2種類あるが、バイアスが小さな場合には妥当性が高いという。一方、偶然誤差が小さな場合には信頼性が高いという。

先ほどの的を使って、妥当性と信頼性の関係を見たのが②である³⁾。左上のように、真の値から系統的なずれが少なく、また、偶然誤差も少ない場合には、「妥当性も高く、信頼性も高い」という。右上のように、真ん中を中心にばらつくが、その偶然誤差が大きいものを「信頼性が低い」という。逆に、左下のように、偶然誤差は小さいが、系統的なずれが大きい場合には、「信頼性はあるが妥当性が低い」ことになる。右下のような場合には「妥当性も低く、信頼性も低い」となる。

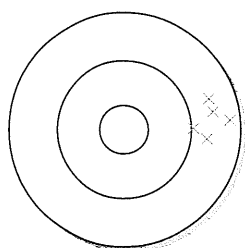
つまり、知ろうとしている実質・真値からずれていないことを、妥当性 (validity) という。もしくは適切性 (relevancy) があるという⁴⁾。それに対して、同じものを測定すると同じような値が得られる再現性 (reproducibility) が高い、精度 (precision) が高いといった概念をもつのが信頼性 (reliability) である。



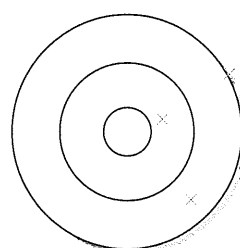
妥当性：高い 信頼性：高い



妥当性：高い 信頼性：低い



妥当性：低い 信頼性：高い



妥当性：低い 信頼性：低い

②妥当性と信頼性

バイアス分類の例

バイアスは、観察したデータからというよりは、研究デザインから見ていかなくてもいけないが、知られているバイアスの数は実に多い。「〇〇バイアス」と名のついているものはもちろん、ホーンソン効果、プラセボ効果などのように「〇〇効果」とよばれるものも、いま見ようとしている真の姿からずれを生じさせるものであればすべてバイアスであり、その名前や種類は何十とある。また、その分類の方法も何種類かあるが、ここでは前述した妥当性を「外的妥当性」と「内的妥当性」の2つに分けて考えるバイアスの分類法を紹介する⁵⁾。ごく簡便な分類法なので、観察研究でのバイアスを理解するのに参考にしてほしい。

◆外的妥当性を脅かすバイアス

その研究結果が、ほかの集団、ほかの状況でも真実である度合いを「外的妥当性」という。一般化可

能性、普遍性といってもよい。この妥当性を脅かすバイアスは、結果を一般化して利用しようとする想定集団での真の姿と、実際に観察対象となった集団で得られた結果とのギャップを引き起こすバイアスである。

想定している集団から、観察対象集団をうまく抽出できなかったという意味で、「抽出バイアス (sampling bias)」とよぶこともある。たとえば、高血圧症の患者について、治療方法とその予後との関連を知りたいとしよう。ある1つの研究における対象者は、漠然と世界中の高血圧症の患者なのではなく、当然、ある限られた医療機関において、年齢、重症度、治療方針なども限定された患者であろう。また、観察の対象

になることに同意するといった、治療に積極的な特徴をもった患者についての観察結果かもしれない。結果の一般化を考えると、自分が結果を利用している患者集団、たとえば自分が勤務している医療機関の外来患者と、その研究がなされた研究対象集団との特徴の差が大きな問題となるであろう。

特徴が似ていなければ、その結果を一般化して利用することを躊躇する、つまり外的妥当性が低いと判断するであろう。医学の研究や調査では、結果を一般化するであろう想定母集団をきちんと定義できないこともある。また、定義できたとしても、想定した母集団から全例、もしくは無作為に抽出して観察することが困難な場合も多い。そのため、外的妥当性を高めるには、異なる集団でも同様の結果が得られているか、いくつかの研究結果を比べて、その結果の一貫性をみるのがよい。

◆内的妥当性を脅かすバイアス

一方、その研究の対象として抽出された集団には問題がないのだが、比較のためのグループ化、集団の中での測定方法の不均一、注目している事柄以外の要因による影響などによって、真の姿からずれて集団が観察されてしまうこともある。得られた観察の結果が、研究対象集団の内できかに真の姿に近いものが得られたかという意味で、「内的妥当性」とよばれる。文献5)では、内的妥当性を脅かすバイアスを、「選択バイアス」「測定バイアス」「交絡」(攪乱バイアスと訳されることもある)、の3つに大別している。

「選択バイアス」(selection bias)とは、観察対象者のデータ収集時や、比較のためのグループ化のときに、研究対象と定義した集団の一部が抜け落ちてしまうことによって生ずるものである。たとえば、コホート研究として一般住民集団を対象に疾患の新規発生を追跡観察しているとしよう。最初に設定したコホート全員ではなく、追跡可能な人だけを選んで観察したとしたら、その研究から得られる結果は大きく真の姿と異なるかもしれない。もしかすると、病気のため死亡したり、入院したりした人が追跡不能になり、健康で元気な人だけを選んで追跡していることになるかもしれない。また、ケース・コントロール研究でも、同様に選択バイアスが生ずる可能性がある。特に、コントロール選択の手順で、このバイアスが混入しやすい。

次に、「測定バイアス」(measurement bias, information bias)とは、観察や測定が研究対象者の間で不均一なときに生ずる。肥満と心疾患発生の関係をみるコホート研究において、きっと関係があるだろうと考え肥満の人をより注意深く観察し、肥満者にばかり心電図などの心機能検査を繰り返したとする。心疾患の発生の観察手順が集団内で不均一となってしまう、得られた結果は、やはりバイアスの影響を受けたものになるだろう。

また、ケース・コントロール研究では、この測定バイアスに分類される有名なバイアスとして、「思い出しバイアス」(recall bias)が知られている。疾患など健康事象が発生したケースのグループと、発生しなかったコントロールのグループに、過去の要因への曝露を調査したとしよう。疾患が発生した患者グループの人たちと、発生しなかったコントロールの人たちとで、曝露の思い出しの程度が異なる。一般に、患者やその家族のほうが、様々な過去の要因への曝露を一生懸命に思い出そうとするだろう。調査する側も、ケースの人たちにより詳しくインタビューしてしまうかもしれない。そのため、ケースもコントロールも均一な思い出しになるよう、対象者全員に共通に行う質問方式などを決めたりするが、完全に排除することはなかなか難しい。

あなたならどうする

「交絡」については、前号で詳しく述べたが、いま一度復習してみよう。

因果性の推論にとって交絡がどんなに厄介なものか、具体例で考えてみる(設定場面や数字はフィクションである)。

いま、全国で標準的に利用されている看護法Sに対して、あなたが勤める医療機関では新しい看護法Nを考案した。その効果を、観察研究で評価したとしよう。同じ期間に看護法NもしくはSが実施された、合計400人にのぼる患者のデータを収集することができたので、さっそく分析してみた。③に示すように、新看護法Nでは「効果あり」の割合は60.0% (120

看護法	効果あり	効果なし	合計
新看護法N	120(60.0%)	80(40.0%)	200
標準看護法S	95(47.5%)	105(52.5%)	200

③ 400例での看護法評価の比較

年齢層	看護法	効果あり	効果なし	合計
70歳以上	新看護法N	105 (60.0%)	45 (40.0%)	150
	標準看護法S	35 (70.0%)	15 (30.0%)	50
70歳未満	新看護法N	15 (30.0%)	35 (60.0%)	50
	標準看護法S	60 (40.0%)	90 (50.0%)	150

④年齢層別の看護法評価の比較

人／200人)、今までの看護法Sでは47.5% (95人／200人)であった。カイ2乗検定⁶⁾の結果も統計学的に有意となり、この新看護法Nが従来の方法より優れていると自信満々でいえると考えた。そこで、看護専門家が集まる研究会で発表して、皆に勧めることにした。

発表にあたり、患者の年齢層で効果が異なるようだとの意見があったので、70歳以上と70歳未満の2つの年齢層に分けて、効果を比較した。その結果が④である。

70歳以上層では、新看護法Nの「効果あり」の割合は60.0%、標準看護法Sでは70.0%であった。70歳未満層では、新看護法Nの「効果あり」の割合は30.0%、標準看護法Sでは40.0%であった。いずれの看護法においても、高齢者の層のほうで効果があるようである。でも、ちょっと待て。看護法の比較では、何と、両方の年齢層とも新看護法Nが劣っているではないか。決して、足し算・引き算を間違え

たわけではない。④の両方の年齢層を加えると、③と同じ人数になる。明らかに、効果の有無には年齢層の影響が大きく、かつ70歳以上では新看護法Nが、70歳未満では標準看護法Sが多くなっていて、年齢層と看護法の間に関連ができてしまった。

さて、あなたは、どのような結論を出すだろう。新看護法Nがよいと考えるか、それとも標準看護法Sがよいと考えるか。

これは、「シンプソンンのパラドックス」とよばれるものである。できれば、前号の交絡の解説も引っ張り出して、少し考えてみてほしい。この例では、たまたま年齢が情報としてあったので、このような矛盾した結果が起きていることが判明した。しかし、予後に影響を与えそうな変数が観察されていなければ、どうなるのだろう。影響を与える変数が何なのかさえわからない場合もある。そういった変数の影響を受けずに、自分が比較したい要因の効果だけを見ることはできないだろうか。課題としている事柄（この場合は看護法の比較）をきちんと推定できるようにするため、介入研究では観察研究とは異なるバイアス対処策がある。次号では、介入研究（もしくは実験的研究）とよばれる研究デザインを紹介していく。

文献

- 1) Rothman KJ: Inferring causal connections-Habit, faith or logic? In: Rothman KJ, editor: Causal inference. MA: Epidemiology Resources Inc.; 1988.
- 2) 別府宏園: くすりとエビデンス5 バイアスに気をつけろ. EB NURSING 2002; 2(1): 102-109.
- 3) Glaser AN: High Yield Biostatistics. PE: Lippincott Williams & Wilkins; 2001.
- 4) 佐久間昭: 医学統計Q & A. 金原出版; 1987.
- 5) Fletcher RH, Fletcher SW, Wagner EH: Clinical epidemiology-the essentials. Baltimore: Williams & Wilkins; 1982. 久道茂, 清水弘之, 深尾彰訳: 臨床のための疫学. 医学書院; 1986.
- 6) 高木廣文, 林邦彦: EBNのための統計の読み方5 交絡因子の影響を除く方法. EB NURSING 2002; 2(1): 94-100.