

医学研究の方法 基本の



承認後に観察研究で効果と安全性の確認をすることを条件に、「新規製剤」を少数で試しに使って承認する仕組みが日本で動き始めています（76号のEditorial参照）。「試用」などという手法が現実のものとなり、観察研究をランダム化比較試験の代用とすることは、科学的根拠に基づく医療（EBM）の危機と言えるでしょう。

そこで、「医学研究の方法の基本とは何か」をシリーズで解説します。読者が「基本」の重要性を知ること、観察研究が悪用されないように監視し、留意し、読者自身や家族、患者の健康を守ることに、本連載を役立ててください。

目次

第1回 「二度あることは三度ある」	2
第2回 「因果関係を直接には経験できない」	4
第3回 「比較試験による差は因果関係を示す」	6
第4回 「研究の進め方、仮説から出版へ」	9
第5回 「バイアスと時間軸」 選択バイアスと情報バイアス	11
第6回 「交絡バイアスに特に注意を」	14

第1回：二度あることは三度ある

医学研究の目的は何のため？

医学の研究の目的は、実験であれ、疫学であれ、「あるものが体に作用すると、体に決まってある変化が起こる」という法則（因果関係）を見つける（証明する）こと、といってもよいでしょう。

「研究」とか「証明」というと、なにやら、専門的な話のように思えるかもしれませんが、私たちは日常的に、この法則（因果関係）の有無を無意識のうちに判断して、物事に対処しています。

二度あることは三度ある

例えば、「二度あることは三度ある」と言います。何かの後で、まれなことが一度だけ起こるのは単に偶然かもしれない。けれども、二度続けて、まれなあることが起こると、三度目があるかもしれないと思ったほうがよい、ということの意味しています。

「ない」とっていた公文書が出てきた、ということが連続して二度あると、またあるぞ、と考えたくなります。薬害エイズの時に、厚生省（当時）が「ない」とっていた文書が出てきました。

これは、まれな薬剤の害反応を考える際にも、非常に重要なヒントになります。サリドマイド胎芽症は1961年12月のレンツ報告（ドイツ）が世界で初めてとなっていますが、実は、その半年前の日本で、「二度あった」ことから「三度目」を予測して、サリドマイドを服用した妊婦の胎児がサリドマイド胎芽症であることを確かめて、その後は妊婦へのサリドマイド使用を中止した産科病院がありました。

これは「統計学」と関連があります。「統計学」は、確率を考える学問です。その統計

学を使うと、「二度あることは三度ある」は、実に適切な表現なのです（詳しい説明は次回以降に）。

ただし、ある薬剤が2人に効いたから3人目にも効くだろう、という意味ではありません。病気や症状は自然治癒やプラセボ効果がありますので、効き目に関しては、厳密な科学的評価が必要です。

時間に正確な人とルーズな人

もっと具体的な例を挙げます。何かの会で、開始時間に正確なAさんが時間通りに来ない場合には、何かトラブルがあったのか？とあなたは心配になります。しかし、いつも遅刻がちなBさんが時刻になって来なくても、あなたは、その人に何か事故が起こったかもしれないとは思わないでしょう。知らないうちに、あなたは、これまでの経験から得た「法則」を当てはめて判断しているのです。

Aさんは時間を守る人だ（という法則）を当てはめると、時間通りに来るはずだ。

Bさんは時間にルーズな人だ（という法則）を当てはめると、遅れても不思議はない、と判断しているわけです。

Aさんは、これまでに全く例外なく時間を守ったわけではなくとも、かなりの頻度で時間を守っていれば、「Aさんは時間を守る人」という法則性（因果関係）があなたの頭の中では成り立ちます。Bさんの場合も、100%ルーズというわけではなくとも、かなりの頻度で時間を守らなければ「ルーズな人」という法則性（因果関係）があなたの頭の中に出来上がります。

100%ではないとしても、ある程度の確率で言えそうなことを暗黙のうちに、ある法則（因果関係）として判定しているのです。この

ように、日常生活においては、意識するかしないかは別にして、確率を考えて判断しています。

AさんとBさんの法則を正確に知ろうと思えば、何回遅れたかをみて判断します。明らかに違う場合には10回でも違いが出るかもしれませんが、微妙な違いだと100回か200回くらい見なければ判断ができないかもしれません。

遅れた割合を見て、差があるかどうかを見るのが、統計学です。

思い込みや先入観

日常生活で、どの程度に時間を守っていれば、「Aさんは時間を守る人」となるのかに関して単純に確率だけでは判断できません。

あなたとAさんやBさんとの付き合いの程度も影響することでしょう。AさんもBさんも同じ頻度で遅れてきていても、あなたがAさんをBさんよりも好ましいと思っていれば、「Aさんは時間を守る人」で「Bさんは時間にルーズな人」と判断するかもしれません。こういうのを、「思い込み」とか「先入観」あるいは「色眼鏡でみる」などと表現しています。

統計の分野では、これを、バイアス＝偏り、という言葉で表現します。

好ましいと思っている人が時間を守らないことがあっても、何か事情があるに違いないので「Aさんは時間を守る人」となるでしょう。嫌な人なら、1回でも遅れたことがあったら、「時間を守らない人」と判断するかもしれません。あるいはAさんに対しては2～3分の遅れは、「ま、いいか」と見逃し、Bさんが1分でも遅れたら厳しく判定するということも起こります。これもバイアスです。

さらには、実はAさんのほうが遅刻がちであっても、それを公表しないことにすると、周りの人には本当のことが伝わりません。

科学で「付度」をしないために

さて、このような偏った判定や未報告が医学研究の分野で行われると、どうなるでしょう。ある治療方法や薬剤に不利な研究結果は公表しない、論文として出版しないということもあります。

承認後に観察研究で効果と安全性の確認を課すことを条件に、少数に試みてとりあえず承認する、という今回の動きで、特に重視されている「観察研究」は、結果を薬剤（の候補）に都合よく解釈しようと思えば可能であり、前述の「バイアス」が生じやすいといえます。科学的証拠の本当の姿が歪み、医学的な判断が間違ってしまうます。私たちは、それを見抜く力を持たなければなりません。

確率の判定を狂わせ、本当の判断を歪めてしまうバイアスは、どのような種類があるのか、どのように働くのか、本シリーズで丁寧に解説します。

薬剤の体への働きかたの法則

医学とは、「物が体に働いて起こる現象」の法則性を見つける学問だということをも最初に述べました。「あるもの」が薬物（の候補）の場合、「ある薬物を使うと良い結果が得られる」または、「ある薬物を使うと害がある」という法則を見つけることです。

これは、あること（良いことや害）が起こったのは、その薬物を使ったからだ、つまり、「薬物の使用」と「あること」との間には因果関係がある、とも言い換えることができるでしょう。

例えば、74号の「総説」で取り上げたプロトンポンプ阻害剤（PPI）の場合、「PPIは肺炎増加と関係ない。むしろ肺炎を少なくする」という観察研究の結果があります。本誌で検討した結果、偏り＝バイアスの処理方法を間違ったために、間違った結論になったことがわかりました。このように、本誌で取り

上げてきた記事を振り返りながら、医学研究の目的、原因と結果との因果関係をどのように検討するのかを考えていきます。

物事の因果関係を考える際に、統計学的な

確率はたいへん重要です。身近に誰もが経験している判断が、実は「確率」です。2回目で以降で詳しく解説します。

第2回 因果関係を直接には経験できない

前回は、「物が体に働いて起こることの因果関係を知ることが医学研究の目的」と述べました。今回の表題は、それと相反するような「因果関係を直接には経験できない」です。これはどういうことか。今回は、直接経験できない「因果関係」を、確率の問題としてどのように処理するのか、を解説します。日常の判断も無意識のうちに確率を使っていますが、偏見や思い込みが入りやすい面があることを前回に述べました。医学研究では、より厳密に、科学的に適切な確率を求めます。

ヒュームの問題 - 因果関係は直接経験できない

例えば、タミフルを服用したことで、その同じ人に生じた異常行動や突然死との間の因果関係について、私たちは直接経験することができません。図1に、この関係を示します。

「因果関係を私たちは直接経験することができない」ことを最初に指摘したのはスコットランドの哲学者デイビッド・ヒュームです。哲学者エマニュエル・カントは、ヒュームの著作を読んで衝撃を受けて、「独断のまどろみからたたき起こされた“Hume interrupted my dogmatic slumber”（英訳）」と3度も書いているそうです[1,2]。この問題は、「ヒュームの問題（Hume's Problem of Causation）」として、日本の高校の世界史や倫理で学ぶ内容だそうです[1]。

なるほど、100%完璧に因果関係を証明する

ことは、いかなる科学的手段をもってしても不可能なのですが、因果関係を確率（probability：蓋然性）の問題として扱うことは可能です。因果関係が本当に存在するなら100%に近づけるのではないか。ではどのようにして？

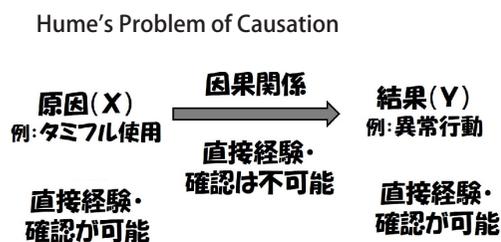
因果関係と確率

原因となりうるXがあった後で、結果となりうるYが起こった。これが、たった1回では、因果関係があるのか、偶然なのかわかりません。しかし、第1回で解説したように、まれなことでも二度続けて起こると、三度目もあるかもしれない、と思うことは、統計学的にみて正しいのです。

統計学的な言葉でいえば、「確率（probability）が大きくなる」からです。今回の解説の主眼はこの確率の問題ですが、その前に、本誌記事にも頻繁に出てくる、「P値」について簡単にご説明しておきます。

Xがあった後にYが生じたとします。「因果関係がある」ことを証明することは統計学的にはできません。そこで統計学には「XがあってもYは起こらない」、つまり「因果関係がない」との仮説を立てて、それが正解である確率

図1：原因と結果、因果関係－ヒュームの問題



(probability の頭文字 P) を計算します。これが P 値です。P 値が 0.05 (5%, または 20 回に 1 回) 未満なら仮説は間違いとみなします。言い換えると、「X があると Y が起こる」、つまり「因果関係がある」となります。

「二度あることは三度ある」を確率でいえば

2002/03 年のシーズンに、インフルエンザにかかった小児がタミフル服用後、睡眠中に突然死したという論文が小児科医向けの学術雑誌に載りました [3]。大阪府内からの報告をまとめたものです。他の薬剤、アマンタジン (抗インフルエンザ剤) あるいはテオフィリン (喘息用薬剤) を服用後に突然死した子もいましたが、薬剤を服用していなかった中で、突然死した子はいませんでした。

全国のインフルエンザ患者推定数と人口構成から、2002/03 年シーズンの 10 歳未満のインフルエンザ患者数は大阪府内で約 30 万人と推測できます。話を分かりやすくするために、このうち 10 万人がタミフルを服用し突然死が起こり、タミフルも他の薬剤も服用しない 10 万人に突然死は起こらなかった (0 人) と仮定して、両者を比較した表 1 を作りました (タミフル以外の薬剤だけを服用していた子もいるはずですが、この点は、今回は考えません)。

表 1 の左は、タミフル服用後の突然死が 1 人の場合です。P 値が >0.9999 は、「10 万人中に 1 人の突然死は、10 万人中 0 人と比べて差がない」との判断が**正解の確率**は限りなく 1 (100%) に近いという意味です (P 値は Fisher の正確法、両側検定で計算)。あくまで、確率を考えた結果です。そしてこれは、薬剤を服用していない 10 万人中 0 人も偶然であり、本当は 1 人くらいは起こるかもしれない、と考えるためです。ただし、タミフルの脳への影響など性質が分かっている場合は、1 人でも生じたら関連を指摘することが可能ですが、それはこの際、考えないことにします。

タミフル服用者中 2 人に突然死が起こると、P 値は 0.5 (50%) となり、偶然である確率が、100% から一気に半分に減ります。二度あることは三度目もありうるとの感覚と、統計学的な確率が一致しています。3 人だと P 値は 0.25 (25%)、4 人で 0.125 (12.5%)、5 人だと 0.0625 (6.25%) と偶然の可能性はさらに減ります。偶然の確率が 10% 未満になることは、逆に、90% 以上因果関係がありそうだという判断になります。

実際 2002/03 年のシーズンに大阪府で、タミフル服用後の突然死が 5 人報告されています。

4 分表または 2 × 2 表と研究方法

表 2 は、前述の数字を一般化して表現したものです。太枠で示した区画が 4 つあるので**4 分表、または 2 × 2 表**ともいいます。薬剤の服用「あり」と「なし」で 2 種類に分け、病気も「あり」と「なし」で 2 種類分けたものです。A) と B) は研究方法の違いです。

表 2 A) は、薬剤服用の有無を先に決めて、

表 1 : 二度あることは三度あるを統計学で示すと

		タミフル				タミフル	
		服用	なし			服用	なし
突然死	あり	1	0	突然死	あり	2	0
	なし	99,999	100,000		なし	99,998	100,000
	合計	100,000	100,000		合計	100,000	100,000
P>0.9999				P=0.5			

3 件で P=0.25、4 件で P=0.125、5 件で p=0.0625、6 件で 0.031。これは、あくまで純粋統計学的な確率。タミフルの性質を考慮すると 1~2 件でも因果関係は推測が可能です。

表 2 : 研究方法の基本

A) RCT/ コホート研究 B) 症例対照研究

		薬剤(曝露)				薬剤(曝露)	
		あり	なし			あり	なし
病気	あり	a	c	病気	あり	a	c
	なし	b	d		なし	b	d
	合計	a+b	c+d				

- a) イベント割合 (リスク) $Ea=a/(a+b)$ と $Ec=c/(c+d)$ を比較
→ リスク比 = (Ea/Ec) を求める。
- b) 曝露オッズ a/c と b/d を比較
→ オッズ比 = $(a/c) / (b/d) = (a \times d) / (b \times c)$ を求める

その後起こる病気を比較する方法で、ランダム化比較試験(RCT)とコホート研究で用います。コホートとは、1群の集団(コホート)のことです。その昔、軍隊の構成を決めたら、その後に入れ替えなしに戦ったそうです。疫学調査で、一旦定めた集団をコホートと称し、2つの集団(コホート)で、その後起こった病気(イベント)を比較する方法に応用したのです。

表2B)は、病気の「あり」「なし」を先に決めて、病気が起こる前に使っていた薬剤の違いを比較する方法で、症例対照研究といいます。

RCTやコホート研究にしろ、症例対照研究にしろ、「研究」では、この4分表に入るa、b、c、dの4つの数字を求めます。その数字を使って、薬剤使用とその結果との間に関連があるかどうか

かを検討します。

今回は、自然にはほとんど起こることのない、きわめてまれなことが二度起こると、三度目がありうる、ということを経験的な確率の問題として考えてみました。しかし、病気は自然にでもよく起こります。その場合に、ある原因Xで結果Yが起こった、というのは、どのようにして判断をするのでしょうか。今回は、これを考えましょう。

参考文献

- 1) 津田敏秀、意見書、タミフルと異常行動死に関する名古屋地方裁判所の判決に対する意見書、2015.4.16
- 2) Kant and Hume on Causality, Stanford Encyclopedia of Philosophy, <https://plato.stanford.edu/entries/kant-hume-causality/>
- 3) 塩見正司、インフルエンザ脳症の臨床スペクトラム、小児内科、35(10):1676-1681、2003

第3回 比較試験による差は因果関係を示す

自然にもよくある結果(Y)が、ある原因(X)で起こったという因果関係はどのように判断をするのでしょうか。

ランダム化比較試験(RCT)

XとYとの因果関係を判断するために最適な方法はランダム化比較試験(RCT)です。RCTは実験です。ある集団を構成する人のいろんな特徴(背景因子という)が同じになるように公平に2つのグループに分けます。そして、一方には試験物質(薬剤の候補X)を使い、もう一方には外形は試験物質そっくりで薬剤成分を含まない X_0 を使って、病気の治り具合や、新たな病気の発生を比較する方法です。

理想的に公平な2群に分け、患者も研究者もXか X_0 のどちらが使われたかが判らないように試験がなされている場合、両群に起こった出来事(Y、 Y_0)に統計学的な有意の差(平均値の

差や、割合の差)が認められれば、その違いは、薬剤の候補Xによって起こった、つまり「因果関係がある」と判断できます。患者に働きかけた違いは、Xと X_0 (Xの有無)だけですから。

ただし、「XがYの原因」とする結論が間違いである可能性は、0(ゼロ)にはなりません。間違える確率(P値)が20回に1回(つまり5%)よりも少ないなら、ほぼ間違いないと考えてよかろう、というのが、これまでの生物統計学の常識となっています。

α あわてもの間違いと、 β ぼんやり者の間違い

本当は関係がないのに、「XがYの原因」と結論してしまう間違いをすることを、 α (アルファ)過誤、「あわて者の間違い」といいます。

では、20回に1回ではなく、もっと厳しく100回に1回(1%)程度にするとどうなるのでしょうか。この厳しい基準を採用すると、本当

は「XでYが起こった」のに、その真実を見逃す可能性がでてきます。これを β （ベータ）過誤とか、「ぼんやり者の間違い」といいます。厳しくし過ぎるのも考えものなのです。

特に害反応を検討する場合には、発症がまれであっても重大な害は、見逃すといけない危険です。

そこで、間違う危険度がある程度大きくても、「薬剤候補Xが、生じた害Yの原因」かもしれない、と考えると、他のRCTや疫学調査をはじめ、薬剤の性質（薬理作用）や毒性試験などの情報を精査し、XとYとに因果関係があるかどうかを検討するほうが賢明です。

本当は因果関係があるのを見逃さないためには、間違う確率を20%（5回に1回）程度にしておくのがよい、とう考えかたがありますが、本誌もこの考え方を採用しています。害については、P値0.05（つまり5%）にこだわるのは危険であり、総合的な検討が大切です。

理想的でないRCTには要注意

「XがYを起こした」と結論するためには、大前提が必要です。「薬剤(X)使用の有無以外は、両群の性質が同じ」という前提です。しかも、この前提が、試験の最初から最後まで守られている必要があります。最初というのは、試験開始前、たとえば対照群に何を選ぶかの決定も含めます。最後までとは、観察終了だけでなく、公表されるまでも含めます。

一見、RCTの体裁をとっていても、実は「対照群が何も使っていない」とはいえないとか、「両群の性質が同じ」とはいえない研究が多いのです。

コクラン共同計画のシステマティックレビュー（総当たり調査）では、収集したRCTが、「薬剤(X)使用の有無(XかX₀か)以外は両群の性質が全く同じ」といえるかどうか、偏り(バイアス)の入り込む可能性がないかどうかを、徹底的に調べます。ただし、コクランといえど

も不完全なものが少なくありません(註)。

註：RCTにも偏り(バイアス)はありうる。まして、観察研究では、バイアスを除く工夫がなければ信頼できない。バイアスについては、次回以降に解説する。

コホート研究はもともと分け方が不公平

コホート研究も薬剤を使ったか使わないかで2群に分けますが、薬剤を使うか使わないか(使用の有無)以外の背景因子には大きな違いがあるため、そもそも出発点から、不公平です。

例えば、胃酸を抑制するプロトンポンプ阻害剤(PPI)を服用する人は服用しない人より、①胃酸の分泌が活発、②胃潰瘍・十二指腸潰瘍がある、③ストレスが強い、④喫煙が多い、といった性質があります。

そこで、公平に比べるためには背景因子をできるだけ揃える工夫が必要です。まず、胃潰瘍・十二指腸潰瘍がある人を選びます。その中でPPIを使った人と使わなかった人とで、例えば、肺炎に罹りやすいかどうかを比較します。

背景因子として、健康状態をそろえる工夫が必要です。性や年齢はもちろん、他の薬剤の使用頻度など微妙な違いを、解析の際には調整する必要があります。

症例-対照研究では何をそろえるか？

症例対照研究では、病気の有無以外の性質-背景因子をできる限り揃えるように2群を設けます。

まず、病気のある人を選びます。例えば、PPIと肺炎との関係を調べる場合は、「肺炎になって入院した人」を選びます。次に、肺炎で入院した人と同時期に肺炎以外の病気で入院した人の中から、肺炎で入院した人と性が同じで、年齢が5歳以内(症例の年齢が62歳なら、60歳から64歳)の人を選びます。

そうして、過去に服用してきた薬剤を比較します。抗がん剤やステロイド剤を服用していると肺炎になりやすいので、そのような、はっきり

りと肺炎になりやすい薬剤を用いている人は、症例からも対照からも除いておくほうが、PPIだけの影響を調べるためには必要でしょう。もしも、除いていない場合には、肺炎が起こったのはPPIの服用が原因か、抗がん剤やステロイド剤の影響であったのかがわからなくなるからです。

症例シリーズはさらに質が悪い

症例だけを集めて検討する方法を症例シリーズまたは「サーベイランス」といいます。例えば、インフルエンザで入院した例ばかりを集めて検討する方法です。この方法では、ある薬剤(例：タミフル)を服用した人も服用しない人も含まれ、また、軽症例や重症例や死亡する人も含まれます。

そのため、「薬剤服用ありーなし」で分けて、死亡の割合を比較するコホートの解析や、死亡のありーなしで分けて、タミフルの使用割合を比較する症例対照研究的な分析も可能です。しかし、計画されたコホート研究や、症例対照研究よりも、背景因子をそろえることは難しく、適切な比較は困難です。

「リアルワールド・データ」と称する調査は、この手法によるものが多く、信頼できません。タミフル服用後に突然死が増えるという適切に処理された比較調査の結果は、動物実験の結果と一致していて信頼できます。しかし症例シリーズが根拠の「タミフル使用が死亡減少につながる」という結論は、信頼できません。

すべて確率によって判断する

表3を見てください。RCTまたはコホート研究と、症例-対照研究を同じ表の中に示しました。縦にみれば、**A)** ランダム化比較試験(RCT)やコホート研究、横にみれば**B)** 症例-対照研究です。

適切に選択された対象を用いて、適切に収集された情報を、背景をそろえてきちんと比較し

て、関連性(因果関係)の確率を計算します。

表3：RCTとコホート研究、症例対照研究の方法

A) RCT・コホート研究 ↓

		薬剤(曝露)	
		あり(X)	なし(X ₀)
病気	あり(Y)	a	c
	なし(Y ₀)	b	d
合計		a+b	c+d

B) 症例-対照研究 →

A) RCTまたはコホート研究：薬剤あり群(X)となし群(対照群：X₀)とでイベントの割合(これをリスクと言いいEで表わす)求め、その比(リスク比)を求める。それぞれのリスクは、

あり群Ea=a ÷ (a+b)、対照群Ec=c ÷ (c+d)

リスク比=Ea ÷ Ec (Ea/Ecと表す)

B) 症例-対照研究：症例(病気ありY)と対照(病気なしY₀)で薬剤曝露のオッズ(薬剤あり÷なし)：a ÷ c、b ÷ dを求め、

その比(オッズ比)を計算=(a ÷ c) ÷ (b ÷ d) = (a × d) ÷ (b × c)

※なお、EBMで用いる主な計算式や指標を以下に示しましたので参考にしてください。

付表：EBMで用いる主な用語と計算式

A) ランダム化比較試験(RCT)、コホート研究

1. リスク：Ea=a / (a+b), Ec=c / (c+d)
2. リスク比(相対リスク)=Ea / Ec
3. 絶対リスク減少(ARR) = Ec - Ea
4. 相対リスク減少(RRR)=(Ec - Ea) / Ec
5. NNTB= 何人に1人に利益か=1 / ARR
6. NNTH= 何人に1人に害か= -1 / ARR

B) 症例-対照研究

1. 症例の曝露オッズ=a / c
対照の曝露オッズ=b / d
2. オッズ比=(a/c) / (b/d) = (a×d) / (b×c)

C) 両者共通

1. 95%信頼区間：95%の確率で推定値がとりうる範囲。その範囲に「無関係を示す1.0が含まれる」と有意でない、「含まれなければ有意」と判断する。
2. p値：まず、「原因と結果は無関係」との仮説を立てる(これが「帰無仮説」)。実際に観察された結果が偶然に起こりうる確率(p値)を計算し、0.05未満なら観察結果は偶然でないと考え、帰無仮説を棄却し、対立仮説「関連あり」を採用する。
3. 不一致度(Inconsistency: I²)：同一目的の複数のRCTどうし、複数のコホート研究どうし、複数の症例対照研究どうしで、調査結果(リスク比やオッズ比)が異質かどうかを%で示した指標。小さいほど一致し、大きいほど不一致を意味する。コクラン・ハンドブックでは、おおむね30%(40%)を超えると不一致とされる。

第4回 研究の進め方、仮説から出版へ

前回は、自然にもよくある結果 (Y) が、ある原因 (X) で起こったという因果関係を判定する方法として、ランダム化比較試験、コホート研究、症例 - 対照研究についてみました。

今回は、これらの研究を具体的に進める作業、つまり、仮説の設定から論文の出版までを示し、結果に誤りをもたらしやすいバイアス (偏り、または系統エラー) がどの段階で入りこむのかを考えます。

誤差とバイアスについて

「誤差」という言葉も、統計学的な用語としてしばしば出てきます。偏りとどう違うのでしょうか。図2を見ながら考えましょう。

誤差はランダムエラーとも言います。つまり、偶然に起こりうる真の推定値からのずれ (差) です。偶然ですから、コインの裏と表が出る確率は、多数回行くと 50% 対 50% に近づいてきますが、2 回や 3 回では、すべて表ということもあり得ます。少数例では真の推定値がえられませんが、多数例を集めれば集めるほど、平均すると、真の推定値に近づいていきます。

一方、バイアスは、系統的エラーとも言われます。コインの表に重りをつけておくと、裏が出やすくなって、多数回では 60% 対 40% になったりします。つまり、いくら多数を集めて平均しても、真の推定値 (50% 対 50%) に近づくことはできません。図2の上の段は、バイアスがない例で、下の段が、バイアスのある例です。

疫学研究の具体的作業とバイアス

図3に、疫学研究における具体的な作

業と、それぞれの段階で入りうるバイアスの可能性を示します。

1. 仮説を設定する

医学研究一般についていえることですが、研究では、まず仮説を立てます。「ある物質が、ある病気の原因になっている」「ある物質を使用すると、ある病気を少なくすることができる」などです。ここで、すこし混乱しそうなのですが、統計学的な仮説の立て方はすこし異なるということについて触れておきます。

図2:ランダムエラー (誤差) と系統エラー (バイアス)

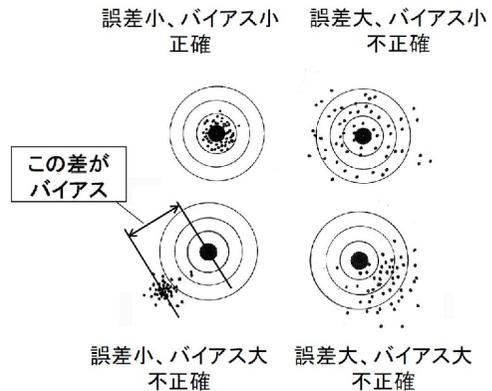
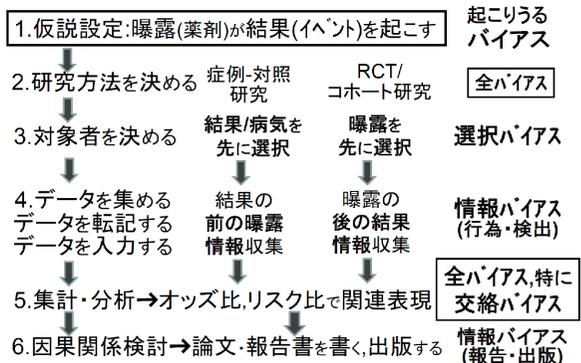


図3: 疫学研究の具体的作業とバイアス



*統計学には「要因が結果と無関係」という仮説 (帰無仮説) を立てて、その仮説が統計学的に棄却された場合に、その反対の仮説 (対立仮説) 「要因が結果と関連あり」を採用する。

統計学的には、実際の仮説とは逆に、「要因が結果と無関係」という仮説を立てます。これを「帰無仮説」といいます。そして、現実に収集されたデータは、「要因が結果と無関係」という仮説のもとで起こりうる確率を計算して、きわめてまれにしか起こりえないなら、「帰無仮説」を否定し、反対の仮説（対立仮説）、つまり「要因が結果と関連あり」を採用します。「帰無仮説」を否定することを、統計学的には「棄却する」といいます。

2. 研究方法を決める

研究方法として、ランダム化比較試験、コホート研究、症例-対照研究のいずれかを選びます。それぞれの特徴は、疫学研究の教科書にゆづりますが、大ざっぱに分けると、次のようなことになります。

ランダム化比較試験（RCT）は、効果や安全性が未知の物質が人で効果があるか、害がないかどうかを、最も少数の人で適切に知ることができる実験です。

コホート研究は、ある程度の頻度が見込まれる害作用が疑われる場合に、その因果関係を調べます。

症例-対照研究は、まれな害反応が疑われる場合に、因果関係を確認する方法です。

これらの研究を実施するために、研究計画を立てますが、その際に各種のバイアスが入りうる余地があります。特に、交絡バイアスが入りうる点に注意が必要です。

次項で対象者の決め方、次々項で、情報の集め方について簡単に触れておきます（前回と重複しますが）。

3. 対象者を決める

RCTやコホート研究では、原因としての「薬剤」を使用する人の群と、使用しない人の群を選びます。症例-対照研究では、結果としての「病気」を持つ人の群（症例群）と、症例と、同性・

同年齢で病気を持たない人を「対照群」として選びます。この際、選び方が不適切だと偏り（バイアス）が生じます。これが、選択バイアスです。RCTでも、完全に公平に2群に分けられている保証がない場合には、選択バイアスの危険性が大きくなります。

4. データを集める・転記する・入力する

対象者が決まると次に、因果関係の検討に必要なデータを集め、転記し、入力をします。

RCTやコホート研究では、薬剤曝露以降に、対象者に起こった出来事（イベント、結果、アウトカムなどという）に関する情報を収集・転記・入力します。

一方、症例-対照研究では、結果としての病気が起こる前に服用した曝露情報を収集・転記・入力します。この情報収集に際して起こる偏り（バイアス）が情報バイアスです。RCTでも、二重遮蔽が何らかの理由で崩れている場合、薬剤を使用した人の行為（行動）や、観察する医師による病気の検出力への影響が出るので、それぞれを行為バイアス、検出バイアスなど呼びます。症例-対照研究では、病気のある人の方が薬剤使用の記憶がよく残っているので、「思い出しバイアス」と呼ばれます。しかし、いずれにしてもこれらは情報バイアスです。

5. データの集計・分析→オッズ比、リスク比等で関連表現

集められ入力されたデータを集計し、分析をします。その際に、コホート研究で、曝露の前に起こったイベント（病気や症状）をイベントとしたり、症例-対照研究で、病気の起こった後に用いた薬剤を薬剤使用に入れたりする分類の誤りが結構起こります。

例えば、HPVワクチン接種後の症状を無関係とした鈴木らによる名古屋調査では、曝露前のイベントも接種後の症状として誤分類されていました。接種前に起こったイベントはワクチ

ン接種とは無関係ですから明らかに誤分類です。接種群と非接種群の選択の間違いですから、選択バイアスの一種といえます。

このデータの集計・分析で最も重大なバイアスは、交絡バイアスが入り込む余地です（詳しくは第6回）。健康状態が悪い人はワクチンを接種しない人が多いために、接種群にはもともと健康な人、非接種群にはもともと病気がちの人が多くという、大きな偏りがあります。接種率が90%近くにもなると、この影響は、たいへん大きいので、この影響を取り除く工夫が、解析の際には求められます。逆に、ワクチンの効果をよく見せ、ことさら安全性を強調したい人は、この影響（交絡バイアス）を考慮せずに解析していますので、注意が必要です。

6. 結果の記載、因果関係の検討、執筆、出版

以上の解析結果を記載し、各種バイアスの影

響（時間性とバイアス排除）を考慮して関連の有無を検討し、他の報告との関連が一致するかどうかを比較し（関連の一致性）、関連の強さを評価し（関連の強固性）、臨床・疫学的な知見以外の実験結果や毒性試験、薬物動態などの結果と矛盾しないかどうか（整合性）などを検討して、因果関係の有無を考察します（詳細は次回）。これらを記載した論文・報告書を作成し、出版します。

報告書作成の過程で、結果の一部、特に重要な結果が報告されないことがあり、これを報告バイアスといいます。また、結果が意図したものでない、あるいはメーカーにとって不都合な結果は論文・報告書として出版しないことが起こります。これを出版バイアスといいます。

いずれも、現実には私たちが利用できる情報の一部が歪んでいるために、まさしく「バイアス」と言えます。

第5回 バイアスと時間軸 選択バイアスと情報バイアス

前回は、仮説の設定からデータ収集・解析、論文執筆・出版までを示しながら、結果を誤らせるバイアス（偏り）が、どの段階で入るかを考えました。今回は、バイアスの種類と、そのバイアスが時間軸のどこで入るのか、選択バイアスと情報バイアスを中心に考えます。

研究方法別にみた選択バイアスと情報バイアス

最も基本的なバイアスは、選択バイアス、情報バイアス、交絡バイアスです。研究方法によって、選択の内容、情報の内容が異なりますので、混乱しないために、整理しておきます。

図4（次頁）に、研究方法別にみたバイアスを示します。また図5（次頁）に、医学研究でバイアスが起るタイミングを、薬剤の使用時間や

結果が起こる時間との関係で示します。図4と図5を見比べながら考えましょう。

A. ランダム化比較試験（RCT）やコホート研究では、薬剤の有無で群分け（対象薬剤の有無の選択）を先にしたうえで、起こった病気の情報（結果に関する情報）を比較します（図4）。

B. 症例 - 対照研究では、病気の有無で群分け（対象疾患の有無の選択）を先にしたうえで、病気が起こる前に使用した薬剤（曝露薬剤の情報）を比較します（図4）。

したがって、**A)**RCTやコホート研究では、対象薬剤の有無に関して選択バイアスがかかり、起こった病気に関して情報バイアスがかかります（図4）。一方、**B)**症例 - 対照研究では、問題とする疾患の有無に関して選択バイアスがかかり、

図4：研究方法別にみた起こりうるバイアス

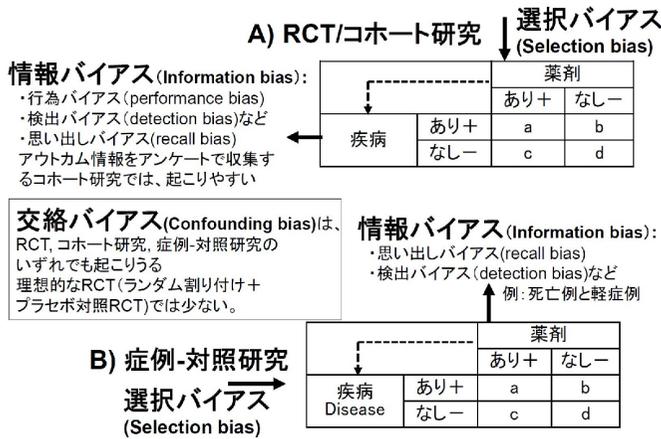
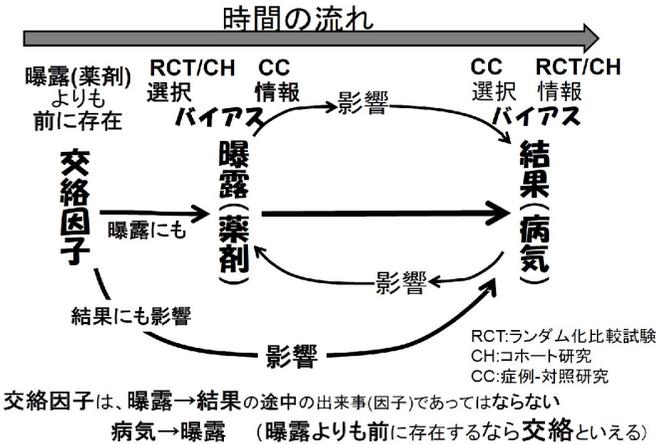


図5：疫学研究におけるバイアスと時間軸



疾患が起こる前に使用した薬剤に関して情報バイアスがかかります(図4)。選択バイアスは対象集団の歪みを意味し、標本歪みバイアスともよばれます。

交絡バイアスについて

交絡バイアスとは、Rothman[1]によれば、

1. その疾患に対する外因性の危険因子である。
2. 調査集団において曝露に関連している。
3. 交絡因子は、曝露因子や病気に影響を受けていない。曝露と疾患との間の因果経路における途中の状態ではない。曝露や疾病の発症に先行する因子なら、この基準は自動的

に満たされる。

とあります。図5で説明すると、交絡要因は、曝露(ここでは薬剤)とイベント(何らかの症状や病気)の前から存在し(Rothmanの3)、薬剤の使用(同2)にも、病気の発症(同1)にも影響する要因です(具体例は次回に述べます)。

選択バイアス・情報バイアスの例

A 1. ランダム化比較試験 (RCT)

RCTでは、薬剤が使用される前に対象者が選ばれ、薬剤が使用されるかどうかで、2群に分けられます。その群分けが偏りなく(バイアスがなく)公平に行われたら問題はありません。これを公平に割り振る作業がランダム化の手続きです。コンピュータで乱数を発生させるなど、基本的に忠実にランダム化されたRCTには、選択バイアスは基本的にないと言えます。

一方、公平に割り付けされ、選択バイアスはなかったとしても、対象となった人や医師が、問題の薬剤(の候補)を用いたか、用い

なかったかを知っていると、対象者や医師は、そのつもりで行動しがちになるため、病気の現れ方、つまり結果の情報に影響します。これは情報バイアスの一つで、行為バイアス(performance bias)と呼ばれます。

例えば、胃・十二指腸潰瘍(以下、胃潰瘍)に抗潰瘍剤のプロトンポンプ阻害剤(PPI)が使われたかプラセボが使われたかを知っていると、腹痛を訴える(行為)の割合は、どちらが使われたかを知らない場合よりも、プラセボ群でより大きく、PPI群でより小さくなるでしょう。医師が、患者に症状を尋ねる行為にも、先入観が入りやすくなります。これが行為バイア

ス (performance bias) です。

また、ある薬剤を使用後に起こった病気を評価する人が、薬剤 (の候補) を用いたか、用いなかかったかを知っていると、その病気を検出し評価する情報に偏りが生じます。これも情報バイアスの一つで、検出バイアス (detection bias) です。

胃潰瘍に PPI を使用した場合でいえば、内視鏡で胃潰瘍の治癒状態の判定情報を得る (検出する) 際にも、PPI の使用の有無が分かっていると、PPI 群の治癒率が高めに出ることになります。薬剤に甘くしようという意図の有無にかかわらずこのバイアスは起こりえます。

A 2. コホート研究

コホート研究でも、RCT と同じく、基本的には薬剤の使用の有無で 2 群に分けて、その後起こる病気 (アウトカム) を比較します。しかし、RCT と根本的に異なる点は、薬剤を使う人にはもともと何らかの病気のある人が含まれる一方、その薬剤を使ってはいけない人 (禁忌) には使われないなど、分け方は公平ではありません。そのために選択バイアスがよく起こります。そして、対象者も処方した医師も、病気を検出する際に、使用した薬剤の有無は分かっているので、行為バイアスや検出バイアスがよくかかります。

例えば、非ステロイド抗炎症剤 (NSAID) の使用者と非使用者で胃潰瘍の頻度を比較するコホート研究を考えてみましょう。

胃潰瘍の既往のある人には、NSAID は処方されにくいという選択が、医師の処方段階で働きます (選択バイアス)。しかし、NSAID が処方されると、患者も胃痛や潰瘍が起こりやすくなることを知っているため胃痛を訴える行為が増え、医師はそれに応えて内視鏡をする行為が増えるでしょう (行為バイアス)。そのために潰瘍が検出される機会は増えます (検出バイアス)。

コホート研究では、研究期間内のいずれかの時点で薬剤を使った人を使用群とし、期間中全く使わなかった人を非使用群として解析がなさ

れることがあります。しかし、使用群の人でも、薬剤を開始するまでは非使用期間として解析が必要です。使用・不使用の選択間違い (誤分類による標本の歪み) のために使用群のイベント発生率が本来よりも少なくなり、評価が (有効性も安全性も) 薬剤に有利になります。これも選択バイアスの一種で、無イベント時間バイアス (immortal time bias) と呼ばれます。

B . 症例 - 対照研究

症例 - 対照研究では、病気の有無の選択の際に、選択バイアスが起り、病気が起こる前に使用した薬剤情報を収集する際に情報バイアスが生じます。

具体的には、例えば NSAID を使っていて腹痛のある人は、NSAID を使わず腹痛のある人よりも胃潰瘍の危険性が考えられて、医師は内視鏡検査を奨め実施するため潰瘍がよく発見されます。そのため、胃潰瘍がある人を症例に選び、胃潰瘍のない人を対照に選んで症例 - 対照研究をすると、胃潰瘍に対する NSAID の関与が過大評価されることになります。対象集団の胃潰瘍の有無の頻度に歪みが生じているため、選択バイアスといえます。

病気になる前に使用した薬剤の情報を患者から収集する際には、病気のある人のほうがより多く思い出します。病気が重症ほど、よく思い出します。これは、症例 - 対照研究における典型的な情報バイアスの一つで、思い出しバイアス (recall bias) と呼ばれます。患者の記憶に頼るのではなくて、カルテ情報に基づいて情報を収集することで、このバイアスを少なくすることができます。

情報を収集する時にインタビューする人が消化性潰瘍のある人には詳しく聞き、対照群の人には詳しく聞かない場合にかかるバイアスは検出バイアスと言えるでしょう。

参考文献

- 1) Rothman ら、Modern Epidemiology、2008

第6回 交絡バイアスに特に注意を

第5回では、バイアスの種類には基本的に、選択バイアスと情報バイアス、それに交絡バイアスがあり、そのバイアスが時間軸のどこで入るのか、選択バイアスと情報バイアスの具体例を示して考えました。最終回は、特に注意が必要な交絡バイアスの具体例について述べます。

交絡バイアスの時間軸

薬剤の使用と、結果が発生する時間的な関係に焦点を当てて、バイアスが入るタイミングを、もう一度、第5回の図5(12頁)を見ながら振り返ってください。

交絡バイアスとは、Rothman[1]によれば、

1. その疾患に対する外因性の危険因子である。
2. 調査集団において曝露に関連している。
3. 交絡因子は、曝露や病気の影響を受けていない。曝露と病気の因果経路における途中の状態ではない。曝露や疾病の発症に先行するならば、3.の基準は自動的に満たされる。

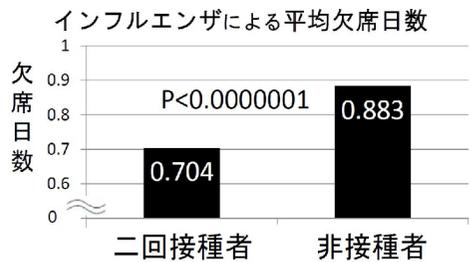
です。交絡因子は、曝露と結果(病気)が起こる前から存在して(Rothman 3番)、薬剤の使用(同2番)にも、病気の発症(同1番)にも影響のある対象集団の性質です。ここまでが前回のおさらいです。

ワクチン接種では交絡バイアスが必発

例えば、インフルエンザワクチンの効果を比較する際、インフルエンザによる平均欠席日数を、2回接種した児と接種しなかった児を単純に比較すると図6のように、P値は1億分の1未満という確実な差となります[2]。

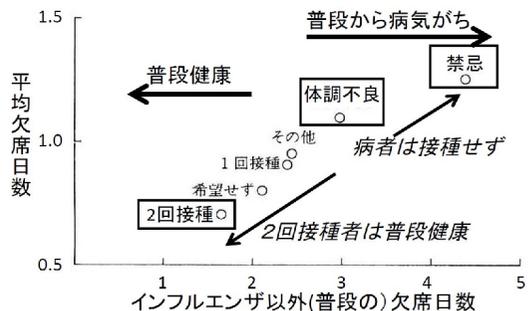
しかし、もともと2回接種した(できた)児はふだんから元気なので、インフルエンザにもかかりにくい。一方、当日熱があればワクチンを打ちませんし、アレルギーのある児も禁忌なので打ちません。そのような児は、ふだんからかぜにかかって休みがちで、インフルエンザにか

図6: 交絡バイアスを無視すると「ワクチンが効く」



2回接種群と非接種群との比較：
 t 検定の $t = 5.73$ ；Wilcoxonの順位和検定 $Z = 5.76$
 いずれも両側検定、 $p < 0.0000001$ である。

図7: 普段の健康が、接種と結果に交絡



普段から接種児は欠席少なく、非接種児は欠席日数が多い。これを調整すると差がなくなる。図1は「ワクチンが効く」証拠にならない。図1、2とも丹後俊郎著「医学への統計学」[2]から作図。

かりやすい(図7)。

これは、ふだん健康か否かが、ワクチンの接種の有無と、インフルエンザにかかるかどうかという結果に大きく影響する、つまり交絡しているためです[2]。これを「健康者使用バイアス」といいます。図7のように、ふだんの健康状態と調整すると、インフルエンザワクチンは無効

となり、図6の結果が事実と異なることが分かります。

このようにインフルエンザワクチン自体は無効であるのに、残念ながら、ワクチンがインフルエンザを予防した、というデータを国も採用して、それを根拠に接種が勧められてきました[3]。

交絡バイアスは害の評価にも関係する

図6と図7で示したことは、ワクチン全般について言えることなので、図8に一般化して示しました。交絡バイアスは、効果だけでなく、害についても、ふだんの健康状態が大きく交絡して、「効く」、あるいは「安全」であるかのように見せ、結果をゆがめるので、この点に注意しておいてください。

HPV ワクチンの害の評価がその典型です。名古屋市の HPV ワクチンに関するアンケート調査を詳しく分析したところ、HPV ワクチンを接種しなかった人は接種した人より、元々いろんな症状の訴えが多かったことがわかりました。

例えば、「すぐ疲れる」という訴えは、非接種者が接種者の約3割増し、「なかなか眠れない」との訴えは約5割増し、目の症状（視野の異常や視力の低下）は6割増し、「手や足に力が入らない」は8割増しでした。「杖や車いすが必要になった」という症状のために受診したことのある人は4.9倍いました[4]。このため、接種者では、非接種者よりも4.9倍多く起こってはじめて、非接種者と同じになります。

実際には、「杖や車いすが必要になった」ために受診した人は、接種しなかった人の9.6倍いましたので、本当は、HPVの接種でワクチン $4.9 \times 9.6 = 47$ 倍起こりやすくなると推定されます[4]。

このことを考慮しないで単純に比較すると、インフルエンザワクチンの場合の、図

6のように、真実とはかけ離れた結論になるのです。

このバイアスを無視して解析すると変な結果が出るのは、次のことから分かります。

HPV ワクチンを接種しても、神経の病気が少なくなるなどということはありませんし、子宮頸がんがん以外のがんは減る可能性があるかもしれませんが、他の部位のがんが減ることはありません。ところが、神経の病気が少なくなったとか、他の部位のがんまで少なくなったとの報告がでてくるのは、この健康者使用バイアスという交絡バイアスを無視した結果です。

治療用薬剤の適応による交絡

一般に薬剤は、何らかの病気の症状を和らげ、合併症を予防するために用います。そして、病

図8：交絡の例(1) 健康者接種交絡バイアス

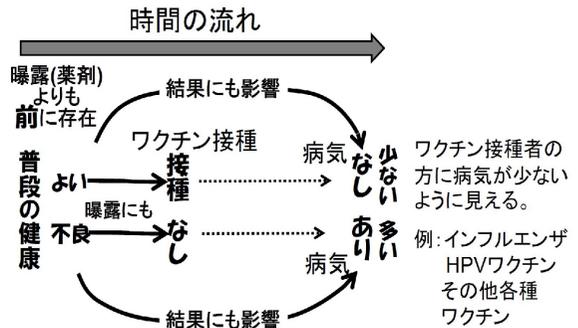


図9：交絡の例(2) 治療薬剤による交絡バイアス

その薬剤の適応症が結果に影響する場合

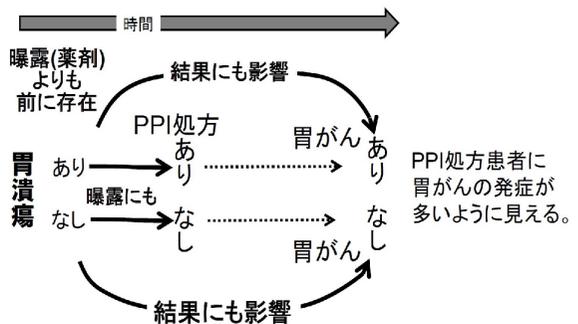
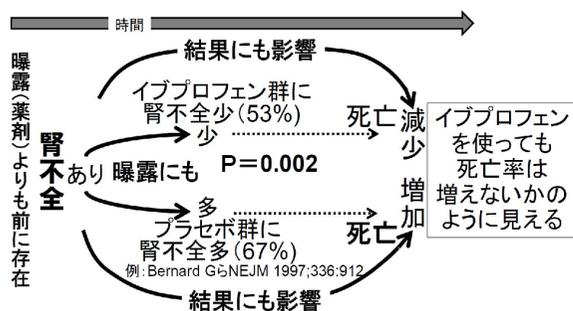


図 10：交絡の例 (3) RCT における交絡バイアスの例



急性腎不全 (AKI) のある熱傷患者の死亡危険度は AKI のない熱傷患者の 11 倍高い [5] と報告されている。イブプロフェン群の腎不全患者の割合が有意に少なければ死亡率は低くなり、プラセボとの差がなくなったり、有意に低くさなりうることに注意が必要である。

気がある程度重くなってきてはじめて薬剤を使います。したがって、薬剤使用群は不使用群に比べてももとの病気が重くなります。

プロトンポンプ阻害剤 (PPI) の使用と胃がんとの関連を調べるために、PPI 使用者と、非使用者を胃潰瘍の有無で比較する場合を考えてみましょう。胃潰瘍があると早期胃がんが紛れ込んでいる場合があります。だから、胃潰瘍のために PPI を使った人と、胃潰瘍がなくて PPI を使わなかった人を単純に比較すると、PPI で胃がんができやすくなるという結果になるでしょう (図 9、前頁)。

適切な比較するためには、胃潰瘍や胃・食道逆流症など適応症の割合が PPI の使用者と非使用者で同じになるようにして追跡をし、胃がんの発生率を比較します。そのようにした最近の研究で、長期の PPI 使用がやはり胃がんを 3 倍超増やすことが分かりました [5]。

RCT でも重大な交絡バイアスがありうる

敗血症の患者にイブプロフェン (I) で解熱した場合とプラセボ (P) を用いた場合で死亡の危険度を評価したランダム化比較試験 (RCT) があります [6]。この試験では、死亡率に差はなかったとされています (I 群:37%対 P 群 40%)。

ランダム化されているはずなのですが、試験開始時に腎不全のある人の割合がプラセボ群に有意に多かったのです (I 群:53%対 P 群 67%、 $p=0.002$) (図 10)。

重症の熱傷で腎不全がある人は、ない人より死亡の危険度が 11 倍とされています [7]。だから、開始時の腎不全の程度をそろえると、イブプロフェン群の死亡率が高くなるはずですが。

不適切な研究では、結果に関係する最も重要な交絡要因を補正していない例がたいへん多く、図 10 のように、ランダム化比較試験といえども、目的とする予後に大きく関係する背景因子に大きな違いがあれば、結果は真実を表していないことになります。まして、観察研究を検討する場合には、補正をした、と報告されている試験でも実際には肝腎の重要な交絡因子で補正していないことがしばしばあります。

くれぐれも注意が必要です。

参考文献

- 1) Rothman ら、Modern Epidemiology、2008
- 2) 丹後俊郎、「医学への統計学」朝倉書店、1993
- 3) 厚生労働省
<https://www.mhlw.go.jp/bunya/kenkou/kekakukansenshou01/qa.html>
- 4) Hama 未公表データ
- 5) Brusselsaers N et al BMJ Open. 2017 Oct 30;7(10):e017739.
- 6) Bernard G, et al. N Engl J Med 1997; 336: 912-9
- 7) Folkestad T et al. Crit Care. 2020 Jan 2;24(1):2

★掲載内容に関して疑問や質問等がある方は、事務局までご一報ください。

E-mail : nopjip@mbr.nifty.com

FAX : 06-6771-6347

