

交絡因子の影響を除く方法

高木廣文・林 邦彦

交絡因子の影響を除く方法

高木廣文 TAKAGI Hirofumi 林 邦彦 HAYASHI Kunihiro
新潟大学医学部保健学科 群馬大学医学部保健学科



- ある要因の結果への影響を示す、効果の指標の数値を変化させる第3の因子を「効果の修飾因子」という。
- 特定の要因の、ある結果への影響を調べる場合に、要因と結果の両者に影響を及ぼし、期待される要因の推定値を得られなくさせる第3の因子を「交絡因子」という。
- 交絡因子は、原因と結果の両方に関係がなければならない。
- 交絡因子は、因果関係を正しく評価できない状況をもたらすので、除去すべきものである。
- 交絡因子を除くためには、「マクマネー検定」や「マンテルーヘンツェル検定」を行う。
- マクマネー検定は、ケース・コントロール研究で、ケースとコントロールが1対1でマッチされていて、特定の要因の有無を比較するときに用いる。
- マンテルーヘンツェル検定は、評価したい要因の影響を正しく推定するために、交絡因子別に層別化し（複数の四分表をつくり）、各層内の交絡因子を均一にすることで、交絡因子の影響を除く方法である。層の数がk個あるときの第k層のオッズ比 ψ_k は層によらず一定である（均一性の仮定）とした場合に、共通オッズ比を求めて検定を行う。
- 特定の事象に関して、多数の研究が行われている場合には、個々の研究の精度や状況が同一でないことを踏まえ、メタ解析の方法として、マンテルーヘンツェル検定を考慮する必要がある。

研究の因果関係に影響を及ぼす「交絡」と「効果の修飾」

ある要因と疾患の発生などの関係を調べる場合、研究下にはないほかの要因により、研究中の因果関係が影響を受けることがある。研究結果に影響を及ぼすような状況には、「交絡」(confounding)と「効果の修飾」(effect modification)の2つがある。

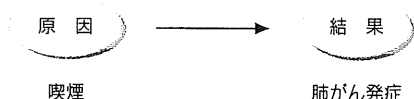
ある要因の結果への影響を示す、効果の指標（リスク差、リスク比、オッズ比など）の数値を変化させるような第3の要因がある場合、「効果の修飾がある」という。そのような第3の要因は「効果の修飾因子」(effect-modifier)とよばれ、詳細な影響評価研究のためには探索すべき因子である。たとえば、

喫煙と肺がん死亡の関係において、緑黄色野菜の多量摂取が喫煙者の死亡を減少させるような場合、緑黄色野菜の摂取は、喫煙と肺がん死亡の関係についての効果の修飾因子と考えられる。

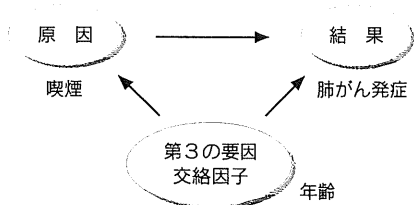
一方、「交絡」は、研究している因果関係を正しく評価できないような状況をもたらすため、疫学研究においては除去すべきものである。

「交絡」とは何か

特定の要因の、ある結果への影響を調べる場合、第3の要因が要因と結果の両者に影響を及ぼし、期待される要因の推定値が得られない場合、その第3の要因は「交絡因子」(confounding factor)とよばれる。



①因果関係のパスモデル



②交絡のある因果関係

ある要因が、研究上で交絡因子となるのは、以下の3条件をすべて満たす場合である¹⁾。

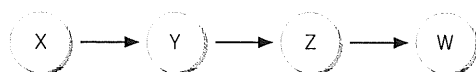
- ①交絡因子は問題となっている結果（疾患の発生，死亡等）のリスク因子である。
- ②問題となっている要因と関係がある。
- ③問題となっている要因と結果の因果連鎖の中間変数ではない。

これらの条件がどのような意味をもつのかを説明するために、変数間の因果関係を図示してみよう。

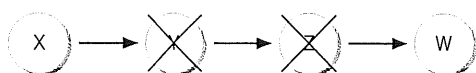
①は、因果関係のパスモデルとよばれるものであり、「原因」である喫煙と、「結果」である肺がん発症の間をパスとよばれる矢印で結んだものである。このような図は、「パス図」，「パスダイアグラム」などともよばれている。原因と考えられる変数と結果を矢印で結び、仮説的な因果関係モデルを示すためによく用いられている。

②は、原因と結果に第3の要因（交絡因子）が影響を及ぼしている関係を示したものである。

交絡因子は、原因と結果の両方に関係がなければならない。たとえば、喫煙の肺がん発症への影響を調べる場合、年齢という因子は肺がん発症に影響するものと考えられる。さらに、喫煙が年齢の影響を受けるのならば（たとえば、高齢者ほど喫煙者が多いなど）、年齢は交絡因子となる。このように、交絡因子は問題となっている結果に対してリスク因子の



③因果連鎖モデル



④中間変数を除いた場合

一つでなければならない。したがって、研究中の原因変数の結果への影響の大きさを正しく推定するためには、交絡因子の影響を除かなければならない。

③は、因果連鎖モデルを示したものである。変数Xは変数Yに影響し、変数Yは変数Zに影響し、変数Zは変数Wに影響するという、変数間の連鎖的な因果関係を示している。

それでは、なぜ交絡因子は因果連鎖の中間変数ではないのだろうか。

④は、③に示した4つの変数のうち、変数YとZを除いた場合を考えるために作成したものである。変数XとWの関係を考える場合、変数YとZの影響を除くと、④からわかるように、変数Xの影響は変数Wまで到達できないことになり、2変数YとZの影響を除いた結果は、変数XとWは無関係ということになる。このように、因果連鎖の中間変数は、交絡因子ではなく、因果関係の解明上の重要因子である。

交絡因子を除く方法

それでは、交絡因子を除くには、どのような方法を用いればよいのだろうか。本連載前号で簡単に解説したように、ケース・コントロール研究ではマッチングを行うという方法がある²⁾。マッチングを行ったデータを解析するには、マクネマー検定 (McNemar's test) やマンテル—ヘンツェル検定 (Mantel-Haenszel test) を用いるのが一般的である。また、複数の変数の結果への影響を分析するには、多変量解析を用いればよい。特に、発症、死亡

などのリスクに関しては、多重ロジスティックモデル、条件付き多重ロジスティックモデル、比例ハザードモデルなどが多用されている。多変量解析に関しては、今後、取り扱う予定である。今回は、マクネマー検定とマンテル―ヘンツェル検定について説明しよう。

ケースとコントロールが1対1でマッチされているときにはマクネマー検定を使う

ケースとコントロールが1対1でマッチされている場合、特定の要因の有無を比較するには、通常のクロス表のカイ2乗検定ではなく、マクネマー検定を用いる³⁾。

マッチングした場合、ケースとコントロールはペアで考えなければならない。

⑤は、ケース（患者）とコントロール（対照）が1対1のペアにマッチされている場合、ある要因の有無によりN組のペアがどのように分類されるかを示している。⑤に示すa, b, c, dはそれぞれのペアの組数を示している。たとえば、ケースもコントロールも共に要因があるペアはa組、どちらも要因がないペアはd組いることを示している。

リスク比の推定値であるオッズ比とその標準誤差は、通常の場合と異なり、以下の式により求められる。

すなわち、オッズ比ORは、

$$OR=b/c$$

により求められ、その標準誤差SE(OR)は、

$$SE(OR) = (1+OR) \times \sqrt{\frac{OR}{b+c}}$$

により求められる。

オッズ比とその標準誤差から、母集団でのオッズ比（母オッズ比 ψ ）の95%信頼区間は、標準正規分布を仮定して、

$$OR - 1.96 \times SE(OR) < \psi < OR + 1.96 \times SE(OR)$$

のように簡単に計算できる。求めた信頼区間が1を含

		対 照		
		○	×	計
患者	○	a	b	a+b
	×	c	d	c+d
計		a+c	b+d	N

○：要因あり，×：要因なし

⑤対応のある場合の四分表

		対 照		
		○	×	計
患者	○	10	40	50
	×	10	40	50
計		20	80	100

○：ピロリ菌あり，×：要因なし

⑥ピロリ菌の有無

まなければ、オッズ比はリスク因子として、統計学的に有意なものと考えてよいことになる。

帰無仮説^{注1)}「母オッズ比=1」、すなわち、要因の有無と疾患には関係がない、という検定は、

$$\chi^2_M = \frac{(|b-c|-1)^2}{b+c}$$

が自由度1のカイ2乗分布^{注2)}に従うことで行える。

注1) 帰無仮説と検定：検定は、標本データを用いて、母集団の各種特性について言及するための方法である。

母集団について、ある仮説を設定し、仮説が正しい場合に、データで観察されるようなある特性値、たとえば平均値の差、相関係数、オッズ比などが偶然得られる確率を計算する。その確率がある基準値より小さい場合、その事象が偶然起きるとは考え難いことになる。そのような場合には、確率計算のもとになった「仮説は正しい」という設定は間違っているものとして、仮説を「棄却する」。その結果、2グループの平均値であれば「有意差あり」、相関係数であれば「有意な相関がある」などのように結論を得ることができる。

判断の基準となる確率の大きさは、「有意確率（ p 値）」とよばれる。検定では仮説は棄却されるために存在し、そのため「帰無仮説 null hypothesis」とよばれている。研究上の作業仮説では、「喫煙は肺がん発症に影響する」のような表現になるが、帰無仮説では「喫煙は肺がん発症に影響しない」という表現になる。これを棄却することで、喫煙が肺がん発症に影響することを示すという理論展開になっている。

すなわち、前記の χ^2_M が3.84（自由度1のカイ2乗分布の5%点）より大きければ、5%の有意水準で帰無仮説を棄却し、問題にしている要因は当該疾患のリスク因子と考えてよいことになる。

前述の説明でわかるように、計算に使用されているのは、⑤の中のbとcのみである。すなわち、ケースとコントロールが共に要因をもつ場合（a）ともない場合（d）はまったく計算に使用されていない。このように、情報として価値があるのは、ケースとコントロールで要因の有無が異なるペアの個数のみである点に注意したい。

⑥に、性、年齢をマッチした胃潰瘍患者と対照者のペア100組について、ヘリコバクターピロリの有無を例としてあげる。

オッズ比は、 $OR=40/10=4$ 、とすぐに求められる。また、オッズ比の標準誤差は、

$$SE(OR) = (1+4) \times \sqrt{4/(40+10)} \\ = 5 \times \sqrt{0.08} = 1.4142 \dots$$

と求められるので、母オッズ比の95%信頼区間は、

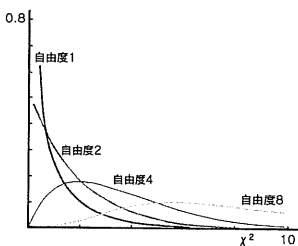
$$1.96 \times SE(OR) = 1.96 \times 1.4142 = 2.772$$

をまず計算し、求めたオッズ比に加減することで、

注2) 検定とカイ2乗分布：帰無仮説が成立していると仮定して、観測されるようなデータが偶然得られる確率を直接計算するのはそれほど簡単ではない。そのため、確率計算に便利ように、データからある統計量を計算して、その統計量が従う特定の分布を利用して確率を求めるのが一般的である。

使用する統計量によって、その統計量が従う分布の名称から、z検定、t検定、 χ^2 検定、F検定などとよばれている。

質問紙調査などにより収集されたデータでは、クロス表に基づく検定が多用されている。クロス表の検定のための統計量は χ^2 （カイ2乗）値であり、カイ2乗分布に従うことから、確率を求める。もともと、カイ2乗分布は標準正規分布に従う変数から独立に抽出されたデータの2乗和の分布である。そのため、何個のデータの2乗和なのかによって、その分布の形状などが大きく異なっている。そのデータの個数は「自由度」とよばれており、図のように、カイ2乗分布は自由度により形状が大きく変化する。



$$1.2 < \psi < 6.8$$

のように求められる。上記の95%信頼区間には1を含んでいないので、ピロリ菌の有無と胃潰瘍には関係があると考えてよいことになる。

また、マクネマー検定では、

$$\chi^2_M = \frac{(|40-10|-1)^2}{40+10} = \frac{29^2}{50} = 16.82$$

となり、自由度1のカイ2乗分布の5%点3.84、また1%点6.63より大きいので、高度に有意という結果になり、信頼区間による結果と一致している。

マンテル・ヘンツェルの方法 交絡因子のカテゴリーごとに層別化して交絡因子の影響を除く

マッチングの方法は、比較するケースとコントロールの条件をそろえるという意味をもっている。一般に、性、年齢、喫煙習慣などの交絡因子の影響を除き、評価したい要因の影響を正しく推定するための方法として、前号で簡単に触れた層別解析という方法がある。これは、交絡因子の各カテゴリー（層）ごとに集計を行い、各層内では交絡因子が均一になるようにすることで、交絡因子の影響を除くという方法である。

⑦は、Avadexという殺虫剤の発がん性をマウスの2系列XとYで雌雄別に実験した結果を示したものである⁴⁾。このような実験では、薬剤の効果が性別やマウスの系列で異なる可能性が高いので、性別や系列を無視して、一つのクロス表にまとめてしまうのは、適切な集計方法ではないと考えられる。

このような場合、複数の四分表をまとめて、各層に共通するオッズ比（共通オッズ比）を推定するための近似的な方法が、Mantel-Haenszelによって提唱されている^{5, 6)}。

全体で層の個数がk個あり、第k層のオッズ比 ψ_k は層によらず一定（均一性の仮定）であるとした場

層		投与	非投与	計
系X：雄	肺がん	4	5	9
	なし	12	74	86
	小計	16	79	95
系X：雌	肺がん	2	3	5
	なし	14	84	98
	小計	16	87	103
系Y：雄	肺がん	4	10	14
	なし	14	80	94
	小計	18	90	108
系Y：雌	肺がん	1	3	4
	なし	14	79	93
	小計	15	82	97

⑦ Avadex と肺がん

	要 因		
	あり	なし	計
患者	a _k	b _k	m _k
対照	c _k	d _k	n _k
計	x _k	y _k	N _k

(k=1,2,...,K)

⑧ 第k層の四分表

合、共通オッズ比 ϕ を推定しよう。

⑧は、第k層について、患者と対照の、要因の有無別の四分表を示したものである。a_k, b_kなどのkは第k層の値であることを示すための添字である。共通オッズ比の推定量 ϕ_{MH} は、

$$\phi_{MH} = \frac{\sum_k (a_k d_k / N_k)}{\sum_k (b_k c_k / N_k)} = \frac{\sum_k R_k}{\sum_k S_k} = \frac{R_+}{S_+}$$

により求められる（上記の右側の式中の記号は、後の計算式で使用するために用いている）。上記の計算の方法は、オッズ比の計算で、各層で分子、分母にくる値をそれぞれ層の人数で割るという操作を行っ

ているものである。したがって、層が一つしかない場合には、通常のオッズ比と一致する。

上記のマンテル－ヘンツェル推定量の信頼区間を求める方法は、多数提唱されているが⁶⁾、ここではRobinsらの方法^{3, 7)}を紹介する。

共通オッズ比の信頼区間を構成するためには、オッズ比の標準誤差を推定する必要がある。しかし、帰無仮説のもとでのオッズ比の分布は、大きく右にゆがんでいるので、オッズ比の対数をとった値、すなわち対数オッズ比 $\log \phi_{MH}$ とその分散V（標準誤差SEの2乗）を使用する。

Robinsらは、対数オッズ比の分散Vとして、

$$V = \frac{1}{2} \sum_k \left[\frac{P_k R_k}{R_+^2} + \frac{Q_k R_k + P_k S_k}{R_+ S_+} + \frac{Q_k S_k}{S_+^2} \right]$$

を提唱した。ここで、

$$P_k = (a_k + d_k) / N_k, \quad Q_k = (b_k + c_k) / N_k,$$

である。

共通オッズ比の95%信頼区間は、標準正規分布に近似することで、

$$\phi_{MH} \cdot \exp[\pm 1.96 \cdot \sqrt{V}]$$

により、その下限と上限が求められる。

⑦のAvadexデータの共通オッズ比の計算を行ってみよう。

各層のオッズ比は、

$$\text{第1層: } (4 \times 74) / (5 \times 12) = 4.933$$

$$\text{第2層: } (2 \times 84) / (3 \times 14) = 4.000$$

$$\text{第3層: } (4 \times 80) / (10 \times 14) = 2.286$$

$$\text{第4層: } (1 \times 79) / (3 \times 14) = 1.881$$

のように求められる。オッズ比は、層によって2倍以上の相違がある。各層のオッズ比が異なる場合、共通オッズ比を仮定することの妥当性に関する検定（均一性の検定：Breslow-Day検定^{8, 9)}）がある。ここでは詳細にはふれないが、均一性の検定はあまり検出力が高くないので、均一性が否定されるのはまれである。

次に、共通オッズ比を求めてみよう。まず、R₊と

S_+ を求める.

$$R_+ = (4 \times 74)/95 + (2 \times 84)/103 + (4 \times 80)/108 + (1 \times 79)/97 = 8.524$$

$$S_+ = (5 \times 12)/95 + (3 \times 14)/103 + (10 \times 14)/108 + (3 \times 14)/97 = 2.769$$

と求められるので, 共通オッズ比は,

$$\psi_{MH} = R_+/S_+ = 8.524/2.769 = 3.079$$

と推定できる.

同様に, 対数オッズ比の分散 V を求めると,

$$SE(\log OR) = \sqrt{V} = 0.406$$

と求められる (なお, 実際の計算は煩雑で複雑であり, コンピュータの使用が必須なので, ここでは省略した).

前述の式にあてはめると, 95%信頼区間は, $1.39 < \psi < 6.82$, と求められる. 実際に計算を確かめたい読者は, 筆者のホームページ (<http://www.clg.niigata-u.ac.jp/~takagi/>) の「簡単統計計算」のなかの層別四分表の計算ルーティンを試してみるとよいだろう.

帰無仮説「母共通オッズ比=1」の検定は, 第 k 層の a_k の期待値 $E(a_k)$ と実際値との相違から, カイ2乗分布を仮定して行われる. 第 k 層に関して, a_k の期待値 $E(a_k)$ と分散 $V(a_k)$ は,

$$E(a_k) = x_k \cdot m_k / N_k$$

および,

$$V(a_k) = \frac{m_k n_k x_k y_k}{N_k^2 (N_k - 1)}$$

によって求められる. 検定のための統計量は,

$$\chi_{CMH}^2 = \frac{(|\sum_k (a_k - E(a_k))| - 0.5)^2}{\sum_k V(a_k)}$$

となり, これが, 自由度1のカイ2乗分布に従うことから, 5%点3.84, 1%点6.63と比較し, それ以上ならば有意とすればよい.

⑦のAvadexのデータについて計算すると,

観察値の合計: $\sum_k a_k = 11$

期待値の合計:

$$\sum_k E(a_k) = (9 \times 16)/95 + (5 \times 16)/103 + (14 \times 18)/108 + (4 \times 15)/97 = 5.244$$

分散の合計:

$$\sum_k V(a_k) = (9 \times 86 \times 16 \times 79)/(95^2 \times 94) + (5 \times 98 \times 16 \times 87)/(103^2 \times 102) + (14 \times 94 \times 18 \times 90)/(108^2 \times 107) + (4 \times 93 \times 15 \times 82)/(97^2 \times 96) = 3.998$$

となるので,

$$\chi_{CMH}^2 = \frac{(|11 - 5.244| - 0.5)^2}{3.998} = 6.908$$

と求められる. この値は, 自由度1のカイ2乗分布の上側1%点6.63より大きいので, Avadexと肺がんには有意な関係があると考えてよいことになる.

マンテル—ヘンツェルの方法 のメタ解析への応用

マンテル—ヘンツェルの方法は, 多数の四分表がある場合に, 各四分表の各セルの数値を足し合わせて表を作るのではなく, そのままの状態でそれらをまとめるための指標を与えることができる. この意味から, 特定の事象に関して, 多数の研究が行われている場合, 個々の研究の精度や状況が同一ではないため, 単純な足し算が適当でないと考えられる場合などに, マンテル—ヘンツェルの方法を用いることができる.

⑨は, 新生児がアトピーに対してハイリスクである

研究報告	実験群	対照群	オッズ比	95%信頼区間
Chandra 1986	5/35*	11/36	0.379	0.116~1.236
Chandra 1989	11/49	21/48	0.372	0.154~0.898
Lovegrove 1994	5/12	8/14	0.536	0.112~2.553
総計	21/96	40/98	0.406	0.216~0.762

*数字は, アトピー児/対象者数

⑨母親の抗原曝露回避とアトピー

と考えられる母親に、授乳期間中に抗原を含まない食事を摂取した場合、児のアトピー罹患の予防に有効かを研究した3論文をまとめたものである¹⁰⁾。

オッズ比は、すべての研究で1より小さいが、95%信頼区間に1を含まないのは2番目の研究のみである。単純に3論文の各人数を加えた場合を総計に示した。3論文の結果を累積した場合、オッズ比は0.406となり、また95%信頼区間の上限も1未満となり、抗原を含まない食事の摂取はアトピー予防に有効のようである。

しかし、3論文の実験条件は必ずしも同一の精度、状況で行われたわけではないので、単純な足し算は問題かもしれない。前述のように各層（ここでは各研究）の母オッズ比の均一性に関する検定として、Breslow-Day検定があるが、あまり感度のよい検定ではなく、均一性が否定されることは少ない。ここでは前述のマンテル―ヘンツェルの方法で、共通オッズ比とその95%信頼区間を求めてみよう。

共通オッズ比ORは、

$$\begin{aligned} OR &= \frac{(5 \times 25)/71 + (11 \times 27)/97 + (5 \times 6)/26}{(30 \times 11)/71 + (38 \times 21)/97 + (7 \times 8)/26} \\ &= \frac{5.97627}{15.02854} = 0.398 \end{aligned}$$

と求められる。詳細な計算の過程は省略するが、母共通オッズ比の95%信頼区間は、0.209～0.756と求められる。

これらの結果は、マンテル―ヘンツェルの方法による数値のほうだが、オッズ比が若干小さい側に寄っていたが、単純に3つの研究結果を累積して得られた結果と大差はなかった。

今回のように2つの結果に差がない場合は特に問題はないが、場合によっては結果が大きく異なることもある¹¹⁾。そのような場合、メタ解析として単純に研究結果を累積することに問題があることを示すものと考えられるので、マンテル―ヘンツェルの方法などを考慮する必要があるだろう。

文献

- 1) 高木廣文：疫学の考え方と研究デザイン。第4回JEA疫学セミナー抄録集。日本疫学会；1997。p.6-13。
- 2) 高木廣文、林邦彦：EBNのための統計の読み方④ケース・コントロール研究。EB NURSING 2001；1(4)：474-479。
- 3) 豊川裕之、丸井英二、高木廣文ほか：保健学講座 ⑥ 疫学・保健統計。メチカルフレンド社；1997。p.148-156。
- 4) Innes JRM, et al. : Bioassay of pesticides and industrial chemicals for tumorigenicity in mice : a preliminary note. Journal of the National Cancer Institute 1969 ; 42 : 1101-1114.
- 5) Mantel N, Haenszel W : Statistical aspects of the analysis of data from retrospective studies of disease. Journal of the National Cancer Institute 1959 ; 22 : 719-748.
- 6) 佐藤俊哉、高木廣文、柳川 堯ほか：Mantel-Haenszel の方法による複数の2×2表の要約。統計数理 1998；46(1)：153-177。
- 7) Robins JM, Breslow NE, Greenland S : Estimators of the Mantel-Haenszel variance consistent in both sparse data and large-strata limiting model. Biometrics 1986；42：311-323。
- 8) Breslow NE, Day NE : Statistical Methods in Cancer Research, Volume1-The Analysis of Case-Control Studies. Lyon : IARC ; 1990. p.142-146.
- 9) 柳川 堯：離散多変量データの解析。共立出版；1986。p.73-74。
- 10) 別府宏園：くすりとエビデンス④ 看護現場のエビデンス。EB NURSING 2001；1(4)：486-492。
- 11) Simpson EH : The Interpretation of Interaction in Contingency Tables. Journal of the Royal Statistical Society 1951；B13：238-241。