

8. 交絡とバイアス

高木廣文
東邦大学看護学部
国際保健看護学研究室

交絡と効果の修飾について

ある要因と疾患の発生等の関係を調べる場合、他の要因の存在の影響により、正しくその要因のリスク等を評価できないことがある。

- (1) 交絡 confounding
- (2) 効果の修飾 effect modification

交絡とは

- ・ 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。
- ・ 第3の要因：
交絡因子 confounding factor

交絡因子の条件

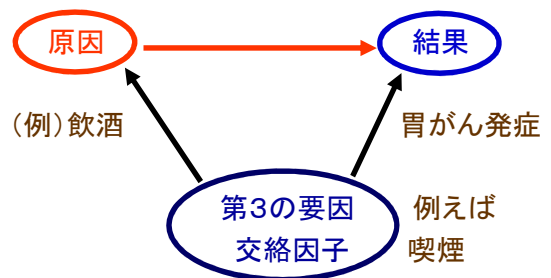
- 1) 交絡因子は問題となっている結果（疾患の発生、死亡等）のリスク因子である。
- 2) 問題となっている要因（原因）と関係がある。
- 3) 問題となっている要因と結果の因果連鎖の中間変数ではない。

因果関係のパスモデル



- ・ 原因と結果を矢印で結び、因果関係があることを示す（パス図、パスダイアグラム）

交絡のある因果関係

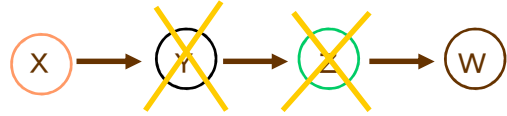


因果連鎖モデル



- 一つの結果が、次の原因になり、鎖のように繋がっている関係

交絡因子はなぜ因果連鎖の中間変数ではないのか？



- 変数YやZがなければ、変数Xの影響は結果変数Wに及ばない！

効果の修飾とは何か？

- ある要因の結果への影響を示す効果の指標（リスク差やリスク比等）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。
 - ⇒ 第3の要因 = 効果の修飾因子 effect-modifier
- 交絡 ⇒ 除去すべきもの
- 効果の修飾 ⇒ 詳細な影響評価

交絡因子を除く方法は？

- 交絡因子を除き、効果の修飾を見つけるための方法は？
 - ◆ 層別解析 :
 - Mantel-Haenszel検定 / 推定
 - マッチング matching
 - ◆ 多変量解析 :
 - 条件付きロジスティックモデル
 - 比例ハザードモデル

層別解析

- 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

交絡因子の各カテゴリ（層）ごとに集計し、各層内では交絡因子が均一になるようにする。

(例) 性別、年齢別、喫煙習慣別など

表. Avadex と肺癌

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

表. 層がない場合のAvadex と肺癌

	投与	非投与	計
肺癌	11	21	32
なし	54	317	371
計	65	338	403

- ・ リスク比 = $(11/65) \div (21/338) = 2.724$
95%信頼区間(1.315, 5.497)
- ・ オッズ比 = $11 \times 317 \div (21 \times 54) = 3.075$
95%信頼区間(1.304, 7.162)

Mantel-Haenszel 推定量について

- ・ 層別四分表について, 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

- ・ 全体で層の個数はK個
- ・ 第k層のオッズ比 ψ_k は層によらず一定 (均一性の仮定)
- ・ 共通オッズ比 ψ を推定する

表. 第k層の四分表

	要因		計
	あり	なし	
患者	a_k	b_k	m_k
対照	c_k	d_k	n_k
計	x_k	y_k	N_k

($k=1, 2, \dots, K$)

共通オッズ比の推定

- ・ Mantel-Haenszel 推定量 ψ_{MH}

$$\psi_{MH} = \frac{\sum_{k=1}^K (a_k d_k / N_k)}{\sum_{k=1}^K (b_k c_k / N_k)} = \frac{\sum_{k=1}^K R_k}{\sum_{k=1}^K S_k}$$

$$= \frac{R_+}{S_+}$$

共通オッズ比の信頼区間の構成

- 対数オッズ比 $\log \psi_{MH}$ とその分散を使用 (Robins, Greenland, Breslow: Biometrics, 1986)

$$V_{RGB} = \frac{1}{2} \sum_{k=1}^K \left[\frac{P_k R_k}{R_+^2} + \frac{Q_k R_k + P_k S_k}{R_+ S_+} + \frac{Q_k S_k}{S_+^2} \right]$$

$$P_k = (a_k + d_k) / N_k$$

$$Q_k = (b_k + c_k) / N_k$$

標準正規分布と区間推定

z 値の分布は近似的に標準正規分布に従う。

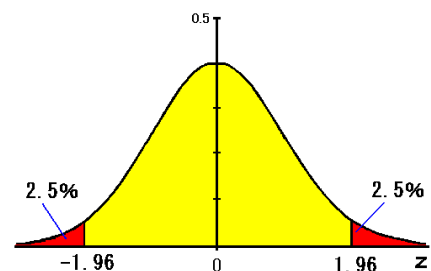


図 標準正規分布

共通オッズ比の信頼区間(1)

- ・対数オッズ比を標準正規分布で近似：
95%信頼区間

$$\log(\psi_{MH}) \pm 1.96 \times \sqrt{V_{RGH}}$$

99%信頼区間

$$\log(\psi_{MH}) \pm 2.58 \times \sqrt{V_{RGH}}$$

共通オッズ比の信頼区間(2)

- ・対数オッズ比を元に戻す
95%信頼区間

$$\psi_{MH} \times \exp\left[\pm 1.96 \times \sqrt{V_{RGH}}\right]$$

99%信頼区間

$$\psi_{MH} \times \exp\left[\pm 2.58 \times \sqrt{V_{RGH}}\right]$$

共通オッズ比の計算例-1

Avadexデータの共通オッズ比の計算を行う。

各層のオッズ比：

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

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

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

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

共通オッズ比の計算例-2

- ・ R_+ , S_+ を求める。

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

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

共通オッズ比の推定値：

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

共通オッズ比の信頼区間計算例

Avadexデータの共通オッズ比の信頼区間の計算を行う：

Robins-Greenland-Breslowによる標準誤差
SE=0.406

$$95\% \text{信頼区間} : 1.39 < \psi < 6.82$$

$$99\% \text{信頼区間} : 1.08 < \psi < 8.75$$

Cochran-Mantel-Haenszel検定(1)

- ・「共通オッズ比=1」の検定

第k層の a_k の期待値

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

分散

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

Cochran-Mantel-Haenszel検定 (2)

- 第 k 層の a_k の期待値 $E(a_k)$ と分散 $V(a_k)$ を用いて :

$$\chi_{CMH}^2 = \frac{\left[\sum_{k=1}^K (a_k - E(a_k)) \right] - 0.5}{\sum_{k=1}^K V(a_k)}$$

自由度 1 のカイ 2 乗分布 :
5%点 3.84, 1%点 6.63 と比較する

Cochran-Mantel-Haenszel検定の例(1)

Avadexのデータ :

- 観察値の合計 : $\sum_{k=1}^4 a_k = 11$
- 期待値の合計 : $\sum_{k=1}^4 E(a_k) = 9 \times 16 / 95 + 5 \times 16 / 103 + 14 \times 18 / 108 + 4 \times 15 / 97 = 5.244$
- 分散の合計 : $\sum_{k=1}^4 V(a_k) = 9 \times 86 \times 16 \times 79 / (95^2 \times 94) + 5 \times 98 \times 16 \times 87 / (103^2 \times 102) + 14 \times 94 \times 18 \times 90 / (108^2 \times 107) + 4 \times 93 \times 15 \times 82 / (97^2 \times 96) = 3.998$

Cochran-Mantel-Haenszel検定の例(2)

期待値と分散から, 検定のための統計量 :

$$\chi_{CMH}^2 = \frac{[11 - 5.244] - 0.5}{3.998} = 6.908$$

自由度 1 のカイ 2 乗分布の上側 1%点 6.63 より大きい \Rightarrow 有意な関係がある
($p = 0.0086$)

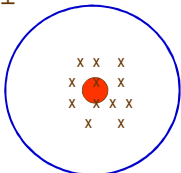
バイアスの問題

バイアス :

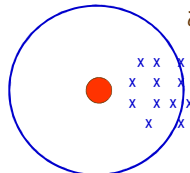
真実からの結果や影響の偏り, もしくはそのような偏りをもたらすプロセス

真実とは体系的に異なる結論を導くデータ収集, 分析, 解釈, 出版, レビューにおけるあるゆる傾向

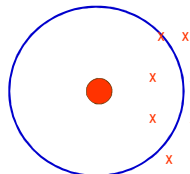
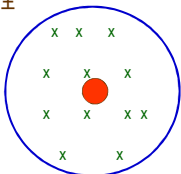
信頼性
高い



バイアス
あり



信頼性
低い



バイアスの分類

- 情報バイアス, 観察バイアス
- 選択バイアス

そのほかにも多くのバイアスがある

1. 情報バイアス, 観察バイアス

曝露や結果のデータの測定における欠陥。比較するグループ間の情報の質（正確性）が異なるという結果をもたらす。

なぜ情報バイアスが生じるのか？

1. リコール・バイアス(思いだしバイアス)
過去のイベントや経験の記憶を思い出すことの正確さや完全さでの差による, 系統的エラー
(患者と健常者での病気に関する意識の差など)
2. 面接者によるバイアス
選択的にデータを集めてしまうという面接者の潜在意識や意識による系統的エラー
3. 観察者によるバイアス
観察者の変動(能力などのばらつき)による真実の値と実際に観察された値の間の系統的差

情報バイアスを避けるには

1. より客観的な情報を得る
2. 盲検（マスク化）の方法を用いる
3. より正確な情報を得るための方法について観察者の訓練をする

2. 選択バイアス

研究のために選ばれた人と選ばれなかった人の中での特性についての系統的差によるエラー

なぜ選択バイアスが生じるのか？

1. Berksonのバイアス
2. デザイン・バイアス
3. 健康労働者効果

1. Berksonのバイアス

選択バイアスの一つで, ケース・コントロール研究において病院のケースとコントロールが互いに系統的に異なってしまう。

研究上の曝露と疾病の組み合わせが, 病院への管理のリスクを増加させる場合に, このバイアスが生じる。この場合, 病院のケースでの曝露率が病院のコントロールよりも系統的に大きくなり, これが順にオッズ比を系統的に歪めてしまう。

2. デザイン・バイアス

真の値と誤った研究デザインの結果として得られた値の間の差。

例には、制御ができていない研究がある。2つ以上のプロセスを分離できない(交絡のある)研究や、母集団の定義のあいまいな研究、不適当なコントロール群で実施された研究である。

3. 健康労働者効果

職業病の研究で初めに観察された現象。労働者は通常、一般人口集団よりも、全ての死亡率で低値を示す。これは、重大な病気や慢性的な不具合は、通常は雇用から除外されるからである。

選択バイアスを避けるには

1. 無作為標本抽出
2. 研究上の疾病以外には、ケースとコントロールを選ぶために、同一基準を用いる。