

# 交絡と層化

## 交絡 (confounding)とは？

- 結果 に対する暴露 とは独立した危険因子であり、なおかつ暴露とも相関関係 (有意差がある必要はない) をもつ場合を交絡と呼ぶ (図1)。

# Confounding

この関係を調べたい

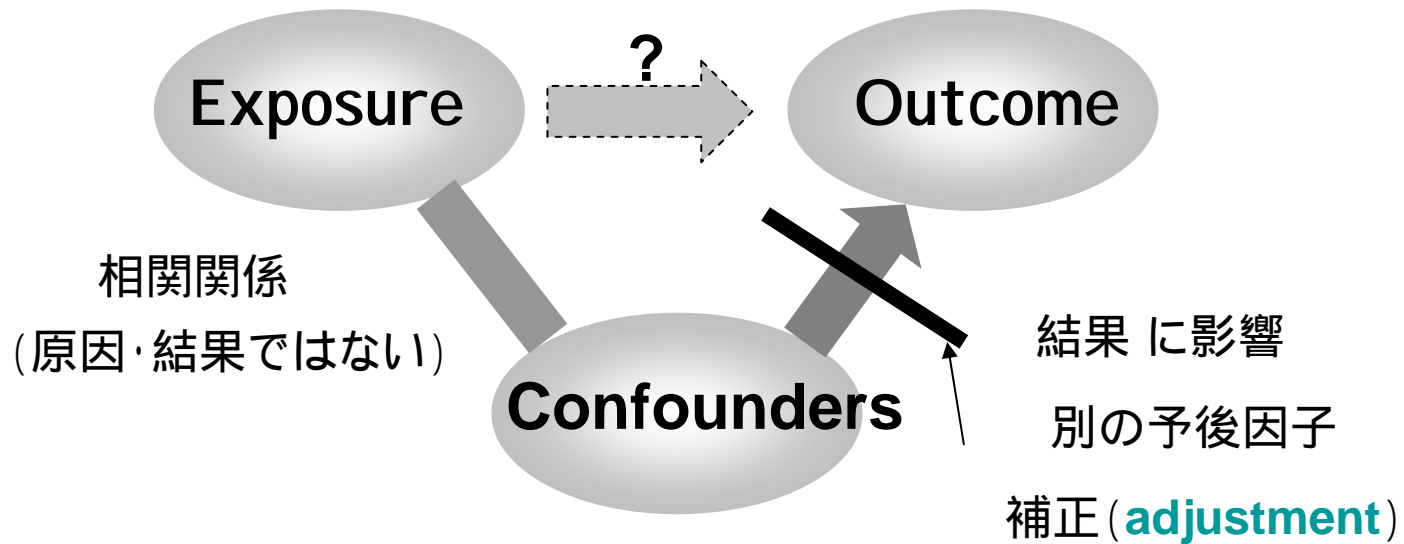


図1 交絡因子(confounders)の概念図

- 一方、これから調べようとしている原因因子とほぼイコールであったり、既に因果関係があるようなものでは、交絡因子とならない(図2)。

# Not Confounding

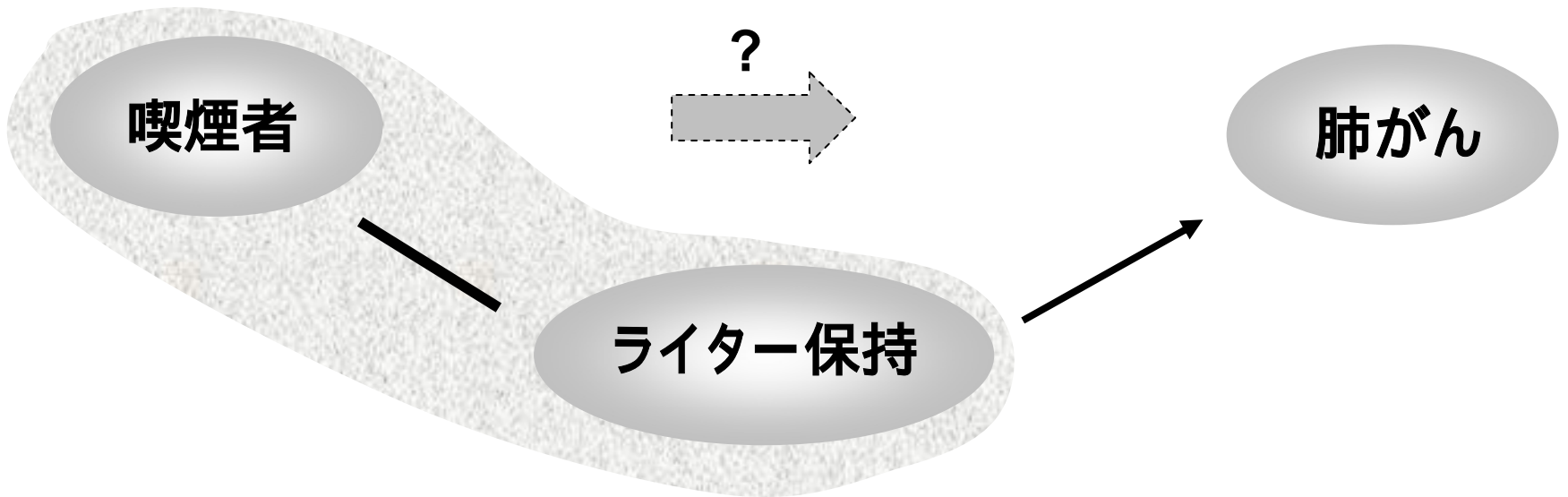


図2 上記のようにexposure と3rd factor がほぼイコールである場合には、その3rd factor は交絡因子とはならない。あるいは、交絡因子として補正すると結果をおかしくしてしまう。

- 「病原性大腸菌による胃腸炎に対して抗生物質を使うと溶血性尿毒症のリスクが上がるかどうか」を例に交絡因子について考えてみる (NEJM 2000; 342: 1930-6.)。この場合の暴露とは抗生物質投与であり、結果は溶血性尿毒症であり、交絡因子は重症度である。つまり抗生物質を処方する医師の基準が予後因子や重症度(重症感)にあるとすると、暴露と交絡の間には相関関係がある。例えば医師が、「重症感が強い」と感じたら抗生物質を処方する傾向にあるなどである。そして、重症度は当然の結果として溶血性尿毒症の合併と関係する。即ち別の予後因子である。この交絡との関係を、ウエイトをおいて断ち切る作業が補正(adjustment)である。臨床研究論文において、しばしばunivariate or crude relative risk (univariate or crude OR)とadjusted relative risk (or adjusted OR)が表に示されているが、交絡を補正する前と後の値を意味している。すなわち、補正後の値の方が真の暴露と結果の関係を表していることになる。時には、年齢、性別、喫煙、肥満度によりmultivariate adjustmentを行ったなどと表の注釈として記載しながら、adjusted valueのみを示す論文も多い。そして、補正前後で暴露と結果の関係を表す比が10%以上変化する場合を交絡と考える。

## 交絡の方向性と程度

- 抗生物質を使用していると溶血性尿毒症が増えることがわかっている (OR = 14.3; 95%CI: 2.9 – 70.7)。そして、白血球数が高いと抗生物質を使用される頻度が高いとしよう(図3)。実際有意差はないが、抗生物質使用の平均白血球数は13,100であり、非使用では11,600であり、プラスの関係にある。この相関関係には必ずしも有意差を必要とはしない。また白血球数が13,000以上の場合、溶血性尿毒症発症は3.9倍高くなる。つまりプラスの関係である。よって、抗生物質使用と溶血性尿毒症の間において、白血球数は交絡因子である。このような暴露と交絡がプラスの関係、交絡と結果がプラスの関係である場合、暴露と結果の関係、すなわち見かけ上抗生物質使用が溶血性尿毒症に対する影響を見かけ上過大に評価してしまう(プラス・プラス、マイナス・マイナスは過大評価、プラス・マイナス、マイナス・プラスは過小評価と覚えておくと便利)。よって、単に抗生物質使用と溶血性尿毒症のリスク比より白血球数で補正したリスク比の方が小さくなる(この論文では白血球単独では補正されていない)。

# Overestimation

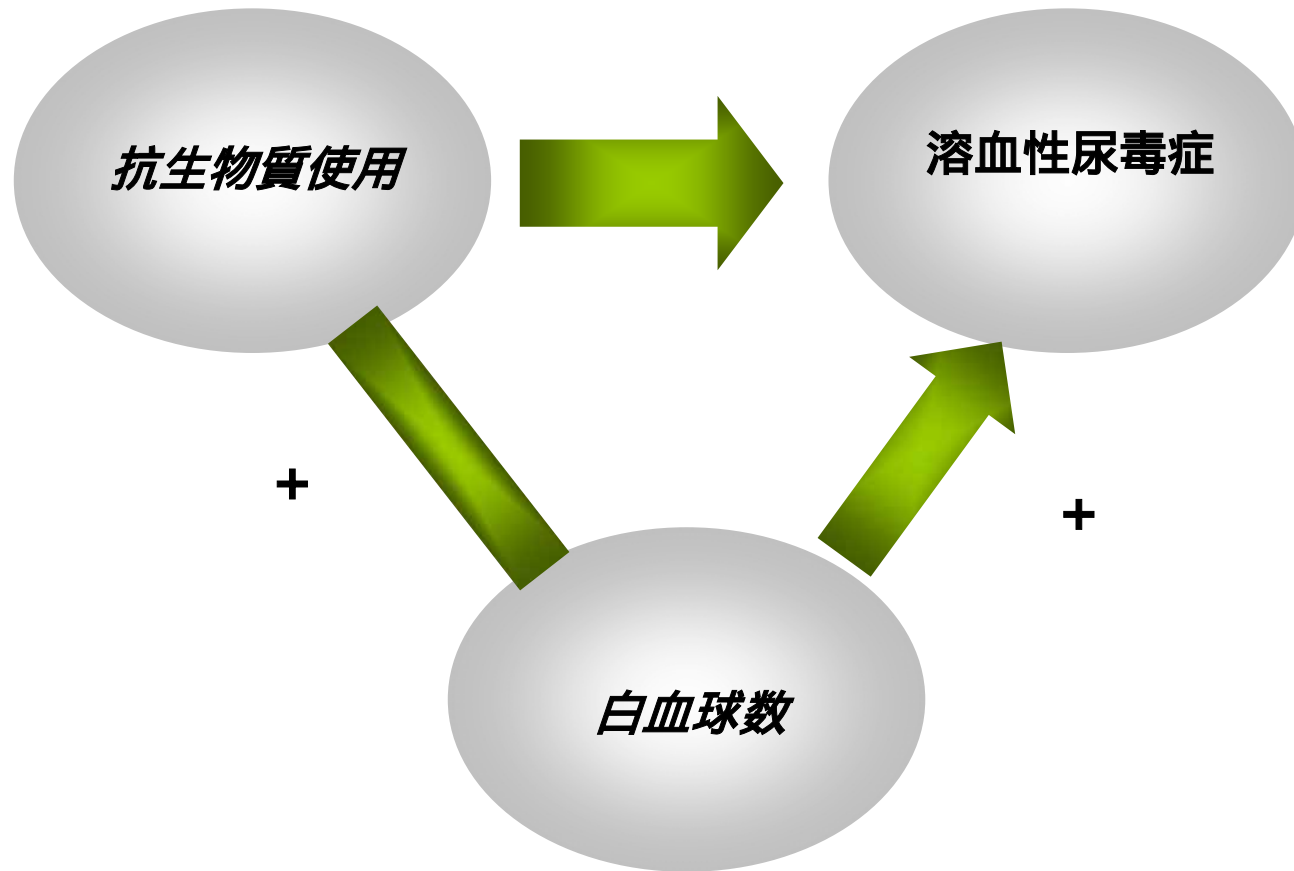


図3 抗生物質使用、白血球数、溶血尿毒症候群の関係



- 論文では、白血球数だけではなく、症状が出始めてから便培養を採るまでの日数も予後因子として重要視されている(図4)。抗生物質使用者では、有意差は認めないものの、この日数が短い傾向にある。よって両者の関係はマイナスである。そして、この日数が短い程溶血性尿毒症になりやすいため(RR=0.3; 95%CI: 0.2 – 0.8)、両者の関係はマイナスである。マイナスとマイナスなので抗生物質の溶血性尿毒症への影響を過大評価する可能性があるが、補正前後のリスク比はほとんど変化していない(<10%)ため、便培養までの日数は単独では少なくとも交絡因子とはなっていない。
- 本論文では便培養までの日数単独では補正されていないが、白血球数と併せた2つの因子で同時補正している。

# Overestimation

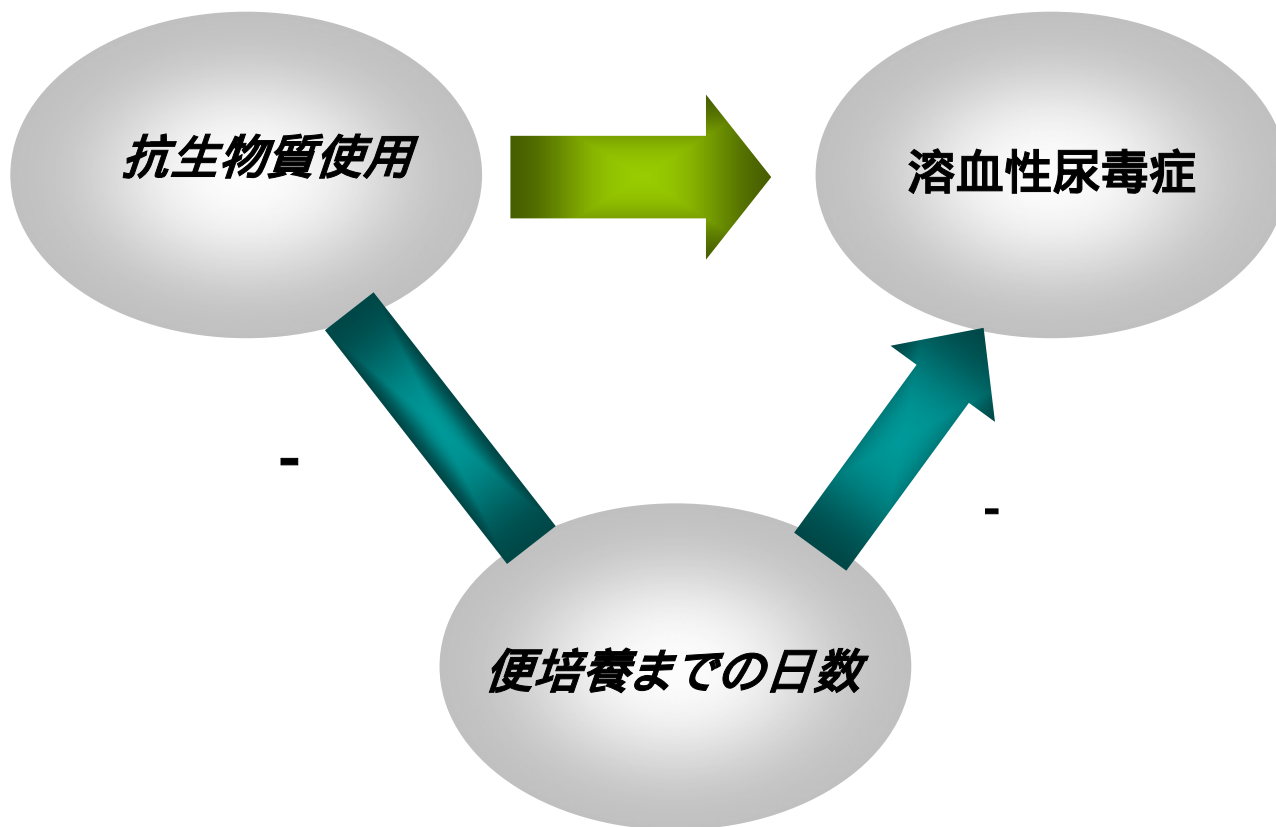


図4 抗生物質使用、便培養までの日数、溶血尿毒症候群の関係

- 本論文では示されていないが、2つの交絡因子を併せた場合、抗生物質使用との関係はマイナスとなり、溶血性尿毒症との関係は逆にプラスとなっていたかもしれない(図5)。このように暴露 - 交絡、交絡 - 結果の関係がマイナス、プラス、あるいはプラス、マイナスのように逆の符号である場合、暴露 - 結果の関係を見かけ上過小評価してしまう。補正前のリスク比が14.3であったのに、補正後のそれは17.3に増加しているのはそのためである。
- 以上みてきたように暴露と交絡、交絡と結果の関係により暴露と結果の関係は過大評価あるいは過小評価される。またその程度は交絡の暴露と結果との関係によって規定される。すなわち暴露と交絡の間に強い関係が存在しても交絡が結果に極軽度しか影響しなければたいしたことではない。逆も真なりである。両者が強く関係していれば交絡の影響は当然大きくなる。バイアスと交絡との相違は、前者は既にデータに含有される真の値とのずれであり、後者は他の因子の影響で発生する真の関係のずれである。よって、バイアスが一旦データに含有された場合にはこれを除去することはできないが、交絡に関しては、いくつかの方法によりこれをある程度除去することができる。

# Underestimation

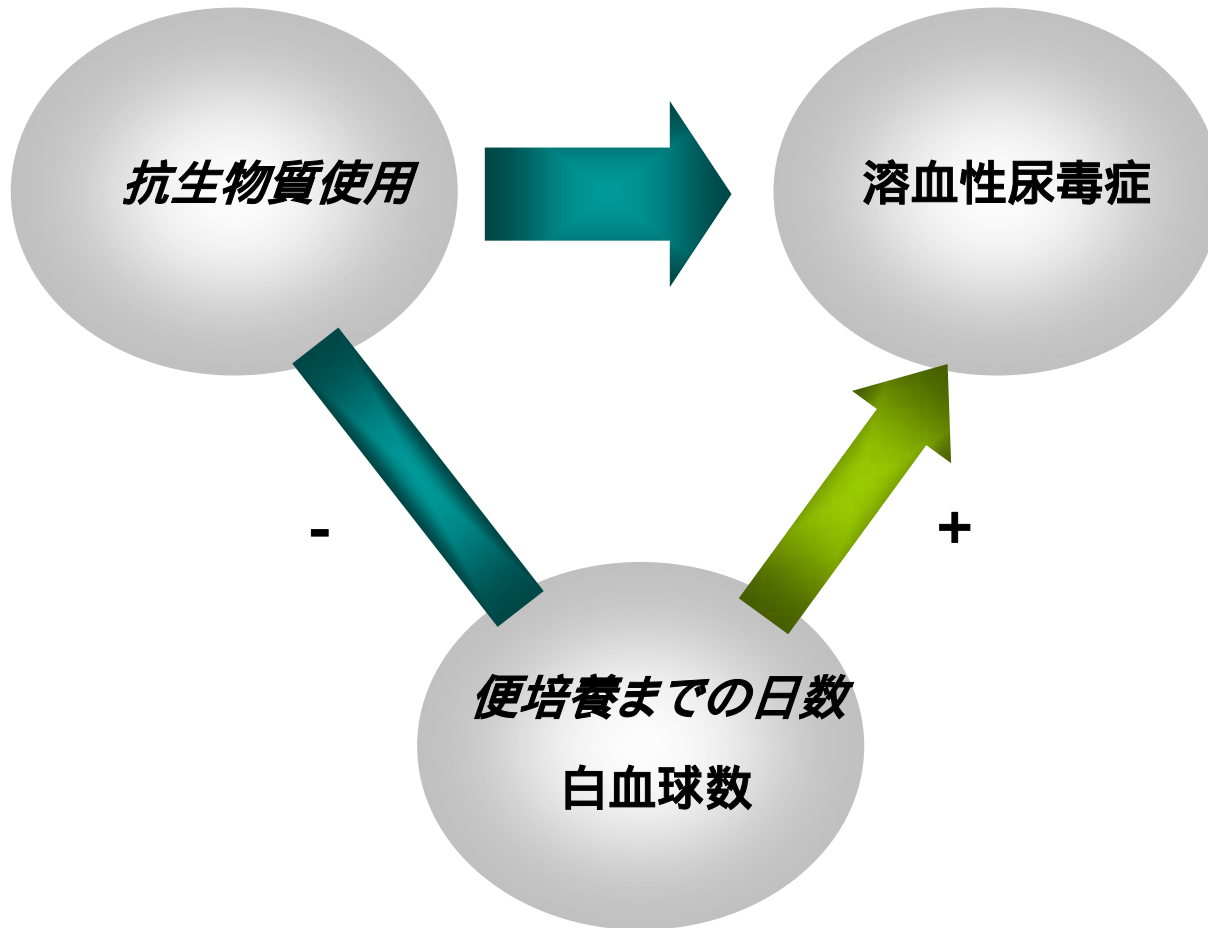


図5 抗生物質使用、便培養までの日数および白血球数、溶血尿毒症候群の関係

## 交絡因子の影響を除去するためには

- 交絡は危険因子であり生物学的特徴として存在するものなので、その関係自体を変えることはできない。しかしながら、暴露と交絡の関係を解析の段階で補正(adjustment)することができることを示した。更に、研究デザインの段階で交絡因子を消去することもできる(図6)。限定(restriction)は結果を得たあとでも可能であるが、他の2つは研究デザインの段階で想定しなくてはならない。
- マッチング
- 限定
- ランダム化
- マッチングは補正よりも強力に交絡の影響を除外してくれる。さらにランダム化は交絡因子が不明であっても交絡因子の影響を打ち消してくれる。先に述べた補正は、交絡と結果との関係を、研究結果を得たあとで補正するものである。よって、ある疾病発生のリスク因子として既に報告があれば、それでマッチングするのも一法であろう。

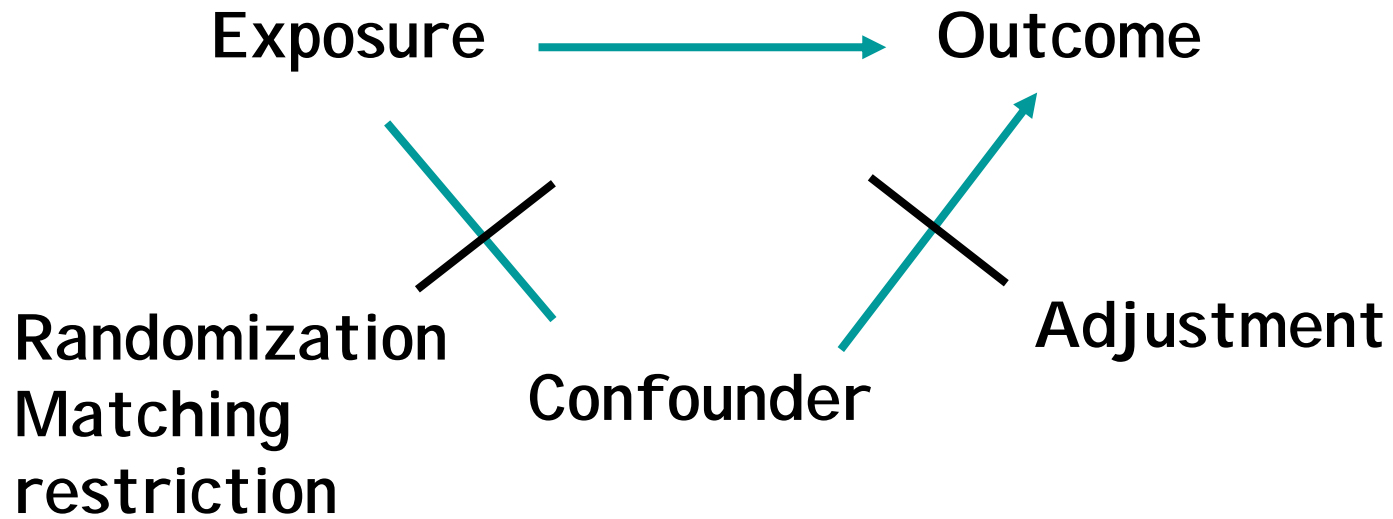


図6 交絡因子(confounders)を打ち消すための方法

## Confounder vs. Effect modification

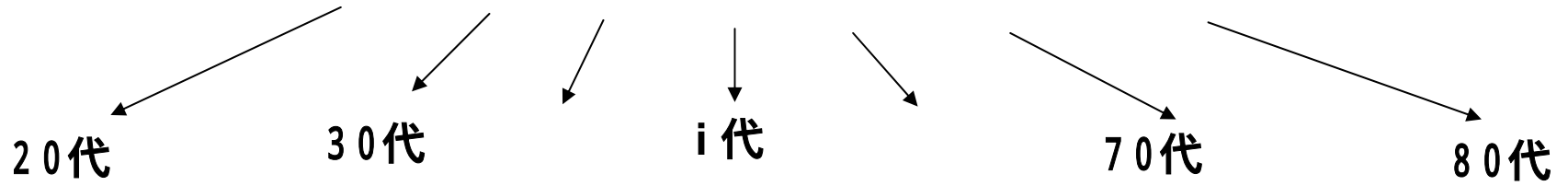
- Confounder (交絡因子) : exposure と関連し、outcome に影響する場合をいう(図1)。
- Effect modification: Effect modifier は層化因子によって結果に及ぼす影響が異なる場合をいう。両者は、exposure, outcome からみれば第三の因子となる(直接因果関係に関与しない)。理論だけをきいてもイメージがわからないと思うので、例を用いて説明したい。

# Crude data vs. Mantel-Haenszel adjustment

- 単純な暴露因子の結果に対するリスク値をcrude data と呼び、交絡因子で補正したリスク値をadjusted data と呼ぶ。復習になるが、ある exposure factor とdisease 発症の関係を調べる際(図7上)、risk ratio (RR), odds ratio (OR)は以下のようなになる。
- $RR_{crude} = (a/N_1) / (c/N_0)$
- $OR_{crude} = ad/bc$
- ある因子、例えば年齢層によって20代、30代、。。70代、の6つの表を上  
の表と同様に創ったとする(図7下)。そうしたとき、adjustment は以下の  
方法によって、簡単な計算機(もちろん筆算でもそろばんでも)できる。こ  
の方法をMantel-Haenszel adjustmentと呼ぶ(MH)。
- $RR_{MH} = [ (a_i)(N_{0i})/T_i ] / [ (c_i)(N_{1i})/T_i ] = [ (w_i)(RR_i) ] / [ (w_i) ]$
- $w_i = (c_i)(N_{1i})/T_i = [T_i][N_{1i}N_{0i}]/(T_iT_i)[c_i/N_{0i}]$
- $OR_{MH} = [ (a_i)(d_i)/T_i ] / [ (b_i)(c_i)/T_i ] = [ (w_i)(OR_i) ] / [ w_i ]$
- $w_i = (b_i)(c_i)/T_i$
- $RR_{crude} = RR_{MH}$ ,  $OR_{crude} = OR_{MH}$  であればその因子は交絡因子ではな  
い。
- $RR_{crude} \neq RR_{MH}$ ,  $OR_{crude} \neq OR_{MH}$  であればその因子は交絡因子である。



	Disease +	Disease -	合計
Exposure +	a	b	$N_1$
Exposure -	c	d	$N_0$
合計	$a + c$	$b + d$	T



	Disease +	Disease -	合計
Exposure +	$a_i$	$b_i$	$N_{1i}$
Exposure -	$c_i$	$d_i$	$N_{0i}$
合計	$a_i + c_i$	$b_i + d_i$	$T_i$

図7 closed cohort study の層化:  
 例えば、全体の表を年代によって分けて作り直してみる。

- 例えば、乳児期人工乳にだけ頼っていると乳児下痢症に罹患しやすいか否かについて研究し以下の結果を得たとする(図8)。

図8 乳児期人工乳にだけ頼っていると乳児下痢症に罹患しやすいか？ パターン1

	下痢あり	下痢なし	合計
人工乳	210	290	500
母乳	90	410	500
合計	300	700	1000

$$RR_{\text{crude}} = (210/500) / (90/500) = 2.33$$

男児

	下痢あり	下痢なし	合計
人工乳	120	80	200
母乳	80	320	400
合計	200	400	600

$$RR_{\text{boy}} = (120/200) / (80/400) = 3.00$$

女児

	下痢あり	下痢なし	合計
人工乳	90	210	300
母乳	10	90	100
合計	100	300	400

$$RR_{\text{girl}} = (90/300) / (10/100) = 3.00$$

- さらに、Mantel Haenszel 法による補正を行う(図9)。
- 結論
- $RR_{crude} \neq RR_{adjusted}$  によって、性別はconfounder (交絡因子)となっている。
- $RR_{boy} = RR_{girl}$  によって、性別はeffect modifier ではない。

図9 MH adjustmentパターン1

boys

	下痢あり	下痢なし	合計
人工乳	120	80	200
母乳	80	320	400
合計	200	400	600

girls

	下痢あり	下痢なし	合計
人工乳	90	210	300
母乳	10	90	100
合計	100	300	400

$$(120 \times 400)/600 + (90 \times 100)/400 \\ = 80 + 22.5 = 102.5$$

$$(80 \times 200)/600 + (10 \times 300)/400 \\ = 26.7 + 7.5 = 34.2$$

$$102.5 / 34.2 = 3.00 = RR_{\text{adjusted}}$$

- 次の例(図10、図11)では、はたして性別はconfounderだろうか、それともeffect modifier だろうか？

図10 乳児期人工乳にだけ頼っていると乳児下痢症に罹患しやすいか?パターン2

	下痢あり	下痢なし	合計
人工乳	210	290	500
母乳	70	430	500
合計	280	720	1000

$$RR_{\text{crude}} = (210/500) / (70/500) = 3.00$$

男児

	下痢あり	下痢なし	合計
人工乳	120	80	200
母乳	40	160	200
合計	160	240	400

$$RR_{\text{boy}} = (120/200) / (40/200) = 3.00$$

女児

	下痢あり	下痢なし	合計
人工乳	90	210	300
母乳	30	270	300
合計	120	480	600

$$RR_{\text{girl}} = (90/300) / (30/300) = 3.00$$

- 結論
- $RR_{crude} = RR_{adjusted}$  よって、性別はconfounder (交絡因子) ではない。
- $RR_{boy} = RR_{girl}$  よって、性別はeffect modifier ではない。



図11 MH adjustmentパターン2

boys

	下痢あり	下痢なし	合計
人工乳	120	80	200
母乳	40	160	200
合計	200	400	400

girls

	下痢あり	下痢なし	合計
人工乳	90	210	300
母乳	30	270	300
合計	120	480	600

$$(120 \times 200)/400 + (90 \times 300)/600$$

$$= 60 + 45 = 105$$

$$(40 \times 200)/400 + (30 \times 300)/600$$

$$= 20 + 15 = 35$$

$$105 / 35 = 3.00 = \text{RR}_{\text{adjusted}}$$

- 次の例(図12、図13)では、はたして性別はconfounderだろうか、それともeffect modifier だろうか？

図12 乳児期人工乳にだけ頼っていると乳児下痢症に罹患しやすいか?パターン3

	下痢あり	下痢なし	合計
人工乳	210	290	500
母乳	70	430	500
合計	280	720	1000

$$RR_{\text{crude}} = (210/500) / (70/500) = 3.00$$

男児

	下痢あり	下痢なし	合計
人工乳	160	40	200
母乳	40	160	200
合計	160	240	400

$$RR_{\text{boy}} = (160/200) / (40/200) = 4.00$$

女児

	下痢あり	下痢なし	合計
人工乳	50	250	300
母乳	30	270	300
合計	80	520	600

$$RR_{\text{girl}} = (50/300) / (30/300) = 1.67$$

- 結論
- $RR_{crude} = RR_{adjusted}$  によって、性別はconfounder (交絡因子) ではない。
- $RR_{boy} \quad RR_{girl}$  によって、性別はeffect modifier である。
- 
- このように交絡因子は補正することにより、crude の値とずれを生じる (10%以上) ことを特徴とし、またそのように定義される。一方、effect modification では補正ししてもcrude の値とずれを生じない。しかし、因子(上の例では性別)によって値がずれる(これを統計学的に証明するにはhomogeneity test で有意差を認める必要がある)。交絡因子の存在するときはeffect modification は存在せず、逆にeffect modification が存在するときは交絡因子を伴わない傾向にある(必ずではない)。交絡因子の場合には補正(前)後の数値を、effect modification の場合にはその因子で層化して、それぞれの値を記載する。交絡因子はその集団の偏りによって生じるもので補正するべきものであり、effect modification は生物学的作用により発生するため、そのまま記載するべきだからである。

図13 MH adjustmentパターン3

boys

	下痢あり	下痢なし	合計
人工乳	160	40	200
母乳	40	160	200
合計	200	200	400

girls

	下痢あり	下痢なし	合計
人工乳	50	250	300
母乳	30	270	300
合計	80	520	600

$$(160 \times 200)/400 + (50 \times 300)/600$$

$$= 80 + 25 = 105$$

$$(40 \times 200)/400 + (30 \times 300)/600$$

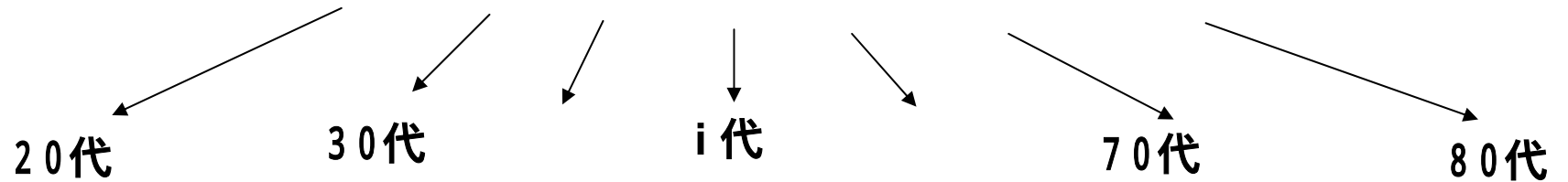
$$= 20 + 15 = 35$$

$$105 / 35 = 3.00 = \text{RR}_{\text{adjusted}}$$

# 層化 Stratification

- 通常のopen cohort study は図14上のような形であらわされる。
- 性別や年齢などの変数によって層化すると、いくつかの似たような表ができあがる。
- $i = 1, \dots, I$  番目のstrata とするとstratum  $i$  は図14下される。
- 1つの臨床研究で交絡因子はしばしば複数考えられる。例えば喫煙の冠動脈疾患による死亡リスクについて調べた場合、結果は当然年齢、性別、コレステロール値に影響されるし、アルコールの量にも影響されるかもしれない。

	exposed	unexposed	total
cases person-time	a $N_1$	b $N_0$	$M_1$ T



	exposed	unexposed	total
cases person-time	$a_i$ $N_{1i}$	$b_i$ $N_{0i}$	$M_{1i}$ $T_i$

図14 opened cohort study の層化:  
例えば、全体の表を年代によって分けて作り直してみる。

- 例えばBritish Doctors Study revised についても図15に示す。これを、単純に合計すると図16になる。全体で見ると喫煙者は非喫煙者に比べ1.72倍の頻度で冠動脈疾患によって死亡している。



## Stratification (open cohort study)

年齡	35-44		45-54		55-64		65-74		75-84	
喫煙	s	ns	s	ns	s	ns	s	ns	s	ns
患者	32	2	104	12	206	28	186	28	102	31
PY	52,407	18,790	43,248	10,673	28,612	5,710	12,663	2,585	5,317	1,462

## British Doctors Study

S: smoker, ns: non-smoker, PY: person-years

- これを、単純に合計すると図16になる。全体で見ると喫煙者は非喫煙者に比べ1.72倍の頻度で冠動脈疾患によって死亡している。

Just simply summarized

	S	ns
患者	630	101
PY	142,247	39,220
IR	44.29	25.79
IRR	1.72	

図16 図15を単純に合計するとincidence rate ratio (IRR)は1.72となる。

- もう一度最初の表に戻って、各年齢群での喫煙によるリスクを計算すると図17となる。各年齢層のIRRに注目すると、年齢が若ければ若い程喫煙による冠動脈疾患死亡率が高くなっている。とてもはっきりしたトレンドとしてつかむことができる。よって年齢は冠動脈疾患による死亡のeffect modifierといえる。
- 次に年齢はconfounder となっているのだろうか？もしそうだとすると、交絡因子を打ち消して正しい値を得るにはどうしたらよいだろうか？
- $\chi^2 = (O - E)^2 / \text{var}$
- O: observed, E: expected, var:variance
- が基本である。
- Open cohort study において交絡因子が存在しないと仮定する。
- $\chi^2 = [X - E(X/H_0)]^2 / \text{var}(X/H_0)$
- X = 暴露された中での患者数 = a
- $E(X/H_0)$  = confounder が存在しない状況下で暴露された中から発生することが期待される患者数 =  $M_1(N_1/T)$
- $\text{Var}(X/H_0) = M_1(N_1/T)(1 - N_1/T) = M_1N_1N_0/T_2$
- 全体を統合した表(crude table)から計算すと、
- $= [630 - (142247 \cdot 731/181467)]^2 / (731 \times 142247 \times 39220/1814672)$   
= 26.24
- となる。

- さて交絡因子を打ち消すために層化したが、それぞれの表を合算しなくてはならない。基本公式はcrude tableのものと類似している。
- $$\chi^2 = \sum [ X_i - E_i (X_i / H_0) ]^2 / \text{vari} (X_i / H_0)$$
- は最初から1個の表を全部合算するという意味である。
- $X_i =$  暴露された中での患者数  $= a_i$
- $E_i(X_i/H_0) =$  confounder が存在しない状況下で暴露された中から発生することが期待される患者数  $= M_{1i}(N_{1i}/T_i)$
- $\text{Vari}(X_i/H_0) = M_{1i}(N_{1i}/T_i)(1 - N_{1i}/T_i) = M_{1i}N_{1i}N_{0i}/T_i^2$
- 自由度は1である。
- $H_0$ : 喫煙と冠動脈疾患死亡率に関連がない。
- $H_A$ : 喫煙と冠動脈疾患死亡率に関連がある。
- $X_i$  630

Effect Modification  
in Stratified Table

年齢	35-44		45-54		55-64		65-74		75-84	
喫煙	s	ns	s	ns	s	ns	s	ns	s	ns
患者	32	2	104	12	206	28	186	28	102	31
PY IR	52,407 6.11	18,790 1.06	43,248 24.1	10,673 11.2	28,612 72.0	5,710 49.0	12,663 146.9	2,585 108.3	5,317 191.8	1,462 212.0
IRR	5.76		2.15		1.47		1.36		0.90	

IR: incidence rate, IRR: incidence rate ratio

図 17 個々の年代層においてIR: incidence rate, IRR: incidence rate ratioを算出。

- $E_i(X_i/H_{i0})$  と  $Vari(X_i/H_{i0})$  を表から計算する (図 18)。
- さらに、各年代層の計算値を合算する。
- $E_i(X_i/H_{i0}) = 595.2$
- $Vari(X_i/H_{i0}) = 110.1$
- よって
- $\chi^2 = \frac{[X_i - E_i(X_i/H_{i0})]^2}{Vari(X_i/H_{i0})} = \frac{(630 - 595.2)^2}{110.1} = 11.02$
- $Pr(\chi^2 > 11.02) = 0.001$
- $H_0$ は棄却され、喫煙と冠動脈疾患死亡率の間に関連があると結論できる。もちろん年齢についてはcontrolしているが(より正確にcontrolするには10年でなく、1年単位で区切る方が正確になるがブランクのマスができたりして効率的ではない。よって適当に区切ると交絡因子を完全には除去しきれないことになる:residual confounder)、さらに年齢層を10年でなく更にこまかくすることもできるのであるから、交絡因子が残存していないとも限らない(residual confounder)。また他の交絡因子(飲酒、性別など)、選択バイアス、観察バイアスが存在しないことが前提となる。

- 年齢で層化していない場合のc 2は26.23であったのに対して、年齢(交絡因子)で層化した後のc2は11.02と大分小さくなっている。しかし年齢という交絡因子の存在を度外視しても、喫煙は冠動脈疾患の発生を有意に増加させるといえる。そして年齢は交絡因子としてcrude data を過大評価させていたことになる。であるからこの場合、補正(adjust)すると値は小さくなる。
- さてそれでは何倍に増加させるのだろうか？
- Effect modification 自体重要な所見であるので、多くは層化したまま報告される。また明らかに差のある若年層だけとる研究者もいるかもしれない(restriction)。しかしその場合多くのデータをゴミ箱に捨てることになり賢明とはいえない。交絡因子(年齢)を除外した際の喫煙が冠動脈疾患死亡率を全体で何倍に押し上げるのかを知りたいところである。
- これを計算するためにはそれぞれの年齢層の表を合算しなくてはならない。単純に合算してもよいものだろうか？例えば高齢者のPYは小さくなっているが、均等に合算すると高齢者のIRRの比率(weight)が過大評価されてしまう。それではPYの比率(weight)に従って分配したらよいだろうか。そうすると若い世代は心筋梗塞患者数が少ないのにこれを反映しないことになってしまう。Inverse variance weights  $\{w_i = 1/\text{var}[\ln(\text{RRI})] = 1/(1/a_i + 1/b_i)\}$ でもよいのであるが、a or b が0だとweight も0になってしまうためそのstrata のデータは捨てなくてはならない。よってMantel-Haenszel weights がそれぞれの表を合算する際、どの程度の重みをもって評価するか用いられる。



Adjusting to exclude confounding by age

年齢	35-44		45-54		55-64		65-74		75-84	
喫煙	s	ns	s	ns	s	ns	s	ns	s	ns
患者	32	2	104	12	206	28	186	28	102	31
PY IR	52,407 6.11	18,790 1.06	43,248 24.1	10,673 11.2	28,612 72.0	5,710 49.0	12,663 146.9	2,585 108.3	5,317 191.8	1,462 212.0
IRR	5.76		2.15		1.47		1.36		0.90	
$E_i(X_i/H_{i0})$	25.03 ←		93.04		195.07		177.72		104.32	
$Var_i(X_i/H_{i0})$	6.61 ←		18.41		32.45		30.06		22.50	

$$E(X/H_0) = \frac{(32+2) \times 52407}{52407 + 18790} = 25.03$$

$$V(X/H_0) = \frac{(32+2) \times 52407 \times 18790}{(52407 + 18790)^2} = 6.61$$

図18 個々の年代層における期待値、varianceの計算

- $w_i = b_i N_{1i} / T_i$
- しかしながら effect modification が存在しなければどの weight を選択してもかまわない。
- $RR_{MH} = \frac{w_i' R R_i}{w_i'} = \frac{b_i N_{1i} / T_i (a_i / N_{1i}) / (b_i / N_0)}{b_i N_{1i} / T_i}$
- $= (a_i N_{1i} / T_i) / (b_i N_{1i} / T_i)$
- 
- $a_i N_{1i} / T_i = 116.835$
- $b_i N_{1i} / T_i = 82.09$
- $RR_{MH} = (a_i N_{1i} / T_i) / (b_i N_{1i} / T_i) = 1.42$
- 年齢を調整した結果、IRRは1.72より1.42に変化した。他の交絡因子、バイアスが無いものとする、喫煙により冠動脈疾患による死亡率が42%増加することになる。

- 先に p value を出しているが、95% confidence interval (CI) を出してみよう。
- $\ln IRR_{MH} \pm 1.96 \sqrt{\text{var}(\ln IRR_{MH})}$
- $\text{var}(\ln IRR_{MH}) = A/BC$
- $A = \frac{M1iN1iN0i}{Ti^2}$
- $B = \frac{aiN0i}{Ti}$
- $C = \frac{biN1i}{Ti}$
- $\text{var}(\ln IRR_{MH}) = A/BC = 0.01149$
- $\ln(1.42) \pm 1.96 \sqrt{0.01149} = (0.141, 0.560)$
- $e(0.141, 0.560) = (1.15, 1.76)$
- **結論**
- 交絡因子やバイアスは、存在を認識できるもの、できないものを含め無限大にあり、adjust したからといってこれらの要素が完全に除去されるわけではない。であるから統計学的結論を述べるときには交絡因子やバイアスが存在しないとする仮定あるいは、どの程度存在するかという考察を論文中述べなくてはならない。喫煙によって15%から76%冠動脈疾患による死亡率の有意な増加をみると95%の信頼をもっていうことができる。最初のIRRは1.72倍であり、年齢というpositive confounding により過大評価されていたことになる。

Calculate at each stratified table

年齢	35-44		45-54		55-64		65-74		75-84	
喫煙	s	ns	s	ns	s	ns	s	ns	s	ns
患者	32	2	104	12	206	28	186	28	102	31
PY	52,407	18,790	43,248	10,673	28,612	5,710	12,663	2,585	5,317	1,462
IR	6.11	1.06	24.1	11.2	72.0	49.0	146.9	108.3	191.8	212.0
IRR	5.76		2.15		1.47		1.36		0.90	
InIRR	0.531		0.093		0.041		0.041		0.042	
$a_i N_{1i} / T_i$	8.445		20.59		34.27		31.53		22.00	
$b_i N_{1i} / T_i$	1.47		9.62		23.34		23.25		24.31	

図19 リスク差からの検討

- 次にrate difference (RD)について検討してみましょう。
- $RD = \sum w_i' RD_i / \sum w_i'$
- $Var(RD) = a/N_1^2 + b/N_0^2 = (aN_0^2 + bN_1^2)/N_1^2N_0^2$
- $w_i' = 1/var_i(RD_i) = N_1^2N_0^2 / (N_0^2a_i + N_1^2b_1)$
- 図19
- $RD = \sum w_i' RD_i / \sum w_i' = 0.00061$
- 交絡因子, バイアスがないものと仮定して、10,000PYあたり6.1人の冠動脈疾患死亡は喫煙によると結論できる。
- $RD \pm 1.96 \sqrt{var(RD)}$
- $Var(RD) = 1 / \sum w_i'$
- $0.00061 \pm 1.96 \sqrt{0.0000000154} = (3.7/10^4 \text{ PY}, 8.5/10^4 \text{ PY})$
- $RD_{crude} = 1.85/1,000\text{PY} (12.4/10^4 \text{ PY}, 24.6/10^4 \text{ PY})$
- $RD_{crude}$ は $RD_{adjusted}$ に比較して随分大きいことが判る。このことは年齢がRDに関してconfounderとして強く働いたことを示している。
- 
-

# Calculate at each stratified table

年齢	35-44		45-54		55-64		65-74		75-84	
喫煙	s	ns	s	ns	s	ns	s	ns	s	ns
患者	32	2	104	12	206	28	186	28	102	31
PY IR	52,407 6.11	18,790 1.06	43,248 24.1	10,673 11.2	28,612 72.0	5,710 49.0	12,663 146.9	2,585 108.3	5,317 191.8	1,462 212.0
IRD	0.000504		0.001280		0.002296		0.003856		-0.002020	
Var(RD) x10 <sup>7</sup>	0.173		1.609		11.104		53.502		181.113	

図19 リスク差からの検討