

# 疫学研究における交絡と効果の修飾

統計数理研究所 佐藤 俊 哉

(1993 年 11 月 受付)

## 1. はじめに

表 1 は心筋梗塞の発生に飲酒が与える影響を調べた仮想的なコホート研究の結果である。左端は対象者全体のデータをまとめたものであり、右の 2 つは対象者全体を喫煙の有無で分類した(層別した)結果である。喫煙者グループと非喫煙者グループの表を足すと全体の表になっていることが確認できる。「飲酒なし」のグループに比べて「飲酒あり」のグループの心筋梗塞発生割合が何倍になっているのか、リスク比を用いて飲酒の効果を測ると、全体では 2 倍であり「飲酒により心筋梗塞発生のリスクが 2 倍になる」と解釈できる。ところが喫煙者グループと非喫煙者グループではリスク比はともに 1 で、「喫煙者グループでは飲酒により心筋梗塞のリスクは増加しない」、また「非喫煙者グループでは飲酒により心筋梗塞のリスクは増加しない」と解釈できる。そうすると表 1 の解釈として、

1. 飲酒は心筋梗塞のリスクを 2 倍にする
2. 飲酒は心筋梗塞の発生と無関係
3. 飲酒は心筋梗塞の発生を 2 倍にするが、喫煙の有無別では無関係

の 3 通りが考えられるが、どの解釈が正しいのだろうか? このような一見矛盾した現象は交絡 confounding として疫学では古くから、また統計家の間でも Simpson (1951) のパラドックスとして知られていた。

交絡という現象は一般に「原因(飲酒)と結果(心筋梗塞)を表す要因以外の第 3 の要因(喫煙)の影響により真の因果関係が歪められること」と考えられており、原因と想定されているリスク要因への曝露を受けたグループ(曝露グループ)と曝露を受けなかったグループ(非曝露グループ)の共変量の不均衡 covariate imbalance によって起こると半分誤って認識されている。表 2 は別な対象者で心筋梗塞と飲酒の関係を調べた仮想的な結果である。表 1 では確かに飲酒ありのグループとなしのグループ間で喫煙者の分布が大きく異なっていたかつ交絡が見られた。しかし、表 2 では表 1 と全く同じ喫煙者の分布であるにもかかわらず今度は全体でも、喫煙、非喫煙別にみてもリスク比は 2.0 で、結果は一貫して「心筋梗塞に対する飲酒の効果

表 1. 心筋梗塞と飲酒の関係。

心筋梗塞	全 体			喫 煙			非 喫 煙		
	あり	なし	合計	あり	なし	合計	あり	なし	合計
飲酒あり	100	100	200	96	64	160	4	36	40
なし	50	150	200	36	24	60	14	126	140
リスク比		2.0			1.0			1.0	

表2. 心筋梗塞と飲酒の関係: その2.

心筋梗塞	全 体			喫 煙			非 喫 煙		
	あり	なし	合計	あり	なし	合計	あり	なし	合計
飲酒あり	100	100	200	80	80	160	20	20	40
なし	50	150	200	15	45	60	35	105	140
リスク比	2.0			2.0			2.0		

あり」となっている。交絡とは一体何なのだろうか？ 本論文では、最近の因果推論の発展にともなった交絡の定義と、その定義から導かれる交絡調整の方法についてまとめを行う。(疫学研究における因果推論の考え方に関しては Rothman (1988), Greenland (1990) を参照してほしい。)

交絡の調整は効果の修飾 effect modification が存在するかどうかで方法が異なる。効果の修飾は、効果の指標と交絡変数との(統計的)交互作用に相当するが、交互作用を量的と質的に分けた場合、量的な交互作用のことを指す (Miettinen (1974))。量的な交互作用とはどのサブグループでも一貫して因果的(または逆に予防的)な効果があるが効果の大きさは一定していないことであり、質的な交互作用とはあるサブグループでは因果的な効果があり、別なサブグループでは予防的な効果があるように効果の方向が逆転してしまうことである。効果の修飾がない場合には回帰モデルなどを用いて交絡を調整した共通な効果の指標を求めることができる。一方、効果の修飾が存在する場合には標準化 standardization と呼ばれる方法を用いて曝露の効果に要約する必要がある。効果の修飾がない場合の交絡の調整方法 (Mantel-Haenszel の方法などの層別解析や回帰モデルによる調整) に関しては多くの報告があるので、本論文では主として標準化による交絡の調整を解説する。

以下では、2章で交絡の定義について因果推論の立場から解説し、3章では Miettinen (1972) と Greenland (1982) に従って標準化の意義、特に効果の修飾との関係を述べ、4章では具体的な交絡調整の方法としてノンパラメトリックな標準化による調整、5章でロジスティックモデルを用いた標準化の方法を解説する。6章では交絡変数選択の基本的な考え方を紹介し、7章で実際にモデルを用いた交絡変数の選択、そして最後に8章で注意点を含めたまとめを行う。

## 2. 交絡の定義

交絡の定義として一般に信じられているのは、「ある要因で調整した場合と調整しない場合で効果の指標の大きさが変わる (change in estimate criterion)」というオペレーショナルなものであろう (Kleinbaum et al. (1982), Boivin and Wacholder (1985), Grayson (1987))。効果の指標とは、リスク要因の曝露を受けることで疾病発生がどれだけ増加するのか(あるいは、治療によって疾病がどの程度改善されるか)を定量的に示す指標であり、リスク差、リスク比、オッズ比、疾病発生率の比(ハザード比)などがある (Greenland (1987))。しかし、この交絡の定義では効果の指標として何を用いるか、例えばリスク比を用いるかオッズ比を用いるか、によってある要因が交絡要因であるかどうかの判断が変わってしまう。交絡とは因果推論に結びついたもっと根元的なものであるはずであり、効果の指標として我々が何を選ぶかということに依存してしまっては困る。

喫煙と肺がんとの因果関係を調べることを考えてみよう。東京都足立区に居住する年齢35歳の佐藤俊哉は喫煙者であり、20年後に肺がんになったとする。佐藤俊哉の肺がんは佐藤俊哉の

喫煙が原因だといえるだろうか。もし喫煙者である佐藤俊哉が20年後に肺がんを発生するリスク(確率)が $p_1=0.02$ と測定できたとしよう。この $p_1$ がいくら大きくても「佐藤俊哉の肺がんは喫煙が原因」と結論することはできない。ひょっとしたら佐藤俊哉はもともと肺がんのリスクが高いのかも知れない。そこでさらに、「佐藤俊哉が過去から将来まで全く喫煙していなかったと仮定したとき、20年後に肺がんを発生するリスク $p_0^*$ 」も $p_0^*=0.005$ と測定できたとしよう。佐藤俊哉が喫煙者であったときと喫煙者でなかったときのリスクの差 $p_1-p_0^*=0.015$ が正の値(あるいはリスクの比 $p_1/p_0^*=4$ が1よりも大きい)なので、確かに喫煙によって佐藤俊哉が肺がんになる可能性は増加すると言っていいだろう。しかし、「佐藤俊哉が20年後に肺がんを発生するリスク」などはまず観察できない。因果推論に関する観察できない量を含んだ枠組みが「counterfactualモデル」であり(Rubin(1974), Holland(1989), 佐藤(1993)),  $p_1-p_0^*$ と $p_1/p_0^*$ はそれぞれ因果リスク差, 因果リスク比と呼ばれる。

それでは、特定の集団に関する因果推論を考えてみよう。佐藤俊哉を含む「足立区に居住する年齢30-39歳の男性喫煙者」は、喫煙が原因となって肺がんの発生が増えるだろうか。「足立区に居住する30-39歳の男性喫煙者グループ」が20年後に肺がんを発生する平均リスクを $IP_1=0.02$ としよう。個人レベルの $p_1$ と違って $IP_1$ はこの喫煙者グループを20年間追跡し肺がんの発生数を数え上げることで観察できる。個人レベルの因果推論と同様に、肺がん発生に対する喫煙の効果を調べるためにはもう一つ「足立区に居住する30-39歳の男性喫煙者グループが、もし過去から喫煙していなかった場合」の20年後の平均肺がん発生リスク $IP_0^*$ を知る必要がある。 $IP_0^*=0.005$ であったとすると、集団レベルでの平均因果的効果 average causal effect として、因果リスク差 $IP_1-IP_0^*=0.015$ , 因果リスク比 $IP_1/IP_0^*=4$ が得られ、個人レベルの場合と同様に「足立区に居住する30-39歳の男性喫煙者グループでは喫煙が肺がん発生のリスクを高める原因になっている」と結論していいだろう。

個人レベルと異なって集団レベルでは $IP_1$ (曝露グループの平均リスク)は観察可能となったが、 $IP_0^*$ (曝露グループがもし曝露を受けていなかった場合の平均リスク)は相変わらず観察できない。問題を整理しよう。我々が知りたいことは漠然とした仮説である「喫煙が肺がんのリスクを高める原因になっているかどうか」であるが、この仮説を調べるためにはサンプルを取ってきてそのサンプルにもとづいた推論を行わなければならない。ここではそのサンプルが「足立区に居住する30-39歳の男性喫煙者グループ」である。因果推論を科学的に行うためには適切なコントロールグループが不可欠であり、「足立区に居住する30-39歳の男性喫煙者グループ」の理想のコントロールは「足立区に居住する30-39歳の男性喫煙者グループがもし喫煙をしていなかった場合」である。我々はこの理想のコントロールグループを調べることはできないので、「足立区に居住する30-39歳の男性非喫煙者グループ」という測定可能な現実のコントロールを調べることになる。

コントロールとして「足立区に居住する30-39歳の男性非喫煙者グループ」をとり、この非喫煙者グループの20年後の肺がん発生リスクを $IP_0$ としよう。 $IP_0$ も非喫煙者グループを20年間追跡して肺がん発生数を数え上げることで調べることができる。ようやくリスク差 $IP_1-IP_0$ , リスク比 $IP_1/IP_0$ を求めることができるようになった。このリスク差やリスク比が正の値であったり1より大であったら、「足立区に居住する30-39歳の男性喫煙者グループ」では喫煙が肺がんのリスクを高める原因になっていると結論して、…いいだろうか? 言い換えると、我々が測定できたリスク差やリスク比は、因果リスク差, 因果リスク比を正しく反映しているだろうか。もし $IP_0^*=IP_0$ であればリスク差, リスク比は因果リスク差, 因果リスク比を正しく求めている。しかし、もし $IP_0^*\neq IP_0$ であれば、もはや曝露の効果は正しく測定できない。つまり因果関係が歪められてしまうことになるので、これはまさしく交絡である。

交絡の定義として、「曝露グループが曝露を受けていなかった場合の平均リスク  $IP_0^*$ 」と「非曝露グループの平均リスク  $IP_0$ 」が等しくない場合交絡あり，等しい場合交絡なし，を用いることにしよう。この定義は理想のコントロールと現実のコントロールが食い違っているかどうかにもとづいているので「コントロールグループの十分性にもとづく交絡の定義」ということができる (Greenland and Robins (1986), Wickramaratne and Holford (1987))。この定義は観察不可能な量  $IP_0^*$  を含んでいるので一見奇妙な定義であるが、「観察不可能な量を含んでいる」ことから「交絡がないことを保証することはできない」ことが分かる。ランダム化臨床試験では要因のランダム割り付けを行うが，この定義からランダム化できえも交絡のない可能性を高めているだけであって，一回一回の試験では交絡がないことを保証しない。疫学的観察研究では曝露がランダムに割り付けられているとはとうてい考えられないので（自然はそれほど親切ではない），交絡は必ずあると考えた方がよい。

ここまでの交絡に関する議論は共変量とか交絡要因といったものとは無関係であった。しかし，実際に交絡の調整ということになると，交絡要因についても定義が必要である。今，非曝露グループの平均リスクをある要因  $C$  で調整した値を  $IP_{0A}$  としよう。調整は後述するように層別解析によってでも，回帰モデルによってでもかまわない。もし，この  $IP_{0A}$  が  $IP_0$  よりも理想のコントロールグループの  $IP_0^*$  に近い場合， $C$  を交絡要因と定義しよう。コントロールグループの十分性の定義から，要因  $C$  が交絡要因であるためには曝露がない場合でも 2 グループ間の平均リスクが異なるのであるから，

1. 要因  $C$  は対象としている疾病のリスク要因でなければならない，

また曝露グループと非曝露グループが最も異なっているのは，実際に曝露を受けたかどうかであるから，

2. 要因  $C$  は曝露と関連していなければならない，

ことが演繹的に導かれる。この 2 点に因果推論の一般的な議論からの，

3. 要因  $C$  は曝露-疾病間の因果連鎖の中間変数であってはならない，

を加えた交絡の必要条件は長い時間をかけて経験的に導かれたが (Breslow and Day (1980), Miettinen and Cook (1981), Rothman (1986))，コントロールグループの十分性の定義からこのように簡単に導くことができる。

この交絡の必要条件から表 1 を解釈すると，喫煙は心筋梗塞のよく知られたリスク要因であること，かつ飲酒者の喫煙割合は高いと予測できることから，飲酒と心筋梗塞の関連を調べる場合喫煙は交絡要因となっている。したがって，喫煙で層別した結果「飲酒と心筋梗塞は関連はない」が正しい解釈となる。（この交絡の有無の解釈はデータにもとづいていないことに注意してほしい。）

### 3. 交絡と標準化

標準化という用語にはなじみがない読者でも，「調整死亡率」，「訂正死亡率」という用語は疾病地図とともに新聞等で見たことがあるだろう。死亡状況の地域比較を行う場合，地域の特性に応じて「過疎化の影響により老人人口の割合が高い地域」，「生産者人口の多い地域」など年齢分布の違いが生じる。年齢は死亡の非常に強いリスク要因であり，かつ地域によって年齢の分布が異なっているので，年齢は死亡状況を比較する際に交絡要因となっている。したがって，

表3. 交絡, counterfactual, SMR.

曝露	全 体		層 1		層 2	
	疾 病	標本数	疾 病	標本数	疾 病	標本数
E	$x$	$n$	$x_1$	$n_1$	$x_2$	$n_2$
E[U]	$x^*$	$n$	$x_1^*$	$n_1$	$x_2^*$	$n_2$
U	$y$	$m$	$y_1$	$m_1$	$y_2$	$m_2$

死亡率の地域比較を行う際には年齢分布が異なることによる影響（交絡）を除去して、正しい比較を行わなければならない。

調整死亡率は、正確には「年齢の影響を調整した死亡率」であり、「標準化」という考え方をを用いて年齢の影響を調整したものである。標準化の方法は、調整死亡率あるいは標準化死亡比 (Standardized Mortality Ratio, SMR) による地域比較に最もよく用いられるため、標準化に関する議論も「複数の死亡率の比較」という観点から行われることが多い (例えば, Fleiss (1982), 柴田 (1985), 福富・橋本 (1989), 新田・佐藤 (1989)). そのためどうしても「標準化した死亡率で複数の地域 (集団) 間の比較を行ってよいか」というところに議論が行きがちである。しかし、交絡の除去に他の方法ではなくなぜわざわざ標準化を行うのか、標準化という考え方の根本的な意味についてはほとんど論じられていない。

表3はリスク要因への曝露と疾病の発生についてのコホート研究全体の結果と層別した結果である (表中の記号は期待される値を表す)。E は曝露グループ、U は非曝露グループ、E[U] は「曝露グループのメンバーが曝露を受けなかった場合」とする。E[U] グループは観察不可能な counterfactual である。曝露グループの疾病発生割合は  $IP_1 = x/n$ 、曝露グループが曝露を受けなかった場合の疾病発生割合は  $IP_0^* = x^*/n$  であり、例えば両者の比をとった真のリスク比  $IP_1/IP_0^* = x/x^*$  は「曝露グループのメンバーが曝露を受けたことによる疾病発生の比例的 (相対的) な増加」という意味を持つ。前章で述べたように、交絡が存在しなければ非曝露グループの疾病発生割合  $IP_0 = y/m$  が  $IP_0^*$  と等しいから  $x^* = n(y/m)$  となり、 $x/x^* = mx/ny$  と真のリスク比と単純な解析でのリスク比が一致する。

「コントロールグループの十分性の定義」から交絡があるということは  $IP_0^* \neq IP_0$  であるから、観測不可能な  $x^*$  を別な情報から求めなければならない。ここで交絡要因によりデータをいくつかの層に分けることによって、それぞれの層のなかでは交絡が無視できたと仮定しよう。各層での非曝露グループの疾病発生割合を  $IP_{01}$ ,  $IP_{02}$  とすると、層内では交絡がないのであるから  $x_1^* = n_1 IP_{01} = n_1(y_1/m_1)$ ,  $x_2^* = n_2 IP_{02} = n_2(y_2/m_2)$  であり、明らかに  $x^* = x_1^* + x_2^*$  より  $x^* = n_1(y_1/m_1) + n_2(y_2/m_2)$  として求められるので、交絡を調整したリスク比は、

$$\frac{x}{x^*} = \frac{x}{n_1 \frac{y_1}{m_1} + n_2 \frac{y_2}{m_2}},$$

で与えられる (Greenland and Robins (1986)). 上式右辺は「曝露グループで実際に観察された疾病発生数 (observed)」と「曝露グループが曝露を受けなかった場合に期待される疾病発生数 (expected)」の比であり、これは古くから知られている 観察発生数/期待発生数 である SMR (Standardized Morbidity/Mortality Ratio) に他ならない。SMR は保険統計で古くから用いられ、また産業衛生の分野では日常的に解析に用いられているリスク評価の指標である (Keiding (1987), Checkoway et al. (1989)). 分母の「期待発生数」はあくまでも「曝露グルー

ブが曝露を受けなかった場合」の期待値であり、帰無仮説（曝露の効果はない）のもとでの期待値ではないことに注意が必要である (Miettinen (1972)).

#### 4. 標準化の方法

交絡要因のレベルで層別した第  $k$  層で、曝露グループと非曝露グループの疾病発生数、サンプル数をそれぞれ  $(x_k, y_k)$ ,  $(n_k, m_k)$  とする。標準化という名称は、適切な標準集団 (standard or reference population) を定めて、曝露グループと非曝露グループの交絡要因の分布が標準集団  $R$  の交絡要因の分布  $r_k$  と同じだと仮定した場合に期待される疾病発生数、

$$E_{R[E]} = \sum_k r_k \frac{x_k}{n_k}, \quad E_{R[U]} = \sum_k r_k \frac{y_k}{m_k},$$

の比較を行うことに由来する。ここで、 $\sum_k$  は全ての層にわたって和を取ることを表し、 $A[B]$  は  $A$  という標準集団の曝露状況が  $B$  だった場合を表す。SMR は標準集団として曝露グループを用いていたことになる。標準集団として用いる集団は、どんな集団でも良いというわけではない。疫学的に意味のある標準集団としては、我々が調べたい仮説に応じて次の3つがある (Greenland (1982, 1991b)).

禁煙による肺がんの予防効果を調べるために、禁煙者グループと喫煙者グループを比較する観察研究を考えよう。この場合曝露は禁煙であり、喫煙し続けることが非曝露に相当する。禁煙するかどうかをランダムに割り付けていない以上、禁煙者グループは何らかの理由があって禁煙したと考えた方が安全である。我々の調べたい仮説の設定が「自発的に喫煙を止めた禁煙者のグループでは、禁煙により肺がんの死亡が減る」であれば、

- 1) 曝露 (禁煙) グループ  $E$  を標準集団とする:  $r_k \equiv n_k$ ,

$$E_{E[E]} \equiv \sum_k x_k = O_E, \quad E_{E[U]} = \sum_k n_k \frac{y_k}{m_k}.$$

また、仮説が「(禁煙できなかった) 喫煙者のグループでは、強制的な禁煙により肺がんの死亡が減る」であれば、

- 2) 非曝露 (喫煙) グループ  $U$  を標準集団とする:  $r_k \equiv m_k$ ,

$$E_{U[E]} = \sum_k m_k \frac{x_k}{n_k}, \quad E_{U[U]} \equiv \sum_k y_k = O_U.$$

もし、禁煙するかどうかをランダム化した介入試験 intervention trial を実施できた場合、禁煙をランダム割り付けした後のグループについて上の2つの仮説を考えることができる。しかしこの場合、より適切な仮説は「ランダム化の対象となった喫煙者全体で、禁煙により肺がんの死亡が減る」であろうから、

- 3) 全体  $P$  を標準集団とする:  $r_k \equiv N_k$  ( $N_k = n_k + m_k$ ),

$$E_{P[E]} = \sum_k N_k \frac{x_k}{n_k}, \quad E_{P[U]} = \sum_k N_k \frac{y_k}{m_k}.$$

このように標準化を行う際には、我々が推測を行いたい標的集団 target population を適切に反映した標準集団を選ばなければならない。

2章, 3章では交絡をコントロールグループ (非曝露グループ) の十分性で定義したが、仮説

の設定にしたがって若干定義を修正する必要がある。上記3つの仮説設定を考えるとコントロールグループの十分性は1)の場合の交絡の定義となる。2)の場合には「非曝露グループが曝露を受けた場合の平均リスク」と「曝露グループの平均リスク」が等しいかが交絡の定義となる。3)の場合にはもう少し複雑で、「曝露グループが曝露を受けなかった場合の平均リスク」と「全体が曝露を受けなかった場合の平均リスク」が等しく、かつ「非曝露グループが曝露を受けた場合の平均リスク」と「全体が曝露を受けた場合の平均リスク」が等しい場合「交絡なし」、そうでなければ「交絡あり」とした定義となる。

#### 4.1 標準化リスク比 Standardized Risk Ratio

これら3通りの標準化の方法によって得られた  $E_{R[E]}$ ,  $E_{R[U]}$  の比が、標準化リスク比  $SRR_R = E_{R[E]}/E_{R[U]}$  であり、それぞれ

- 1) 曝露グループが曝露を受けたことによる疾病発生の比例的な増加,
- 2) 非曝露グループが曝露を受けることによる疾病発生の比例的な増加,
- 3) グループ全体が曝露を受けることによる疾病発生の比例的な増加,

を表す。区別が必要な場合には、曝露グループでの標準化リスク比 (*SRR in the exposed*,  $SRR_E$ , or *SMR*), 非曝露グループでの標準化リスク比 (*SRR in the unexposed*,  $SRR_U$ ), 全体での標準化リスク比 (*SRR in the population*,  $SRR_P$ ) と呼ぶ。この定義では各層ごとのリスク比に特別な仮定 (例えば「リスク比が層を通じて一定」など) を置いていないことに注意してほしい。ここで比較しているのは、曝露グループ、非曝露グループ、あるいはグループ全体が「曝露を受けた場合と曝露を受けなかった場合の疾病発生割合」である。層とリスク比との間にどのような交互作用が存在しても標準化リスク比のこの解釈は成立する。そうはいっても質的な交互作用 (ある層では原因的な効果があり、別な層では予防的な効果がある) がある場合にはそのような解釈に医学的な意味があるかどうかは疑問である。しかし、交互作用は存在するが全ての層で効果の方向がそろっている場合、つまり効果の修飾が存在する場合でも、標準化リスク比は医学的に意味のある指標となる。

$\log(SRR_R)$  の近似的な分散の推定値は Greenland (1982) が,

$$V_G(\log SRR_R) = \frac{\sum_k E_{R[E]k}^2 \left( \frac{1}{x_k} - \frac{1}{n_k} + \frac{1}{y_k} - \frac{1}{m_k} \right)}{E_{R[E]}^2},$$

を与えた。しかしこの推定値には  $x_k, y_k$  がゼロであった場合の補正等に問題があり、次の推定値 (Rothman (1986)),

$$V(\log SRR_R) = \frac{\sum_k E_{R[E]k} (r_k - E_{R[E]k}) / n_k}{E_{R[E]}^2} + \frac{\sum_k E_{R[U]k} (r_k - E_{R[U]k}) / m_k}{E_{R[U]}^2},$$

の方が性質がよいと期待できる。ここで、 $E_{R[E]k} = r_k(x_k/n_k)$ ,  $E_{R[U]k} = r_k(y_k/m_k)$  である。付録1にこれらの推定値の導出を示した。これより、真の  $SRR_R$  の 95% 信頼区間は、

$$SRR_R \exp[\pm 1.96 \times \sqrt{V(\log SRR_R)}],$$

となる。

例. 表4は喫煙と冠動脈性疾患 (CHD) 発症との関係を調べた仮想的なコホート研究の結果である。CHDのローリスクグループとハイリスクグループでリスク比は3.75, 1.95とほぼ2

表4. 喫煙と冠状動脈性疾患 (CHD) の発生.

CHD 発生	ローリスク			ハイリスク		
	あり	なし	標本数	あり	なし	標本数
喫煙	10	30	40	52	18	70
非喫煙	4	56	60	19	31	50
リスク比		3.75			1.95	
リスク差		0.183			0.363	
オッズ比		4.67			4.71	

倍異なっているので効果の修飾が疑われる。このデータで曝露グループを標準集団とした場合 SMR を求めると、 $SMR=2.12$ ,  $V(\log SMR)=0.0343$ , 95% 信頼区間 (1.47, 3.05) となる。非曝露グループを標準集団とすると値は若干異なり、 $SRR_U=2.27$ ,  $V(\log SRR_U)=0.0380$ , 95% 信頼区間 (1.55, 3.32) である。

#### 4.2 標準化リスク差 Standardized Risk Difference

標準化リスク比では、曝露を受けることによる疾病発生の相対的な増加を調べることができたが、疾病発生の絶対的な増加を調べたい場合もあるだろう。曝露グループと非曝露グループの交絡要因の分布が標準集団と同じ場合の疾病発生割合は  $E_{R[E]}/r$ ,  $E_{R[U]}/r$  であるから (ただし、 $r=\sum_k r_k$ )、標準化リスク差とその近似的な分散の推定値は、

$$SRD_r = \frac{E_{R[E]} - E_{R[U]}}{r},$$

$$V(SRD_r) = \frac{1}{r^2} \sum_k \left[ \frac{E_{R[E]k}(r_k - E_{R[E]k})}{n_k} + \frac{E_{R[U]k}(r_k - E_{R[U]k})}{m_k} \right],$$

で与えられる (Rothman (1986), 付録1 参照)。

例 (続き). 表4のデータで、疾病発生割合の差はローリスクが 0.183, ハイリスクが 0.363 とやはり 2 倍程度異なっている。そこで、曝露グループを標準集団とした標準化リスク差を計算すると、 $SRD_E=0.298$ ,  $V(SRD_E)=0.00377$ , 95% 信頼区間 (0.177, 0.418) を得る。

#### 5. ロジスティックモデルを用いた標準化

4章の標準化の方法は疾病発生割合に何の仮定も置かない、いわばノンパラメトリックな標準化であった。もし、データに対して何らかのモデルがよく当てはまる場合には、当てはめたモデルのもとで計算される期待疾病発生数を用いて標準化を行うことができる。このように、モデルでスムージングした標準化は、モデルで用いている仮定が正しければ、ノンパラメトリックな標準化よりも推定精度がよくなることが期待できる (Greenland (1991a), Greenland and Holland (1991)). 本章では、カテゴリカルデータ解析でよく用いられるロジスティックモデルを用いて標準化リスク比、標準化リスク差を推定する方法を述べる。

疾病発生リスクのモデルとしてはいくつかのものが考えられるが (佐藤 (1993)), ここでは観察されたデータにロジスティックモデル、



$$\log \frac{r(\mathbf{x})}{1-r(\mathbf{x})} = \alpha + \beta x + \boldsymbol{\gamma}' \mathbf{z},$$

がよく当てはまったとしよう。ただし、 $r(\mathbf{x})$  は共変量ベクトル  $\mathbf{x} = (1, x, \mathbf{z}')$  を持つ個人が観察期間中に疾病を発生するリスク (確率) を表し、 $x=1$  (曝露あり)、 $0$  (曝露なし)、 $\mathbf{z}$  は交絡変数の組とする。ロジスティックモデルは、 $r(\mathbf{x})$  が必ず  $0$  から  $1$  の間に入る、二項分布モデルのもとでの canonical link である、など統計モデルとしてはたいへん便利な性質を持っている。ただし、ロジスティックモデルから求められる疾病オッズ比  $e^d$  は、疾病の発生がまれな場合は近似的にリスク比として解釈できるが、まれでない場合に疫学的な解釈は困難である (Greenland (1987))。しかし、ロジスティックモデルの当てはまりが良いということは共通オッズ比の仮定が成立することを意味するので、共通リスク比や共通リスク差の仮定は満たされない (リスク比、リスク差と交絡要因との間に効果の修飾が存在する)。このため、共通リスク比、共通リスク差の仮定にもとづく回帰モデルは誤ったモデルになってしまう。したがって、ロジスティックモデルを当てはめた結果を、疫学的な解釈が可能な標準化リスク比や標準化リスク差に変換してレポートする必要がある。

ロジスティック回帰係数  $\boldsymbol{\beta} = (\alpha, \beta, \boldsymbol{\gamma}')$  の最尤推定値を  $\mathbf{b} = (a, b, \mathbf{c}')$  とし、 $\mathbf{b}$  の分散共分散行列の推定値を  $\mathbf{V}_b$  とする。また推定された回帰係数を用いて、第  $i$  番目の個人が曝露を受けたかどうかに関わらず強制的に  $x=1$  とした場合と  $x=0$  とした場合の期待リスクを、

$$P_{1i} = \frac{\exp(a + b + \mathbf{c}'\mathbf{z}_i)}{1 + \exp(a + b + \mathbf{c}'\mathbf{z}_i)}, \quad P_{0i} = \frac{\exp(a + \mathbf{c}'\mathbf{z}_i)}{1 + \exp(a + \mathbf{c}'\mathbf{z}_i)},$$

を用いて表す。また、対象者を曝露グループのメンバー  $n$  人についての和をとる場合  $\Sigma_E$ 、非曝露グループのメンバー  $m$  人についての和をとる場合  $\Sigma_U$ 、対象者全体  $n+m=N$  人の和をとる場合  $\Sigma_P$  で示すことにする。ロジスティックモデルを用いた曝露グループの期待平均リスクと曝露グループが曝露を受けなかった場合の期待平均リスクは、

$$IP_{1E} = \frac{1}{n} \Sigma_E P_{1i}, \quad IP_{0E} = \frac{1}{n} \Sigma_E P_{0i},$$

と書ける。また、非曝露グループが曝露を受けた場合と非曝露グループの期待平均リスクは、

$$IP_{1U} = \frac{1}{m} \Sigma_U P_{1i}, \quad IP_{0U} = \frac{1}{m} \Sigma_U P_{0i},$$

対象者全員が曝露を受けた場合と曝露を受けなかった場合の期待平均リスクは、

$$IP_{1P} = \frac{1}{N} \Sigma_P P_{1i}, \quad IP_{0P} = \frac{1}{N} \Sigma_P P_{0i},$$

となる。

これらの期待平均リスクを用いて、ロジスティックモデルでスムージングした曝露グループでの標準化リスク比である SMR の推定値  $SMR^* = IP_{1E}/IP_{0E}$ 、曝露グループでの標準化リスク差の推定値  $SRD_E^* = IP_{1E} - IP_{0E}$ 、などを求めることができる。また、 $\log SMR^*$ 、 $SRD_E^*$  の近似的な分散の推定値は、

$$V(\log SMR^*) = \mathbf{R}_E' \mathbf{V}_b \mathbf{R}_E, \quad V(SRD_E^*) = \mathbf{D}_E' \mathbf{V}_b \mathbf{D}_E,$$

ただし、

$$R_E = \frac{n}{IP_{1E}} \left[ \sum_E \frac{e^{b'x_{1i}}}{(1+e^{b'x_{1i}})^2} x_{1i} \right] - \frac{n}{IP_{0E}} \left[ \sum_E \frac{e^{b'x_{0i}}}{(1+e^{b'x_{0i}})^2} x_{0i} \right],$$

$$D_E = \frac{1}{n} \sum_E \left[ \frac{e^{b'x_{1i}}}{(1+e^{b'x_{1i}})^2} x_{1i} - \frac{e^{b'x_{0i}}}{(1+e^{b'x_{0i}})^2} x_{0i} \right],$$

$$x_{1i} = (1, 1, z_i)', \quad x_{0i} = (1, 0, z_i)'.$$

これらの分散の導出は付録2に示した。

例(続き). 表4ではローリスクグループとハイリスクグループでリスク比, リスク差ともに2倍程度異なっていたが, オッズ比はそれぞれ4.67, 4.71とほぼ等しいのでロジスティックモデルを用いた標準化を行ってみよう. ロジスティックモデルをあてはめた結果は,

$$\log \frac{r(x, z)}{1-r(x, z)} = -2.64 + 1.55x + 2.16z,$$

ただし,  $x=1$ : 喫煙,  $0$ : 非喫煙,  $z=1$ : ハイリスク,  $0$ : ローリスクである. また, 回帰係数( $a, b, c$ )の分散共分散行列の推定値は,

$$V_b = \begin{pmatrix} 0.1402 & -0.0764 & -0.0996 \\ -0.0764 & 0.1143 & 0.0157 \\ -0.0996 & 0.0157 & 0.1310 \end{pmatrix},$$

となった. これらから, ロジスティックモデルにもとづいたSMRは $SMR^*=2.12$ ,  $V(\log SMR^*)=0.0332$ , 95%信頼区間(1.48, 3.01)となる. また, ロジスティックモデルにもとづいた曝露グループの標準化リスク差は $SRD_E^*=0.297$ ,  $V(SRD_E^*)=0.00369$ , 95%信頼区間(0.178, 0.417)となる. ノンパラメトリックな標準化では $V(\log SMR)=0.0343$ ,  $V(SRD_E)=0.00377$ であったので, わずかではあるが推定値の分散を小さくしている.

## 6. 交絡の調整

単一の曝露が疾病発生の原因となっているかどうか, その因果関係が問題となっているとしよう. 全ての交絡要因で調整した効果の推定値は, 仮定した[曝露, 交絡要因, 疾病]間の関係が正しいときバイアスは入らないが, 不必要に効果の推定値の分散が大きくなってしまふ. 逆に重要な交絡要因を調整し損なうと推定値にバイアスが加わってしまう. したがって, 我々はこの交絡変数を調整すべきで, どの交絡変数を調整すべきではないか, という古典的な変数選択のジレンマに直面する (Robins and Greenland (1986)). 「交絡要因で調整する」とは, 層別解析であれば層別変数に加えること, 回帰モデルを用いた場合にはモデルに加えることを表す. ただし, 曝露の効果を調べることが主目的であるため, 回帰モデルを用いた場合に曝露変数は変数選択の対象とはせず, いつでも強制的にモデルに含めることにする.

### 6.1 有意性検定

ある要因を調整すべきかどうかを判断する手がかりとして, 有意性検定を用いる方法がある (Fleiss (1986)). 例えば, 曝露グループと非曝露グループ間の背景因子のt検定や $\chi^2$ 検定, 回帰モデルにおいて要因の係数がゼロであるかどうかの検定などで有意であれば調整し, 有意でなければ調整しない, という方法がよく使われる. 有意性検定は交絡要因の同定に使ってはいけない (Kleinbaum et al. (1982), Rothman (1986)), といわれているながらも未だに利用されている.

コントロールグループの十分性の定義から、交絡がないことをサンプルだけからは決めることはできないし、交絡がないことを保証するような手続きはない (Greenland and Robins (1986), Wickramaratne and Holford (1987))。繰り返すが、たとえ曝露の割り付けをランダム化できても交絡がない可能性を高めるだけであって (平均的には交絡はないはず)、ランダム割り付けさえも一回一回の研究について交絡なしと主張する根拠にはならない。ましてや観察研究においてはほぼ確実に交絡はあると考えた方がいい。

さて、有意性検定の論理は「交絡の存在を示すために、まず交絡はないという仮定をする。その仮定のもとでは起きる可能性が非常に小さいことが起こった場合に交絡ありと判断する」ものであった。従って、観察研究ではほぼ確実に交絡が存在することが分かっているので、明らかにこの論理はそぐわない。強いて仮説検定を用いるとすれば正当な論理は同等性検定であり、「観察研究では、交絡は必ず存在するけれども、その影響が無視できる程度であればかまわないので『この程度なら無視していい』という範囲を決めて、交絡があるという仮定のもとで求めた統計量が、その指定した範囲に入った場合交絡の影響は無視できる、と判断する」という考え方を用いる必要がある (Greenland (1989a))。

ランダム割り付けを行った場合には、交絡 (あるいは共変量の不均衡) をを見つけるための有意性検定は別な理由で無意味である。なぜなら、この場合には平均的には交絡がないことを我々は知っているので「交絡がないと仮定し、その仮定のもとでは起きる可能性が非常に小さいことが起こった場合」、それはランダム化のプロセスにしたがって、起きる可能性が小さいことが本当に起きているのである (Begg (1990) はパラメータの推定値にバイアスが入ることから有意性検定を用いることの不適切さを述べている)。

## 6.2 推定値の変化による基準

ある変数が交絡要因であるかどうかを判定する基準として、推定値の変化を用いる方法があった。この方法では交絡要因の疑いのある変数で調整した場合とその変数を見捨てた場合で曝露の効果 (リスク比, リスク差など) が大きく変われば交絡あり、ほとんど変化しなければ交絡なしと判定する (例えば調整した場合としない場合とでリスク比が10%以上違うかどうか)。しかし、サンプルから交絡の有無を判断することができない以上、どの基準を用いてもデータから判断することの不適切さは明らかである。

最近の提案は古くからの方法の改良版で、

- 1) 事前に交絡要因と分かっている要因は全て調整する、
- 2) はっきりした証拠はないがひょっとすると交絡要因かもしれないという変数 (例えば、リスク要因であることは分かっているが、曝露グループと非曝露グループで分布が異なるかどうかは分かっていない変数) についてのみ推定値の変化による基準を用いて判定する、

という方法である (Mickey and Greenland (1989))。また、推定値の変化だけではなく、効果の信頼区間の上限と下限が (例えば) 10%以上違ったら交絡要因だと判断する、という基準もうまく働くようである (Greenland (1989b))。非常に強いリスク要因ではあるが曝露グループと非曝露グループで分布がほぼ等しい (つまり、厳密には交絡要因になっていない) 変数で調整した場合、効果の点推定値は変化しないが推定誤差を小さくできる場合がある。信頼区間の変化にもとづく方法はこの点に対応した改良となっている。

推定値の変化による基準では、調整した場合としない場合でどの程度効果の推定値が異なったら交絡ありと判断するかを決めなければならない。一応の目安として10-15%以上と提案さ

れているが (Mickey and Greenland (1989)), 厳密な根拠があるわけではない。しかし、この点は有意性検定を用いる場合でも有意水準を何%にするかを定める必要があり、それと同程度の恣意性にすぎない (Greenland (1989b))。

### 6.3 オッズ比, 疾病発生率の比を効果の指標とする場合

リスク比やリスク差を効果の指標として、層別解析や回帰モデルにより交絡要因の調整を行う場合は、6.2節で述べた手順で調整を行えばよい。オッズ比を効果の指標として用いる場合も、対象としている疾病の発生がまれであればオッズ比はリスク比のよい近似となるし、研究デザインによってはオッズ比でリスク比や疾病発生率の比を推定できるので (佐藤 (1992)), この場合には6.2節と同様に調整を行うことができる。しかし、疾病オッズ比を直接推定し、しかも疾病がまれでない場合、交絡要因だけを調整したのではオッズ比の推定にバイアスが入る場合がある (Hauck et al. (1991))。さらに、オッズ比の場合には「非常に強いリスク要因で調整した結果、点推定値は変化しないが、推定値の分散は小さくなる」という線形モデルにみられるような効率の増加はありえない。ある要因で調整してもしなくてもオッズ比の推定値が変化しない場合には、調整した推定値の分散は調整しない推定値の分散よりも必ず大きくなる (Robinson and Jewell (1991))。

一方、疾病発生率の比 (生存解析で興味のあるパラメータで incidence rate ratio, hazard ratio などと呼ばれる) を効果の指標として用いる場合、研究開始時には曝露グループ、非曝露グループでバランスのとれたリスク要因であり交絡が起きていない場合でも、強いリスク要因を調整しないと疾病発生率の比の推定にバイアスが入る (Chastang et al. (1988))。生存時間解析では「時刻  $t$  において、曝露グループが曝露を受けていなかった場合の時刻  $t$  から観察終了までの生存曲線と、非曝露グループの時刻  $t$  から観察終了までの生存曲線が等しいとき交絡なし」とするのが交絡の定義である (Robins (1989))。この定義は時間に依存しており、したがって研究開始時に交絡がなくても時間の経過とともに交絡が起こる、という状況がありうる。

このためオッズ比や疾病発生率を効果の指標とする場合には、バランスのとれているリスク要因であっても、場合によっては調整しなければならないことがあるので注意が必要である。

## 7. 交絡調整の例

Veterans' Administration Lung Cancer Trial (Kalbfleisch and Prentice (1980)) のデータを使って実際に交絡変数の選択を行ってみよう。これは、化学療法と標準療法のランダム化臨床試験であるから、治療 (1: 化学療法, 0: 標準療法) は必ず解析に含める。また交絡要因の候補は表5に示した5つの予後因子である。前治療は二値データ、組織型はダミー変数で与え、残りの変数は連続量のまま扱った。この研究はランダム化臨床試験であるため事前情報としては交絡はないと考える。ただし、この試験は対象者数が137名とそれほど多くないので強い事前情報ではないのと、生存解析では研究開始時に交絡はなくとも、時間の経過とともに交絡が起

表5. Veterans' Administration Lung Cancer Trial の共変量リスト。

Karnofsky	重症度を示す尺度
罹病期間	肺がん発生からランダム化までの月数
年齢	ランダム化時の年齢
前治療	試験以前の治療の有無
組織型	扁平上皮がん, 小細胞がん, 腺がん, 大細胞がん

表 6. 各変数を1つずつ除いた場合の治療の MRR の比.

治療の MRR	0.75	0.75	0.74	0.77	0.82
除く変数					
罹病期間	1.00				
前治療	1.00	0.99			
年齢	1.04	1.04	1.04		
Karnofsky	1.13	1.12	1.13	1.06	
組織型	1.11	1.11	1.12	1.09	1.20

きる場合があるので、変数減少法により交絡を逐次評価していくことにする。判断の基準としては効果の指標の推定値の変化（15% 以下なら除く）を用いた。この場合の治療効果の指標は化学療法と標準療法の死亡率の比 MRR であり、解析には Cox 回帰を用いた。

結果を表 6 に示す。表中の数字は一つの変数を除いた場合の MRR と除かなかった場合の MRR（最上段の値）の比である。変数減少法により、この比が 0.85-1.15 の間にある変数で最も 1 に近いものから順に除いていった。最初のステップでは罹病期間を除いても 0.75 のままであるので罹病期間を除き、次は前治療を除いて、最終的に残った変数は組織型だけであった。化学療法の MRR の推定値および 95% 信頼区間は、全変数を用いた飽和モデルで 0.75 (0.50, 1.12)、組織型のみを調整した最終モデルで 0.82 (0.56, 1.21) であり、化学療法は標準療法に比べ死亡リスクを 20% 減少させたが、5% 水準で有意ではなかった（カッコ内 95% 信頼区間）。

## 8. ま と め

本論文では、疫学研究で重要な概念である交絡と効果の修飾について、因果推論にもとづく交絡の定義、効果の修飾がある場合の交絡の調整方法としての標準化、交絡変数の選択方法、の解説を行った。

全ての交絡要因のレベルでリスク比またはリスク差が共通でない場合でも、質的な交互作用がなく効果の修飾だけがみられる場合には、曝露の効果を要約した単一の指標がほしいという要求は自然である。全ての層で曝露効果が一定であるという仮定は非常に厳しい仮定であり、この共通性の仮定を支持する生物学的なモデルがない場合には「曝露効果は共通ではなく、交絡要因のレベルによって異なる」と考えるべきである (Greenland (1987))。効果の修飾が存在する場合でも標準化した指標を求めることで疫学的に解釈可能な要約を行うことができる。標準化は便利な手法ではあるが当然限界もある。標準化一般についての注意としては次の 3 点が指摘されている (Greenland (1991a)):

1. 標準集団としては、推測の対象としている標的集団の交絡要因の分布を反映した集団を選ばなければならない。
2. 標準化した指標だけに注目することにより、重要な効果の不均一を見逃す可能性がある。そのため、各層の効果の指標もチェックする必要がある。
3. 「全ての層にわたって曝露の効果はない」という仮説の検定に、標準化した指標とその標準誤差を用いた Wald 検定、

$$\frac{\log SRR}{\sqrt{V(\log SRR)}}, \quad \frac{SRD}{\sqrt{V(SRD)}}$$

を用いることは効率が悪いので避けるべきである。

表7. 組織型によるサブグループ解析.

組織型	死亡率の比	(95% 信頼区間)
扁平上皮がん	1.83	(0.87, 3.86)
小細胞がん	0.61	(0.32, 1.15)
腺がん	0.80	(0.34, 1.90)
大細胞がん	0.64	(0.28, 1.44)
全体	0.85	(0.59, 1.24)
扁平上皮除く	0.67	(0.43, 1.03)

4章ではロジスティックモデルを用いた標準化を紹介したが、Greenland and Holland (1991)では Mantel-Haenszel オッズ比を用いてスムージングした標準化リスク差の推定を議論している。

因果パラメータに関する推測は当然のことだが難しい。6章で述べた交絡要因調整の方法は決定的というわけではなく、推定値の変化による基準も、何パーセント以下なら変数を省くか特に決まりはない。また、どんなに注意深く変数選択を行っても、もともと用いたモデルが誤っていたのでは何にもならない。それを防ぐには、考え得るいくつかの異なったモデルにより曝露の効果を推定し、どのモデルを用いても少なくとも効果の方向は一致することを確認する、いわゆるモデルの sensitivity analysis を行う必要がある。また、強い仮定を必要とするモデルをデータ解析に用いる場合には、より弱い仮定のモデル（例えば層別解析）で妥当性をチェックしながら解析を進める必要がある (Rothman (1986))。

7章の Veterans' Administration Lung Cancer Trial のデータも Cox 回帰という非常に強い仮定を必要とするモデルで解析を行ったので、念のため最後に残った組織型で層別して解析してみよう。Mantel-Haenszel の方法で生存時間解析を行うと (Sato (1992)), 組織型で調整した化学療法と標準療法の死亡率の比の Mantel-Haenszel 推定値は 0.85, 95% 信頼区間は (0.59, 1.24) と Cox 回帰の結果とほとんど変わらなかった。同じことを組織型別に調べた結果を表7に示す。驚いたことに、化学療法の効果は組織型別に一定ではなく、小細胞がん、腺がん、大細胞がんでは死亡率を減らす効果があるが、扁平上皮がんでは逆に死亡率を高くしている (標準療法が効いている)。扁平上皮がんを除いて、小細胞がん、腺がん、大細胞がんだけで層別解析を行うと、死亡率の比と 95% 信頼区間は 0.67 (0.43, 1.03) であり、5% 水準では有意ではないが (logrank 検定の p 値が 0.09), 化学療法の効果はありそうである。(もちろん、この研究は事前に十分なデザインがなされたランダム化臨床試験であるから、たとえ扁平上皮がんを除いた結果が統計的に有意であっても、研究終了後の探索的な解析から「扁平上皮がん以外では化学療法の効果あり」と結論することはできない。必要であれば、「扁平上皮がん以外では化学療法の効果がある」という仮説を検証できる試験デザインを再度検討すべきである。)

このように質的な交互作用が存在する場合にはもはや結果を単一の効果の指標に要約することはできない。したがって、標準化の際の注意2にあるように重大な交互作用を見落とさないために「各層ごとの効果の指標のチェック」をおこたってはならない。また、疫学研究では曝露-疾病発生前の真実のモデルが分かっている例はほとんどないので (Greenland (1989b)), 現在のデータ解析では Cox 回帰, ロジスティック回帰がルーチンとなっているが、これら以外のモデル当てはめの検討, 層別解析の併用による結果の確認, なども必ず実施しなければならない。

## 謝 辞

草稿の段階で適切なコメントをいただいた橋本修二先生、山本精一郎先生に感謝します。

なおこの研究は、統計数理研究所 共同研究 (3-共研-88, 5-共研 A-73), 文部省科学研究費補助金 05770255, の補助を受けた。

## 付録 1. 標準化リスク比, 標準化リスク差の漸近分散

標準化リスク比と標準化リスク差の漸近分散を導くために, 各層の疾病発生数 ( $x_k, y_k$ ) は独立に二項分布  $\text{Bi}(n_k, p_{1k}), \text{Bi}(m_k, p_{0k})$  に従うと仮定する。また, 以下では全ての ( $n_k, m_k$ ) が十分大きいという大標本モデル (large-strata) を仮定する。Greenland (1982) は標準化リスク比が各層のリスク比,

$$\hat{RR}_k = \frac{m_k x_k}{n_k y_k},$$

の重み付き平均,

$$SRR_R = \frac{\sum_k r_k (x_k/n_k)}{\sum_k r_k (y_k/m_k)} = \frac{\sum_k E_{R[E]k}}{\sum_k E_{R[U]k}} = \frac{\sum_k E_{R[U]k} \hat{RR}_k}{\sum_k E_{R[U]k}},$$

で書けることを利用してその漸近分散,

$$\text{var}^A(SRR_R) = \frac{\sum_k [E(E_{R[U]k})]^2 \text{var}^A(\hat{RR}_k)}{[E(\sum_k E_{R[U]k})]^2},$$

を導いた。  $\text{var}^A$  は漸近分散を意味する。次の関係,

$$E(E_{R[U]k}) = r_k p_{0k}, \quad \text{var}^A(\hat{RR}_k) = RR_k^2 \text{var}^A(\log \hat{RR}_k),$$

に注意して,  $\log SRR_R$  の漸近分散は,

$$\begin{aligned} \text{var}^A(\log SRR_R) &= \text{var}^A(SRR_R) / SRR_R^2 \\ &= \frac{\sum_k (r_k p_{1k})^2 \text{var}^A(\log \hat{RR}_k)}{(\sum_k r_k p_{1k})^2}, \end{aligned}$$

であり ( $SRR_R$  は標準集団を R とした場合の真の標準化リスク比),  $r_k p_{1k}$  を  $E_{R[E]k}$  で推定して, 本文の  $V_G(\log SRR_R)$  を得る。

もう一つの漸近分散の推定量は, 直接,

$$\begin{aligned} \text{var}^A(SRR_R) &= \frac{\text{var}(E_{R[E]} - SRR_R E_{R[U]})}{[E(E_{R[U]})]^2} \\ &= \frac{\sum_k \text{var}(E_{R[E]k}) + SRR_R^2 \sum_k \text{var}(E_{R[U]k})}{[E(\sum_k E_{R[U]k})]^2} \\ &= \frac{\sum_k r_k^2 p_{1k}(1-p_{1k})/n_k + SRR_R^2 \sum_k r_k^2 p_{0k}(1-p_{0k})/m_k}{(\sum_k r_k p_{0k})^2}, \end{aligned}$$

を求めて,

$$\text{var}^A(\log SRR_R) = \frac{\sum_k r_k^2 p_{1k}(1-p_{1k})/n_k}{(\sum_k r_k p_{1k})^2} + \frac{\sum_k r_k^2 p_{0k}(1-p_{0k})/m_k}{(\sum_k r_k p_{0k})^2},$$

であるから,  $r_k p_{1k}, r_k p_{0k}$  を  $E_{R[E]k}, E_{R[U]k}$  で推定して得たものである。

同様にして標準化リスク差の漸近分散の推定量は,

$$\begin{aligned} \text{var}^A(SRD_R) &= \text{var}(E_{R[E]} - E_{R[U]})/r^2 \\ &= \frac{\sum_k r_k^2 p_{1k}(1-p_{1k})/n_k + \sum_k r_k^2 p_{0k}(1-p_{0k})/m_k}{r^2}, \end{aligned}$$

から得られる。

## 付録2. ロジスティックモデルにもとづく標準化リスク比, 標準化リスク差の漸近分散

ロジスティックモデルでスムージングした標準化リスク比  $\log SRR_R^*$  の漸近分散は, 多変量デルタ法 (Bishop et al. (1975)) を用いて,

$$\text{var}^A(\log SRR_R^*) = \left( \frac{\partial}{\partial \beta} \log SRR_R \right)' V_\beta \left( \frac{\partial}{\partial \beta} \log SRR_R \right)$$

で表される。この推定値は,

$$V(\log SRR_R^*) = R_R' V_b R_R,$$

ただし,

$$\begin{aligned} R_R &= \frac{1}{\sum_R P_{1i}} \left[ \sum_R \frac{e^{b'x_{1i}}}{(1+e^{b'x_{1i}})^2} x_{1i} \right] - \frac{1}{\sum_R P_{0i}} \left[ \sum_R \frac{e^{b'x_{0i}}}{(1+e^{b'x_{0i}})^2} x_{0i} \right], \\ x_{1i} &= (1, 1, z_i)', \quad x_{0i} = (1, 0, z_i)'. \end{aligned}$$

同様に, ロジスティックモデルでスムージングした標準化リスク差  $SRD_R^*$  の漸近分散の推定値は,

$$V(SRD_R^*) = D_R' V_b D_R,$$

ただし,

$$D_R = \frac{1}{N_R} \sum_R \left[ \frac{e^{b'x_{1i}}}{(1+e^{b'x_{1i}})^2} x_{1i} - \frac{e^{b'x_{0i}}}{(1+e^{b'x_{0i}})^2} x_{0i} \right],$$

また,  $N_R$  は標準集団  $R$  に対応して  $N_R \in \{n, m, N\}$  である。

標準集団を曝露グループにとった  $R=E$  の結果を本文に示した。

## 参 考 文 献

- Begg, C.B. (1990). Significance tests of covariate imbalance in clinical trials, *Controlled Clinical Trials*, **11**, 223-225.
- Bishop, Y.M.M., Fienberg, S.E. and Holland, P.W. (1975). *Discrete Multivariate Analysis*, The MIT Press, Boston.
- Boivin, J.F. and Wacholder, S. (1985). Conditions for confounding of the risk ratio and of the odds ratio, *American Journal of Epidemiology*, **121**, 152-158.
- Breslow, N.E. and Day, N.E. (1980). *Statistical Methods in Cancer Research. Vol. 1—The Analysis of Case-Control Studies*, Oxford University Press, New York.
- Chastang, C., Byar, D. and Piantadosi, S. (1988). A quantitative study of the bias in estimating the treatment effect caused by omitting a balanced covariate in survival models, *Statistics in Medicine*, **7**, 1243-1255.
- Checkoway, H., Pearce, N.E. and Crawford-Brown, D.J. (1989). *Research Methods in Occupational Epidemiology*, Oxford University Press, New York.
- Fleiss, J.L. (1982). *Statistical Methods for Rates and Proportions*, 2nd ed., Wiley, New York.
- Fleiss, J.L. (1986). Significance tests have a role in epidemiologic research: reactions to A.M.



- Walker, *American Journal of Public Health*, **76**, 559-560.
- 福富和夫, 橋本修二 (1989). 標準化死亡比に関する考察, *日本公衆衛生雑誌*, **36**, 155-160.
- Grayson, D.A. (1987). Confounding confounding, *American Journal of Epidemiology*, **126**, 546-553.
- Greenland, S. (1982). Interpretation and estimation of summary ratios under heterogeneity, *Statistics in Medicine*, **1**, 217-227.
- Greenland, S. (1987). Interpretation and choice of effect measures in epidemiologic analysis, *American Journal of Epidemiology*, **125**, 761-768.
- Greenland, S. (1989a). Reader Reaction: Confounding in epidemiologic studies, *Biometrics*, **45**, 1309-1310.
- Greenland, S. (1989b). Modeling and variable selection in epidemiologic analysis, *American Journal of Public Health*, **79**, 340-349.
- Greenland, S. (1990). Randomization, statistics, and causal inference, *Epidemiology*, **1**, 421-429.
- Greenland, S. (1991a). Estimating standardized parameters from generalized linear models, *Statistics in Medicine*, **10**, 1069-1074.
- Greenland, S. (1991b). *Validity Concepts in Epidemiologic Research*, Unpublished course text, UCLA School of Public Health.
- Greenland, S. and Holland, P.W. (1991). Estimating standardized risk differences from odds ratios, *Biometrics*, **47**, 319-322.
- Greenland, S. and Robins, J.M. (1986). Identifiability, exchangeability, and epidemiologic confounding, *International Journal of Epidemiology*, **15**, 413-419.
- Hauck, W.W., Newhaus, J.M., Kalbfleisch, J.D. and Anderson, S. (1991). A consequence of omitted covariates when estimating odds ratios, *Journal of Clinical Epidemiology*, **44**, 77-81.
- Holland, P.W. (1989). Reader Reaction: Confounding in epidemiologic studies, *Biometrics*, **45**, 1310-1316.
- Kalbfleisch, J.D. and Prentice, R.L. (1980). *The Statistical Analysis of Failure Time Data*, Wiley, New York.
- Keiding, N. (1987). The method of expected number of deaths, 1786-1886-1986, *Internat. Statist. Rev.*, **55**, 1-20.
- Kleinbaum, D.G., Kupper, L.L. and Morgenstern, H. (1982). *Epidemiologic Research*, Van Nostrand Reinhold, New York.
- Mickey, R.M. and Greenland, S. (1989). The impact of confounder selection criteria on effect estimation, *American Journal of Epidemiology*, **129**, 125-137.
- Miettinen, O.S. (1972). Components of the crude risk ratio, *American Journal of Epidemiology*, **96**, 168-172.
- Miettinen, O.S. (1974). Confounding and effect-modification, *American Journal of Epidemiology*, **100**, 350-353.
- Miettinen, O.S. and Cook, E.F. (1981). Confounding: essence and detection, *American Journal of Epidemiology*, **114**, 593-603.
- 新田裕史, 佐藤俊哉 (1989). 『生物統計学入門』, 現代数学社, 京都.
- Robins, J.M. (1989). The control of confounding by intermediate variables, *Statistics in Medicine*, **8**, 679-701.
- Robins, J.M. and Greenland, S. (1986). The role of model selection in causal inference from nonexperimental data, *American Journal of Epidemiology*, **123**, 392-402.
- Robinson, L.D. and Jewell, N.P. (1991). Some surprising results about covariate adjustment in logistic regression models, *Internat. Statist. Rev.*, **59**, 227-240.
- Rothman, K.J. (1986). *Modern Epidemiology*, Little, Brown, and Co., Boston.
- Rothman, K.J. (ed.) (1988). *Causal Inference*, Epidemiology Resources Inc., Boston.
- Rubin, D.B. (1974). Estimating causal effects of treatments in randomized and non randomized studies, *Journal of Educational Psychology*, **66**, 688-701.
- Sato, T. (1992). Comment to "A comparison of two simple hazard ratio estimators based on the logrank test", *Statistics in Medicine*, **11**, 487-488.
- 佐藤俊哉 (1992). 疫学研究のデザインと相対リスクの推定, *統計数理*, **40**, 173-184.
- 佐藤俊哉 (1993). 疫学研究における生物統計手法, *日本統計学会誌*, **22** (増刊号), 493-513.
- 柴田義貞 (1985). 標準化死亡比による死亡率比較の妥当性, 文部省科学研究費総合研究 A 昭和 59 年度「生

存時間の統計解析と医学薬学への応用」研究班研究会予稿集, 4-8.

Simpson, E.H. (1951). The interpretation of interaction in contingency tables, *J. Roy. Statist. Soc. Ser. B*, **13**, 238-241.

Wickramaratne, P.J. and Holford, T.R. (1987). Confounding in epidemiologic studies: the adequacy of the control group as a measure of confounding, *Biometrics*, **43**, 751-765.

## Confounding and Effect Modification in Epidemiologic Studies

Tosiya Sato

(The Institute of Statistical Mathematics)

Confounding is a major source of bias in estimation of causal exposure effect in epidemiologic studies. Every epidemiologist knows the impact of confounding. However, until recently, it is not clear what confounding is, and how it occurs. Based on recent developments in causal inference, this paper reviews essential aspects of confounding. It is defined by “the adequacy of the control group” or “comparability”. This paper shows the important role of standardization for adjustment of confounding. Standardization methods are introduced directly from causal inference. It can be used under effect modification. Effect modification is another important concept in epidemiology, which is closely related to but basically independent of confounding. Some technical details for standardization are discussed.

This paper also gives a practical guide to confounder selection problems. It emphasizes that selection will be based on the update version of “change in estimate” criterion rather than significance tests. Through a real example, it strongly recommends to use robust stratified analysis in addition to standard statistical models, such as logistic and Cox regressions.