

応じて定義できる必要がある。

表 2-4-4

	①狭い症 例の定義 に合致	②広い症 例の定義 に合致	③症例の 定義には 合致しな いが	④無症状	合計
曝露あり	a 人	b 人	c 人	d 人	a+b+c+d 人
曝露なし	e 人	f 人	g 人	h 人	e+f+g+h 人
合計	a+e 人	b+f 人	c+g 人	d+h 人	

$$\{(a+b+c+d) \times (e+f+g+h)\} / \{(e+f+g+h) \times (a+b+c+d)\}$$

なお、無症状の定義を行わず、単に症例でないものを無症状にすると影響の程度を過小評価することになる。

多水準の曝露の分析

上記の場合は、症例の定義を複数用意することにより、2かけ2表がnかけ2表になったが、今度は、曝露の水準を2より多く設定することにより、2かけ2表が2かけn表になる場合である。これは簡単な例で示すと、「喫食歴あり」と「喫食歴なし」の2段階に分けていたものが、「たくさん食べた」、「少しだけ食べた（2－3口以内）」、「全く食べなかった」の3段階に分けるような場合である。これを表 2-4-5 に示す。

表 2-4-5

	疾病あり	疾病なし	合計
たくさん食べた	a 人	b 人	a+b 人
2－3口食べた	c 人	d 人	c+d 人
全く食べなかった	e 人	f 人	e+f 人
合計	a+c+e 人	b+d+f 人	

この場合、あくまでも「喫食歴なし」あるいは「全く食べなかった」を基準にして、オッズ比を計算する。そうすると2かけ2表では、1個しかオッズ比は計算できなかったが、表 2-4-5 のような3かけ2表ではオッズ比は、2個計算できるようになる。「喫食歴なし」に比較して「たくさん食べた」場合のオッズ比と、「喫食歴なし」に比較して「少しだけ食べた」場合のオッズ比である。

たくさん食べた場合のオッズ比： $(a \times f) / (b \times e)$

2-3口食べた場合のオッズ比： $(c \times f) / (d \times e)$

まとめ

- ①アウトブレイク調査は、観察研究が基本となり、必然的に「後ろ向き調査」が基本となる。
- ②従って、「後ろ向きコホートデザイン」（全数調査）と「症例対照研究デザイン」（分母のサンプリング）が基本デザインとなる。
- ③いずれのデザインも、2かけ2表とそこから算出される疫学的影響の指標が基本となる。オッズ比の意味を理解する必要がある。
- ④症例対照研究デザインでは、対照の選び方の基本を学ぶ必要がある。病院対照、近所・友人対照、人口ベース・地域ベース対照の三種類の対照の選び方の長所短所を踏まえる。マッチングサンプリングは対照の選び方の効率を上げる。次節に解説するマッチング分析が必要である。
- ⑤仮説の設定は、どのような対策を取るかに大きく影響される。現時点で、原因施設を究明するのか、原因食品を究明する必要があるのかを明確にして、それを仮説に組み込む。対策を念頭において仮説を設定し調査を進めることが無駄を除くことになる。
- ⑥症例の定義と同時に「無症状」の定義を行う。多水準の曝露分析についても試み、量反応関係を描出すると関連が明瞭になることがある。

第5節 バイアスとチャンス

—チャンスと信頼区間、 χ^2 検定

学習目標

- ①生物統計学の推論の基本である検定と推定を学ぶ。
- ②「有意差がある」、「有意差がない」ということの意味を踏まえ、このことが「因果関係がある」、「因果関係がない」ということと混同しないように基本を学ぶ。
- ③疫学的指標の測定誤差の構成成分であるチャンス（偶然の変動）とバイアス（系統的誤差）の基本を学ぶ。
- ④疫学分析全般に関連するバイアスとして、選択バイアス、情報バイアス（誤分類）、交絡バイアスについてアウトブレイク調査との関連で学ぶ。
- ⑤症例の定義を変えることを、情報バイアスとの関連で学ぶ

検定と推定、有意差について

推定されたオッズ比やリスク比などの疫学的影響の指標、つまり何倍多発、が偶然の変

動によりたまたま多めに観察されたものなのか、それとも偶然による変動を考慮しても多発していると判断すべき多発なのか、どちらかであるかを質的に判断する方法は、有意差検定と呼ばれる。2かけ2表における有意差検定としては、カイ二乗検定 (χ^2 乗検定) が最もよく知られている。

検定の基本的考え方は、統計学の教科書を参照にされたいが、ここでもその骨子を簡潔に述べる。検定は、比較する二群の指標（ここでは発症割合）に差がないという帰無仮説 (null hypothesis) をまず設定する。次に、実際に集められたデータ（ここでは2かけ2表から得られたデータ）から判断して、この帰無仮説が成り立つ確率がかなり低い時（しばしば 5%未満という閾値が採用される）、帰無仮説は支持出来ないのをこれを棄却する。この時、統計学的に有意な差があったと称する。また、曝露群の発症割合が高かった場合には、「曝露群に有意な多発があった」というふうに称する。

帰無仮説が成り立つ確率をデータから計算する際に、 χ^2 検定統計量を計算し、 χ^2 分布を用いて判断し、それに相当する確率を用いる場合をカイ二乗検定と称する。 χ^2 分布を近似的に当てはめているので、2かけ2表の中に5未満の少ない数がある場合には、Yatesの補正などの補正を行う。さらに少数がある場合は、Fisherの直接確率法を用いて χ^2 検定統計量を計算せずに直接確率値を計算することが可能である。

検定には、多発と減少の両方を考慮する両側検定とどちらか片方のみを考慮する片側検定がある。カイ二乗検定では両側検定を、Fisherの直接確率法では片側検定を用いる。

次に推定の考え方について、簡潔に述べる。推定は、点推定値と区間推定値を計算するが、このうち区間推定値を用いることにより有意差検定と同様の有意差の判断を行うことができる。ただし、有意差検定では帰無仮説を基本にして判断を行うのに対して、区間推定値は対立仮説を基本にして判断を行うので、両者による有意差の判断は、（しばしば微妙ではあるが）異なる。

点推定値と区間推定値の関係を図 2-5-1 に示す。以下、本稿では、推定する対象を2かけ2表から計算されるオッズ比もしくはリスク比に話題を絞る。つまりこの場合、何倍多発するかという値を、一点の値（点推定値）と区間の値（二点：区間推定値）を用いて我々は推定することになる。特にここでは、オッズ比に絞って説明する。リスク比も同様に考えることができる。なお、特に指定しないかぎり、区間推定は、95%区間推定値を推定していることとして説明する。

図 2-5-1 点推定と区間推定

点推定値は、主に交差積 (ad/bc) を用いて計算されるオッズ比そのものである。これで真のオッズ比を一点で推定していることになる。しかしこれでは偶然による変動を評価出来ないで、95%の確率で真のオッズ比を含むようなオッズ比の区間を推定する。区間であ

るので二点のオッズ比で与えられる。正規分布を仮定したとき、点推定値と区間推定値の関係は次のようになる。

区間推定値の下限の値 = 点推定値 - 1.96 × 標準誤差

区間推定値の上限の値 = 点推定値 + 1.96 × 標準誤差

オッズ比で正規分布を仮定するには、通常、自然対数をとったオッズ比を用いるので、オッズ比の区間推定値の下限値と上限値の関係は次のようになる。多発している場合、オッズ比の区間推定値の下限が 1 を上回る場合（区間推定オッズ比の下限の自然対数が 0 を上回る場合）、有意な多発があったと判断する。

区間推定オッズ比の下限の自然対数

= オッズ比の自然対数 - 1.96 × オッズ比の自然対数の分散の平方根

区間推定オッズ比の上限の自然対数

= オッズ比の自然対数 + 1.96 × オッズ比の自然対数の分散の平方根

ここまで、95%信頼区間に関してのみ述べてきた。なぜ 95%信頼区間を用いるのかという理由を問う質問がしばしばあるが、この理由は特になく、慣習上最も用いられることが多いという理由しかない。これは有意差検定で 5%の閾値が慣習上用いられるのと同様である。一説には、研究者の生涯に 20 回程度の研究が行われ、研究一回につき一回の検定を行うと、その中に偶然による変動による有意差が一回程度入ってくる確率が 5%あるいは 95%という理由などがある。

95%信頼区間の他に、90%信頼区間や 99%信頼区間が用いられることも実際にある。特に、90%信頼区間が用いられていることを見ることは珍しくない。90%信頼区間、95%信頼区間、99%信頼区間の相互の関係を図 2-5-2 に示す。真値が含まれる確率を示す区間なので、区間の幅は、同じデータから計算する場合、99%信頼区間が一番広く、90%信頼区間が一番狭い。それぞれを一つの図（図 2-5-2）に模式的に表すと、90%信頼区間の周りを 95%信頼区間が囲み、その周りを 99%信頼区間が囲んでいることが分かる。90%信頼区間を真値が含まれる確率が高いほど広く、また真値が含まれる確率を上げようとするほど有意差の判断ができにくくなることが分かる。ちなみに、点推定値は幅を持たないので 0%信頼区間であり、オッズ比の場合 0 から無限大の値を取るため、100%信頼区間は 0 から無限大になる。ここで抑えておくべきは、オッズ比の真値が分布する可能性は、オッズ比 0 から無限大まで存在しているということである。そしてオッズ比のそれぞれの値毎の真値がある可能性は、点推定値において最大であり、点推定値の両側にむかって低下する。90%信頼区間や 95%信頼区間、99%信頼区間は、その可能性の合計がそれぞれ、90%、95%、99%になるところで区切っているだけである。従って、このように見えてくると、有意差がある

のかないのかは、どこで区切るか（どのような信頼区間を選択するのか）により異なってくる。このことを踏まえておかないと、「有意差がある」か「有意差がない」かに過度にこだわるような態度に結びついてしまうことになる。

図 2-5-2 90%信頼区間、95%信頼区間、99%信頼区間

次に検定と推定のどちらを用いるべきかについて述べる。2 かけ 2 表から検定を行って得られる情報は、帰無仮説を棄却出来るかできないか、すなわち「有意差がある」か「有意差がない」かの質的判断の情報である。ところが、推定により得られる情報は、「有意差がある」か「有意差がない」かの質的情報の他に、どの程度、非曝露者に比較して曝露者においてどの程度疾患が多発しているかという疫学的影響の指標（オッズ比など）も情報として提供されることになる。このことは具体的にどのようなことを意味するかについて図 2-5-3 を用いて説明する。

図 2-5-3 は 2 つの研究結果をオッズ比とその信頼区間で示したものである。横の矢印の両端が 95%信頼区間を示し、両矢印の真ん中にある点が、点推定値を示している。図の左端の縦線は、オッズ比が 1 の点（null value）を示している。研究 A は有意差があるが、研究結果が示すオッズ比の点推定値は 1.5 程度である。研究 B の信頼区間の下限は 1 をわずかに下回るの有意差はないが、研究の結果が示すオッズ比の点推定値は 4 を超えている。ここで有意差のみを判断の基準とすると、研究 A に有意差があつて、研究 B は有意差がないという情報しか伝わらない。しかし、オッズ比の点推定値と区間推定値で示せば、研究 B には有意差はないが、オッズ比の点推定値が研究 A よりずっと高い値を示しており、実際に大きな多発が起こっている可能性が高いことを示している。有意差があるとかないとかの情報だけでは、多発の程度の情報は伝わらない。一方、点推定値と区間推定値の両方を示せば、多発の程度を示すオッズ比の情報だけでなく、検定に相当する有意差の判断もできることになる。検定結果だけでなく、点推定値と区間推定値で示せと多くの医学雑誌の投稿規定が求める理由はここにある。

図 2-5-3 点推定値は大きくないが有意差がある場合と大きい有意差がない場合

図 2-5-4 は、1978 年にニューイングランド医学雑誌に掲載された論文で示された図である。この図では有意差がないとされた 71 の治療法の効果を検証した論文の結果が区間推定値で示されている。そうするとその中の多くは、治療法の効果を示す傾向があることが分かる。有意差検定だけではこのような情報が捨てられてしまうのである。

図 2-5-4 有意差がなかった 71 臨床試験の結果を区間推定値で示すと

食中毒事件の疫学や感染症の疫学調査では、危険をいち早く察知し、証拠に基づいて対策を講じなければならない。また、有意差が出るまで例数を増やすということは、実際上は不可能である。従って、このような意味からも、有意差の有無だけに頼ることは、実務的でないばかりか危険ですらあることは十分承知しておかなければならない。

2 かけ 2 表における様々な有意差検定と様々な区間推定

2 かけ 2 表における有意差検定では、通常のカイ 2 乗検定の他、Yates の補正、Mantel-Haenszel 検定、Fisher の直接確率法検定、mid-P、などがあるが、基本的に 2 かけ 2 表の各セルの値が大きいつきには（5 より大きいつき）、通常のカイ 2 乗検定で済む。5 以下のセルがあるつきには、様々な補正や直接確率法検定などが必要となってくる。しかしパーソナルコンピューターとそこで使用する統計パッケージが発達した現在では、計算の前にいずれかの方法を選択する必要はなく、計算されて出てきた全ての方法のうち、いずれかを選択すればよいことになる。

区間推定も基本的に同様である。巻末に、様々な有意差検定や区間推定の種類と考え方を記したので参照されたい。

90%信頼区間、95%信頼区間、99%信頼区間とある様々な推定幅のいずれを選択するかについては、あらかじめ決めておく必要があるが、推定という考え方からは有意差の有無はさほど重要ではないことは既に述べたとおりである。

チャンスとバイアス

2 かけ 2 表を作成してオッズ比などの疫学的影響の指標を推定する分析疫学では、オッズ比などの疫学的影響の指標を、誤差を少なくし、できるだけ正確に推定することが一つの目標となる。疫学的影響の指標の誤差を分類すると、2 種類の誤差に分けられる。チャンス（偶然の変動による誤差）とバイアス（系統的誤差）である。さらにバイアスは、情報バイアス (information bias)、交絡バイアス (confounding bias)、選択バイアス (selection bias) の 3 つに分類される。これを図で示すと図 2-5-5 のようになる。情報バイアスはさらに 4 種類に分類出来る。検診の評価の疫学では、さらに lead time bias、length biased sampling の 2 つがあり、またメタアナリシスでは出版バイアス (publication bias) も指摘されているが、食中毒事件・感染症の疫学ではあまり関係がないので、ここでは、チャンス、情報バイアス、交絡バイアス、選択バイアスに絞って概略を説明する。

表 2-5-1 誤差、チャンス、バイアス

誤差 (error)

チャンス (偶然の変動による誤差)

バイアス (系統的誤差)

情報バイアス (information bias) もしくは誤分類によるバイアス

曝露の誤分類 (exposure misclassification)

Non-differential

Differential

疾病の誤分類 (disease misclassification)

Non-differential

Differential

交絡バイアス (confounding bias)

選択バイアス (selection bias)

チャンス (偶然の変動による誤差)

標本数 (調査対象者数) が少なければ少ないほど、偶然の変動による誤差が入りやすい。逆に標本数が多ければ多いほど、真の値に近づく可能性が大きくなる。前者は、信頼区間が広いことを意味し、偶然による変動が含まれている可能性が大きいことを意味する。後者は、逆に信頼区間が狭いことを意味し、偶然の変動による誤差が含まれている可能性が小さいことを意味する。

既に説明したように、検定は影響の程度の大きさという要素と標本数の大きさ (偶然の変動による誤差) という要素の両方の要素を含んでしまう。

バイアスの方向性

バイアスの分類以外に、そのバイアスがオッズ比などの疫学的影響の指標を、どのような方向に系統的誤差を生じさせるかを予測することは、バイアスの考察の上で不可欠である。バイアスがもたらす誤差を3つに分類すると以下ようになる。

1. 疫学的影響の指標 (この場合オッズ比など相対危険度) に無限大の方向へ誤差をもたらすバイアス
2. 疫学的影響の指標 (この場合オッズ比など相対危険度) に0の方向へ誤差をもたらすバイアス
3. 疫学的影響の指標 (この場合オッズ比など相対危険度) に1の方向へ誤差をもたらすバイアス

なお、上記の分類では、疫学的影響の指標としてオッズ比など相対危険度を想定したが、これがリスク差など寄与危険度の場合には、1. の無限大の方向はそのままだが、2. の0の方向はマイナス無限大の方向、3. の1の方向は、0の方向である。このバイアスの方向を、バイアスそれぞれを検討する際に念頭におけば、問題の整理をしやすくなる。以下、食中毒の疫学で問題となりそうな点を中心に、主なバイアスについて説明する。

情報バイアス

食中毒事件の疫学調査における「情報」とは、喫食調査と症状調査から得られる情報である。喫食調査の情報は、疫学一般で言うと曝露情報に相当する。この喫食調査で得られた情報を元にして得た結果が、真の喫食結果と異なる分類がされたときに、曝露の誤分類 (exposure misclassification) という。この誤分類を感度・特異度の説明に用いた2かけ2表(表2-1-2:曝露と疾病の2かけ2表とは異なるので注意!)を用いて表現すると表2-5-2のようになる。一方、症状調査で得られた情報を元にして得た結果が、症状の真の有無と異なる分類がなされたときに、疾病の誤分類 (disease misclassification) という。この誤分類を感度・特異度の説明に用いた2かけ2表(表2-1-2:曝露と疾病の2かけ2表とは異なるので注意!)を用いて表現すると表2-5-3のようになる。

表 2-1-5. 食中毒事件における曝露の誤分類を説明する表

	真に喫食あり	真に喫食なし	計
喫食したと回答	a人	b人	a+b人
喫食せずと回答	c人	d人	c+d人
計	a+c人	b+d人	a+b+c+d人

表 2-1-5. 食中毒事件における曝露の誤分類を説明する表

	真に症状あり	真に症状なし	計
症状なしと回答	a人	b人	a+b人
症状ありと回答	c人	d人	c+d人
計	a+c人	b+d人	a+b+c+d人

それぞれ表で、b人とc人の存在は、喫食調査と症状調査の結果が、それぞれ真の喫食状況と真の症状の有無とは多少のズレ(誤分類)が存在していることを表現している。喫食していたのに喫食していない、喫食していなかったのに喫食していた、症状があったのに症状がない、症状がないのに症状があった、とそれぞれ間違っ情報で分類されているから誤分類と表現する。

問題は、それぞれの誤分類が起こったときに、オッズ比などの疫学の影響の指標がどの

方向にバイアスされるかである。これを整理するために誤分類を2種類に分類して整理する。Non-differentialとDifferentialである。すでに誤分類は、曝露の誤分類と疾病の誤分類の2種類あるので、この分類を加えることにより、誤分類は表2-5-1のように4種類になる。

次に、Non-differentialとDifferentialの誤分類について説明する。曝露情報（喫食情報）と疾病情報（症状情報）は、共に、食中毒事件に限らず疫学調査では必ず集められる情報である。そして曝露情報と疾病情報共に、情報を収集する際に誤分類は多かれ少なかれ生じうる。問題は、その誤分類がどのようにして起こってくるかによって、Non-differentialとDifferentialに分類できる。

通常考えられる曝露情報が正確でないために起こってくる誤分類、あるいは疾病情報が正確でないために起こってくる誤分類は、Non-differentialである。この時、曝露情報の誤分類は疾病の状況（状態）とは関係なしに（独立して）起こり、疾病情報の誤分類は曝露の状況（状態）とは関係なしに（独立して）起こる。食中毒事件での喫食調査と症状調査が、特定の症状と結びつけて喫食調査の特定の食歴の誤分類が生じたり、特定の食歴と結びつけて特定の症状の誤分類が生じたりするようなことは、実際の食中毒事件では以下に述べるような原因施設の特別な例を除いてほとんど起こりえない。この場合を外れる場合をDifferentialな誤分類と呼ぶ。すなわち、曝露情報の誤分類が疾病の状況によって生じるDifferentialな曝露の誤分類と、疾病情報の誤分類が曝露の状況によって生じるDifferentialな疾病の誤分類である。

バイアスの方向性は、Non-differentialはオッズ比など影響の程度指標を1の方向にバイアスする。これは例えば白の水彩絵の具と黒の水彩絵の具を混ぜ合わせれば、間の灰色になるというのと同じことだ。コントラストがない方向（オッズ比1の方向）に行くということである。一方、Differentialな誤分類は無限大とゼロの方向にバイアスする。コントラストが逆にはっきりする方向（オッズ比無限大、もしくはゼロの方向）にバイアスする。オッズ比無限大の方向にバイアスするかゼロの方向にバイアスするかは、状況による。以下、Differentialな誤分類を食中毒事件で想定した説明のための仮想例で示す。

症状のある人と症状のない人のそれぞれから、「どこで喫食したのか」という原因施設を調査したとしよう。この時、ある施設が噂に上っているとき、症状のある人たちはその原因施設での喫食を懸命に思い出そうとするが、症状のない人たちは思い出そうとする努力が相対的に少なくなる。そうすると結果的に「思い出さない」ことが「喫食しなかった」ことにつながり、症状のある人に比べて症状のない人に、相対的に「喫食しなかった」と答える人が多くなり実態を表さなくなる。これは結果的に2かけ2表のdが多くなることを意味してオッズ比が過大評価（無限大の方向にバイアス）されることになる。これがDifferentialな曝露の誤分類である。食中毒事件で食品別の喫食調査を行う場合にはこのようなことは起こりにくい。また、上記のような場合でも、思いだしを症状のない人に対してきちんとするようにすれば現実的にあまり起こりそうにない。また曝露した人と曝露し

なかった人では、調査者が曝露していない人（原因施設を利用していない人）に比べて曝露している人（原因施設を利用する人）に対して、念入りに症状を聞くという場合に、Differential な疾病の誤分類が起こりうる。しかし、食中毒事件の疫学では実際にこのようなことは起こりそうにない。

症例の定義

既に紹介した症例の定義の問題は、チャンスとバイアス、とりわけ情報バイアスの問題と密接に関連している。例えば、腸管出血性大腸菌 O157:H7 のアウトブレイクで、症例の定義に時間的に制約をつけずに、潜伏期間の2-3倍の期間に発生した症例を全て含んだとしよう。この時、原因食品で推定されるオッズ比は、1の方向にバイアスされる。なぜなら、原因調査では1次感染の症例が問題になるのに、長い期間の症例を含めたために2次感染の症例までも含んでしまった。本来これらの症例は、1次感染を念頭に置けば「症状なし」と分類されるべき者を「症状あり」として誤分類したことになる。この誤分類はNon-differential な疾病の誤分類である。同様に腸管出血性大腸菌のアウトブレイクで、症例の定義の症状の部分を狭くして「血性の下痢かつ便菌陽性」として無症状の人と曝露歴を比較すれば、腸管出血性大腸菌以外の下痢が症例の中に入ってくる可能性は非常に低い。しかし、症例の定義を「下痢1回以上」とすると、腸管出血性大腸菌以外の下痢症例が含まれる可能性が生じた分だけ、わずかではあるが誤分類が生じている。この誤分類はNon-differential な疾病の誤分類であり、オッズ比を1の方向にバイアスする。

このように見ると、症例の定義は狭い方が良いのではと思いがちだが、症例の定義を狭くすればするほど、症例の数は少なくなり、その分だけ信頼区間が広がって、推定値が不安定になる。症例の定義を変える以外に症例の数を増やすことはあり得ないので、症例数がさほど多くない食中毒事件では、症例の定義を広く取らざるを得ない。

また、上記のようなNon-differential な疾病の誤分類は、原因食品に対してのみしか明確に起こらないので、症例の定義を広く取った場合のオッズ比と狭く取った場合のオッズ比を比較して、症例の定義を狭く取った場合のオッズ比の上昇がある食品に関して観察できれば、その食品が原因食品であると推論しやすくなる。症例の定義を工夫することは、このようにして原因究明の際に推論を深めることに繋がるので、是非、習得すべきである。

交絡バイアス

交絡要因によって生じてくるバイアスを言う。

交絡要因が交絡要因として交絡バイアスをオッズ比など疫学的影響の指標に対してバイアスを生じさせるためには、以下の3つの条件を満たしていなければならない。

- ① 交絡要因は、疾病のリスク要因でなければならない。
- ② 交絡要因は、曝露と関連していなければならない。

③ 交絡要因は曝露と疾病の中間要因であってはならない。

上記の3つの条件のうち、①と②を図にして示すと、図2-5-5のようになる。

交絡バイアスによって、オッズ比など疫学的影響の指標がバイアスされる方向は、無限大の方向もしくはゼロの方向である。どちらの方向にバイアスされるかは、交絡要因が疾病・症状に及ぼす影響の方向と、交絡要因と原因曝露との関連の方向による。

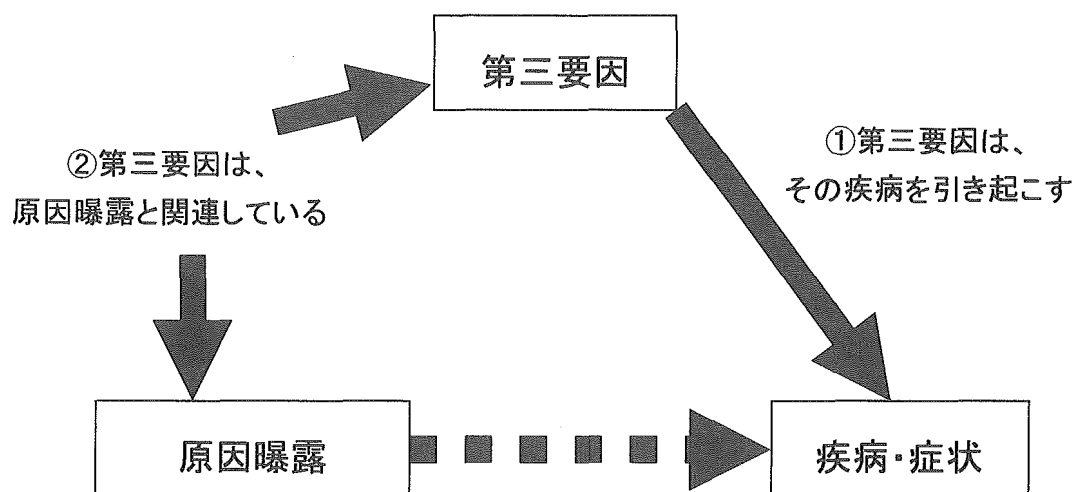


図 2-5-5 交絡要因の定義を示す図

これを事例で示す。曝露をパンとし、疾病を下痢かつおう吐とする（実際はサルモネラ菌による症状）。この食中毒事件の時の原因食品は、生卵を混ぜたオクラ納豆であった。この時、パンと症状の間には、交絡要因であるオクラ納豆により交絡バイアスが成立する。まず、①オクラ納豆は原因食品なので症状を発生させるリスク要因である。また②洋食であるパンを食べる人は、和食であるオクラ納豆を共に食べることはほとんどないので、パンとオクラ納豆は負の関連がある、③オクラ納豆はパンを食べることと症状との中間要因ではない。以上から、交絡要因オクラ納豆により、パンと症状とのオッズ比に対する交絡バイアスが生じ、オッズ比は本来 1 のはずなのに、ゼロの方向に過小評価され、一見、パンが症状の発生を抑制したかのようなオッズ比が推定された。これは交絡要因オクラ納豆により、パンを食べることと症状の発生との間にニセの関連が生じたことになる。

選択バイアス

全数を調べる後ろ向きコホート調査の場合は、欠席者まできちんと調べれば、選択バイアスが生じることは、まずあり得ない。ここでは症例対照調査の例で説明する。

症例対照調査の対照の集め方には、大きく分けて3つの種類があったことはすでに説明

した。そのうち、症例の友人・近所から対照を集める場合には、症例と同じような曝露を受けている可能性がある。ところが、症例対照調査の際の対照の集め方の原則は、曝露とは関連しない（独立した）人たちを集めることである。従って、曝露が症例を彦起こしている場合（曝露が原因食品である場合）には、曝露と症例は関連し、かつ、同じように曝露と対照も関連してしまっているため、曝露と対照との独立性は失われてしまっている。この場合、算出されるオッズ比など疫学的影響の指標は、過小評価される。病院対照を利用する際にも、対照として利用する病名が曝露と関連している病気である場合には、同様に曝露による症例へのオッズ比を過小評価することが生じる。これらは、対照の選択が曝露と関連したものとなってしまったために生じた。対照の選択バイアスとも言う。

まとめ

- ①アウトブレイク調査では検定よりもむしろ疫学指標の推定が用いられる。理由は情報量が多いからである。点推定値と区間推定値以外に、時には2かけ2表も示した方が報告書の情報は多くなる。
- ②「有意差がある」、「有意差がない」ということと「因果関係がある」、「因果関係がない」ということの混同がしばしば観察される。
- ③チャンス（偶然の変動）とバイアス（系統的誤差）は報告書で分けて考察する。
- ④疫学分析全般に関連するバイアスとして、選択バイアス、情報バイアス（誤分類）、交絡バイアスを報告書の考察で行う。
- ⑤理論的には、症例の定義を AND でつなぎ狭くすることは、情報バイアスの non-differential な疾病の誤分類を少なくすることであり、OR でつないで広くすることは、チャンスを小さくすることである。
- ⑥バイアスを考察することは、仮象（関連がなくても見せかけ上関連がある場合と、関連があっても見せかけ上関連がない場合がある）を見抜くことである。