

疫学研究における生物統計手法

佐藤俊哉*

Biostatistical Methods in Epidemiologic Research: A Review

Tosiya Sato*

This paper reviews biostatistical methods in epidemiologic studies for causal inference between a risk factor and a disease. After a brief introduction into measures of exposure effect which are of interest in epidemiologic research, classical epidemiologic designs, *i.e.* cohort and case-control studies, for effect estimation are illustrated. Recent developments in biostatistics and epidemiology allow to propose new study designs, such as nested case-control, case-cohort, and two-stage case-control studies. Some aspects of causal inference in epidemiologic observational studies are discussed. Other important topics, misclassification errors, ecological bias, and longitudinal studies, are briefly reviewed.

リスク要因と疾病発生との因果関係を調べるための疫学研究で用いる生物統計手法に関するレビューを行う。疫学研究に興味のある、リスク要因への曝露の効果の指標を導入した後、曝露効果を推定するための古典的な研究デザインである、コホート研究、ケース・コントロール研究とそこで用いる生物統計手法を解説する。最近では、生物統計学の発展にともなって、コホート研究、ケース・コントロール研究に代わる新しい研究デザインがいくつか提案されているが、新しいデザインのうち代表的なネステッド・ケース・コントロール研究、ケース・コホート研究、2段階ケース・コントロール研究の紹介を行う。また、疫学的観察研究から因果推論を行うための最近の研究成果についても報告する。その他の重要な話題である、誤分類の影響、Ecological bias、経時観察研究、についても簡単ではあるが文献紹介を行う。

1. はじめに

疫学とは「特定の集団内の健康に関連した状態または事象の分布や決定要因に関する研究であり、健康問題をコントロールするためにその研究結果を応用する (Last [46])」研究領域である。疫学研究は具体的には、循環器疾患に影響を与える要因は何か？、がんを予防する生活習慣は？、現在難病を患っている人はどのくらいいるか？、などの具体的な公衆衛生上の問題に対し回答を与えるための応用科学であり、非常に広い範囲を対象としている。

疫学は歴史的にコレラなどの急性感染症の流行を記述し、原因を明らかにし、流行を防ぐ学問として誕生した。疫病学、流行病学という疫学の別名がこのことを物語っている。現在でも予防接種の有効性評価や AIDS のように、感染症の疫学は重要な分野である。しかし、成人病と呼ばれる慢性非感染症の増加にともなって、疫学の重点は、がん、循環器疾患、などの原因と予防を研究することに移りつつある。1950年代から60年代にかけて行われた喫煙と肺がんに関する多数の疫学研究は、この慢性疾患の疫学の確立に貢献した (Breslow and Day [8], p. 9)。現在では喫煙が肺がんの非常に強い原因の一つであることを疑う人はいないであろうが、

当時は喫煙と肺がんの因果関係の判定をめぐるいわゆる「喫煙と肺がん論争」が起こった。著明な統計学者である R. A. Fisher, J. Berkson, H. J. Eysenck らは、喫煙と肺がんの因果関係を立証するためには疫学的観察研究だけでは不十分であると述べた（一般に信じられているように彼らは喫煙と肺がんの因果関係を否定しているわけではない）が、この論争は、観察研究にもとづいてどのように因果推論を行うか、という現在最も重要な話題に直接結びついている。「喫煙と肺がん論争」に興味のある読者は、最近の議論(Vandenbroucke [96], Greenland [27], Stolley [87], Eysenck [16]) およびそれらの論文に引用されている文献を参照してほしい。

確かに疫学的観察研究だけで因果関係を証明することは不十分であるかもしれない。因果関係を科学的に判断するためには要因のランダム化が不可欠である。例えばランダム化臨床試験は、新薬や新治療法の効果を科学的に判定するための手段として、現在では広く承認されている(広津 [39])。しかし、疫学研究では、一般に健康に悪影響を与える要因、リスク要因、を研究の対象としているので、リスク要因を対象者にランダムに与えることは倫理的に許されない。したがって、疫学的観察研究は多くの場合リスク要因と疾病発生の因果関係を調べる唯一の研究方法である。さらに疫学研究では直接ヒトを研究の対象としている、という利点が存在する。動物実験から得られた結果をヒトへ外挿するためにはたくさんの障害があるし、「健康問題をコントロールする」という疫学研究の最終目的からは人間集団の観察に頼らざるを得ない。このような点から、疫学研究は「自然の実験」とも呼ばれている。

本論文では、非常に広い範囲の疫学研究の中でも「ある要因に曝露されることが疾病発生の原因であるかどうか、という因果関係に関する研究」に限定して、その中で用いられている生物統計手法のレビューを行う。曝露-疾病間の関係を調べるためには、疾病発生の状況をどのようにとらえるか、また疾病発生に曝露が与える効果の大きさをどうやって量るかが大きな問題となり、まず最初に2節では疫学研究で用いる疾病発生の指標、曝露の効果の指標について述べる。続いて3節、4節では曝露の効果을推定するための古典的なコホート研究、ケース・コントロール研究という2つの代表的な研究デザインについてまとめを行う。5節ではコホート研究やケース・コントロール研究の利点をミックスした新しい疫学研究のデザインについて簡単な紹介を行い、6節で因果推論に関する最近の話題を紹介する。

その他の重要な話題についても7節で簡単に触れるが、本論文だけで「疫学研究における生物統計手法」を全てカバーすることは不可能である。そこで、基本的な教科書および関連した話題のレビュー論文を紹介し、読者の（というよりも、このレビューの）助けとしたい。急性感染症（あるいはもっと以前）から始まる疫学の歴史については丸井 [60] がよくまとまっていて読みやすい。疫学の教科書としては、入門的なものとして Kelsey et al. [42], Walker [102], 多少専門的なものに Kleinbaum et al. [43], Rothman [76] がある。また、Breslow and Day の教科書 [7, 8] はたいへん高度な内容を含んでいるが、どちらも第1章はそれぞれケース・コントロール研究、コホート研究のたいへんいい（やさしい）紹介となっている。疫学研究における生物統計手法の最近の発展をまとめたものには、Breslow [5], Gail [19] がある。Breslow [5] は5節で紹介する新しい研究デザインと本論文では触れられなかったグラフィック表示を強調したレビューを行っている。Gail [19] はロジスティック回帰、Cox 回帰にもとづくモデリングの立場から1980年代の生物統計手法の発展のレビューを行っている。Gail のレビューは851の論文をテーマ別に分類した文献集にもなっているので、興味のある読者の参考になるだろう。

2. 疫学研究で興味のある指標

2.1. 疾病発生の指標

疾病発生状況をとらえる考え方には2通りある。ある時点において、特定個人が研究対象としている疾病にかかっていることを有病 (prevalence) という。集団検診やスクリーニングで発見された疾病は検診時での有病状況であるし、剖検 (死後に原死因や付随した疾病の確認のために解剖すること) で確診された疾病は死亡時の有病状況である。一方、研究対象としている疾病に現在かかっていない人で、将来その疾病にかかる可能性のある (at risk) 人を追跡して新たに疾病にかかった場合、それを疾病発生 (incidence) という。(例えば子宮がんの発生を調べたい場合、男性は当然として、子宮摘出を行った女性も新たに子宮がんにかかる可能性はないので、子宮がんの「at risk」とはならない。)

疾病の有病数を正確に把握することは、例えば医療費補助を行う場合のコスト計算のように公衆衛生行政の観点から重要な問題である (Walker [102])。しかし、リスク要因への曝露と疾病発生との因果関係を調べる場合には、1) 有病状況は疾病の致命率 (疾病にかかった者が死亡する割合) や治癒する割合など治療技術の進歩に大きく影響を受ける、2) 疾病発生を把握するということは疾病発生前後の状況も観察できることにほかならない、といった理由から有病よりも疾病発生を用いることが好まれる (Rothman [76], pp. 33-34)。ある1時点の有病状況とリスク要因の同時分布を調べる断面研究 (cross-sectional studies) では、3節で述べるコホート研究と同様の手法を用いて、有病割合とリスク要因との関係を調べる場合がある。ただし、断面研究の主たる目的は同時分布の把握という実態調査であり、有病割合とリスク要因との関係は仮説スクリーニングを目的としたあくまでも2次的なものである。

ある集団における疾病発生状況についても、2通りのとらえ方がある (Elandt-Johnson [15])。一つは疾病発生割合 (incidence proportion または cumulative incidence rate) であり、研究開始時に将来疾病にかかる可能性のある対象者の集団が一定期間中に疾病を発する割合の期待値である。以下では「将来疾病にかかる可能性のある対象者の集団 (population at risk)」を「リスク集団」と呼ぶことにする。疾病発生割合は少なくとも転入あるいは登録に関して閉じた集団でなければ定義することはできない。もう一つは、疾病発生率 (incidence rate) であり、リスク集団の観察人-時間1単位当たりの期待疾病発生数である。疾病発生割合は無次元数であるが、疾病発生率は時間の逆数の次元を持つ。したがって疾病発生率の逆数は平均生存時間として解釈できるので、生存時間解析と直接結びつけることができる。疾病発生率は転入に関して閉じた集団でも、転入、転出などを許す開いた集団 (ダイナミックコホート) でも定義可能である。

2.2. 曝露の効果の指標

簡単のためリスク要因への曝露を受けたグループと受けていないグループ間で疾病発生状況を比較することを考えよう。曝露グループの疾病発生割合を IP_1 、非曝露グループの疾病発生割合を IP_0 とする。曝露により疾病の発生がどの程度増加するかを定量的に計る指標としては、2グループの疾病発生割合の単純な対比である、疾病発生割合の差 $IPD = IP_1 - IP_0$ と疾病発生割合の比 $IPR = IP_1 / IP_0$ が用いられる (Greenland [26])。疾病発生割合の差は曝露による疾病発生割合の絶対的な変化であり、「曝露を受けることで疾病発生が何%増えるか」を表す。また疾病発生割合の比は曝露による疾病発生割合の相対的な変化であり、「曝露を受けることで疾病発生が何倍になるか」を表す。疾病発生割合は閉じたりリスク集団でしか定義できないので、疾病発生割合の差や比も閉じたりリスク集団での曝露効果の指標である。例えば、リスク要因への曝露が胎児に与える影響を調べる場合や急性中毒のように、一定の期間の間に疾病を発生するか

どうか(出産時に生存しているかどうか, 短時間に中毒を起こすかどうか)が重要な場合に有効な効果の指標となる。

一方慢性疾患の疫学研究のように観察期間が長期にわたる場合, 対象者の登録には長い時間がかかり, また対象者の死亡, 転出などの観察打ち切りが問題になってくる。このように転入, 転出, 死亡などを許す開いた集団を長期間観察する必要がある場合には, 疾病発生割合は定義できないので曝露グループと非曝露グループ間の疾病発生率を比較する。曝露グループの疾病発生率を IR_1 , 非曝露グループの疾病発生率を IR_0 とすると, 疾病発生割合と同様に, 疾病発生率の差 $IRD = IR_1 - IR_0$ と疾病発生率の比 $IRR = IR_1/IR_0$ を曝露の効果の指標として用いる。慢性疾患の疫学ではある期間内に疾病を発生するかどうかよりも, 疾病が発生するまでの時間がどのくらいかに興味があるので, 疾病発生率と生存解析との関係から IRD や IRR が適切な効果の指標となる。

次節以降では, さまざまな疫学研究のデザインから疾病発生割合の差 IPD , 疾病発生割合の比 IPR , 疾病発生率の差 IRD , 疾病発生率の比 IRR , を推定する方法を述べる。

3. コホート研究

説明を簡単にするためリスク要因への曝露状況は曝露を受けたか ($E=1$), 受けなかったか ($E=0$), 疾病に関しても疾病を発生したか ($D=1$), しなかったか ($D=0$), にそれぞれ2分類された場合を考える。2節で述べた疾病発生割合や疾病発生率を推定するための自然な研究デザインは, 曝露を受けたリスク集団と曝露を受けなかったリスク集団を一定期間追跡して, 研究対象としている疾病を発生した数を数え上げることである。このような研究デザインはコホート研究 (cohort studies, または前向き研究 prospective studies) と呼ばれ, リスク要因への曝露から疾病発生まで, 時間の流れと因果関係を調べる順序が一致していて, 実験に最も近い自然な研究デザインである。コホートとはラテン語で古代ローマの軍隊の一分隊を指すが (Last [46]), 疫学研究では「特定の集団」という意味で用いる。

3.1 単純な解析

曝露グループ, 非曝露グループの研究開始時のリスク集団の総数を N_1, N_0 , 観察期間内の期待疾病発生数を A, B で表そう。すると, 疾病発生割合は曝露グループで $IP_1 = A/N_1$, 非曝露グループで $IP_0 = B/N_0$ となる。コホート研究では IP_1, IP_0 ともに推定することができるので, 疾病発生割合の差 $IPD = IP_1 - IP_0$ も, 疾病発生割合の比 $IPR = IP_1/IP_0$ も簡単に推定できる。6節で詳しく述べる「交絡 (confounding)」が無視できる場合, グループ i ($i=1$: 曝露グループ, $i=0$: 非曝露グループ) の疾病発生数は独立に2項分布 $Bi(IP_i, N_i)$ に従うと考えるとよいことから (Robins [65]), IPD および IPR に関する仮説検定や信頼区間を求めることができる。我々が知りたい「曝露により疾病発生は増加しない」という帰無仮説は「 $IPD=0$ 」または「 $IPR=1$ 」という仮説と同値であり, この帰無仮説の検定には Fisher の正確な p 値, χ^2 検定などを用いる (Rothman [76])。 IPD や IPR の信頼区間には Wald 信頼区間 (IPR の場合は $\log IPR$ に対して) がよく用いられるが, efficient score にもとづく信頼区間も提案されている (Gart and Nam [21, 22])。

同じ対象集団に対し, 曝露グループのメンバー全員の「at risk」の状態を観察できた時間の合計を M_1 , 同様に非曝露グループの「at risk」の合計観察時間を M_0 としよう。曝露グループ, 非曝露グループの疾病発生率は $IR_1 = A/M_1, IR_0 = B/M_0$ であり, コホート研究では疾病発生割合と同様に疾病発生率の差 $IRD = IR_1 - IR_0$, 疾病発生率の比 $IRR = IR_1/IR_0$ も推定することができる。交絡が無視できる場合には, グループ i の疾病発生数は独立に期待値 $M_i IR_i$ のポアソン分布に従うと考えるとよいので (Breslow and Day [8], pp. 131-135), 「曝露により疾病発生

は増加しない」という帰無仮説検定のための Fisher-p 値や χ^2 検定, また IRD , IRR の信頼区間を求めることができる。 IRD および IRR に関する推測は (IR_1, IR_0 についても), 登録について閉じたコホートだけではなく, 転入, 転出などの人口動態を全て許したダイナミックコホートでも行うことができる。

3.2. 層別解析

疫学的観察研究では実験研究と異なり, 曝露-疾病間の因果関係に影響を与えるかも知れない他の要因を全てコントロールして観察を行うことは不可能である。曝露状況と関連している疾病のリスク要因が存在する場合, その要因を無視して曝露-疾病間の解析を行うと, 効果の推定にはバイアスが入る(6節参照)。例えば, 曝露グループに年齢の高い者が集中していると, 曝露が死亡を増やさなくとも見かけ上非曝露グループよりも死亡割合が高くなるだろう。このような現象は「交絡」として疫学では古くから知られているが, バイアスのない効果の推定を行うには交絡要因で層別解析を実施したり, モデルを用いて交絡要因の調整をする必要がある。層別解析は簡単な交絡要因調整の方法であり, コホートデータを交絡要因のレベルに応じて, それぞれの層の中では交絡が無視できるような層に分ける。各層ごとには曝露グループと非曝露グループ間で偏りのない比較ができるようになるが, もし全ての層で疾病発生割合の差 IPD や比 IPR が一定の値であれば共通な IPD , IPR の推定値を求めることが正当化される。

共通 IPD , IPR の推定方法には最尤推定法, 重み付き最小2乗法, Mantel-Haenszel の方法などがある (Rothman [76])。層別化の方法が「層の数は数個で, 各層に十分な大きさの標本数がある (large-strata)」場合, どの推定法からも共通な IPD , IPR の一致推定量が導かれ, しかも最尤推定量, 重み付き最小2乗推定量は漸近有効となる。Mantel-Haenszel 推定量は, large-strata での効率はしばしば最尤推定量に比べてずいぶん落ちるものの, 「各層の標本数は少ないが, 層の数がたくさんある (sparse-data)」場合でも共通 IPD , IPR の一致推定量である。例えば, 曝露を受けた個人に対し交絡要因と考えられる変数 (年齢, 喫煙歴, など) をマッチさせて曝露を受けていないコントロールを数名選ぶデザインは, マッチをとった一組を1つの層とした sparse-data の典型例である。[Sparse-data と小標本を混同してはならない。マッチしたデザインは sparse-data ではあっても, 全体の標本数は十分大きい] 最尤推定量, 重み付き最小2乗推定量は sparse-data のもとではもはや一致性を持たない (Greenland and Robins [34])。共通 IPD , IPR の Mantel-Haenszel 推定量には large-strata, sparse-data 両方で用いることのできる信頼区間の方法が提案されている (Greenland and Robins [34]; Sato [81])。

疾病発生率を問題にしている場合も疾病発生割合と同様に交絡に対処する必要がある。層別解析で交絡を調整する場合, 各層で疾病発生率の差や比 IRD , IRR が一定であれば最尤推定法, 重み付き最小2乗法, Mantel-Haenszel の方法から共通 IRD , IRR の推定値を求めることができる。ただし, 共通 IRD , 共通 IRR の重み付き最小2乗推定量は large-strata では一致推定量であるが, sparse-data では一致推定量ではない。共通 IRR の最尤推定量は large-strata, sparse-data とともに一致推定量であるが, 共通 IRD の最尤推定量は sparse-data で一致推定量とはならない (Sato [83])。一方, 共通 IRD , IRR の Mantel-Haenszel 推定量は large-strata, sparse-data とともに一致推定量であり, どちらの場合にも用いることのできる信頼区間の方法が与えられている (Breslow and Day [8]; Greenland and Robins [34]; Sato [82])。

もし各層で効果の指標が一定とは仮定できない場合, 効果の方向が層によってまちまちであればもはや単一の値に要約することはできない(質的な交互作用)。しかし, 交互作用は存在するが効果の方向は全ての層で同方向である場合 (量的な交互作用, または効果の修飾) には, 交絡の影響をとり除いて全体的な効果を示す意義があり, そのための方法が SMR (standard-

ized morbidity/mortality ratio) などに代表される「標準化 (standardization)」という考え
方である (Miettinen [57]; Greenland [23, 29]).

3.3 疾病発生リスクのモデル化

層別解析による交絡要因の調整は簡便ではあるが、調整すべき交絡要因が全て2値変数であ
ったとしても、10変数の交絡要因を調整すると1024の層ができてしまう。交絡要因でマッチを
とったデザイン以外では、多くの層で標本がゼロとなり、もはや現実的な交絡の対処方法では
ない。多くの交絡要因を同時に調整する必要がある場合、連続量の交絡変数を扱う場合、など
にはモデルを用いた交絡の調整が不可欠である。疫学研究で用いる代表的なモデルとしては、
個人レベルの疾病発生リスク(疾病発生確率)をモデル化した絶対リスクモデル (additive risk
model),

$$P(D=1|E, Z) = \alpha_A + \beta_A E + \gamma_A Z,$$

相対リスクモデル (relative risk model),

$$\log P(D=1|E, Z) = \alpha_R + \beta_R E + \gamma_R Z,$$

がある (Wacholder [98]). ここで、 $P(D=1|E, Z)$ は曝露 E と交絡要因の組 Z を持つ個人が疾
病を発生する確率、 E は曝露を受けたとき1、受けなかったとき0である。これらは曝露と交
絡要因との間に交互作用がない場合のモデルで、一般には交互作用項、例えば δEZ を含めて
モデル化を行う。曝露の効果を表すパラメータは、絶対リスクモデルのもとでは疾病発生リス
ク差、

$$\beta_A = P(D=1|E=1, Z) - P(D=1|E=0, Z),$$

相対リスクモデルのもとでは疾病発生リスク比、

$$\exp(\beta_R) = \frac{P(D=1|E=1, Z)}{P(D=1|E=0, Z)},$$

であるが、これらは集団レベルの疾病発生割合の差 IPD 、疾病発生割合の比 IPR と同じものを
推定している (Greenland [26]).

疾病発生リスクには $0 \leq P(D=1|E, Z) \leq 1$ という制約があるが、絶対リスクモデル、相対リス
クモデルはこの制約を常に満たすとは限らない。疾病発生リスクそのものや対数疾病発生リス
クをモデル化する代わりに対数疾病発生オッズを、

$$\log \frac{P(D=1|E, Z)}{1 - P(D=1|E, Z)} = \alpha_L + \beta_L E + \gamma_L Z,$$

とモデル化することで、カテゴリカルデータ解析で最も用いられているロジスティックモデル
を得る。ロジスティックモデルでは、 $0 \leq P(D=1|E, Z) \leq 1$ の制約を自動的に満たすが、交絡要
因で調整した効果の指標、

$$\exp(\beta_L) = \frac{P(D=1|E=1, Z)[1 - P(D=1|E=0, Z)]}{[1 - P(D=1|E=1, Z)]P(D=1|E=0, Z)},$$

は疾病発生オッズ比であり、疫学的に簡単な解釈はできない。しかし、研究対象としている疾
病の発生が曝露グループ、非曝露グループともに非常にまれである場合、 $1 - P(D=1|E=1, Z)$ 、
 $1 - P(D=1|E=0, Z)$ とともにほぼ1であるから、 $\exp(\beta_L) \cong \exp(\beta_R)$ となって近似的に疾病発生

リスク比として解釈できる (Cornfield [12])。

疾病発生がまれでない場合には、ロジスティックモデル自体に疫学的な解釈を行うことはできないが、ロジスティックモデルが正しければ (データに対して当てはまりがいい場合) 層別解析と同様にロジスティックモデルを用いて、標準化した疾病発生割合の差や比を求めることができる (Greenland [29]; Greenland and Holland [32])。モデルが正しいとき、このような標準化は層別解析でのノンパラメトリックな標準化よりも精度のいい推定値を与える。

疾病発生率を用いる場合も、多くの交絡要因を同時に調整したい場合や連続量の交絡変数を調整する場合、個人の疾病発生率をモデル化する必要が生じる。時刻 t における個人レベルの疾病発生率 $r(t|E, Z)$ に、時間に関係した部分と曝露効果など時間に関係しない部分との乗法モデルを仮定した、

$$\log r(t|E, Z) = \log a(t) + \beta_c E + \gamma_c Z,$$

が生存時間解析で用いられる Cox 回帰 [13] である。Cox 回帰は $E=0, Z=0$ のときの疾病発生率 $a(t)$ (ベースラインの疾病発生率) にはノンパラメトリックに特別な仮定を置かず、曝露効果などにはパラメトリックな仮定を置いたセミパラメトリックモデルであるが、推定には部分尤度 (partial likelihood) にもとづいた複雑な計算を必要とする。このほか放射線被曝や職業曝露のモデル化では疾病発生率に加法モデルを仮定した過剰リスク (excess risk) モデルを用いる場合もある。共変量データが個人レベルではなく、 (E, Z) がカテゴリ化された集団レベルで得られているときには、疾病発生率のモデル化にはポアソン回帰を用いることができる (Breslow and Day [8])。

3.4 コホート研究の例

おそらく最も有名なコホート研究の例は、1948年から Boston 郊外の住宅地 Framingham で約 4500名の白人を対象に実施された Framingham Heart Study であろう (NHLBI [61])。Framingham Study では冠状動脈性疾患発生のリスク要因を探索することを目的に、研究開始時に冠状動脈性疾患を発症していない対象者の喫煙歴、血圧、血清コレステロールレベルなどを測定した。引き続いたフォローアップの結果、研究開始時に血圧及びコレステロールレベルの高い人は冠状動脈性疾患を発生しやすいことが分かった。したがって、血圧やコレステロールレベルが高いことは、それまで考えられていたように疾病の結果ではなく冠状動脈性疾患の病因の一部であることが判明した。疫学的、医学的な知見だけでなく、Framingham 研究ではリスク要因の解析方法としてロジスティックモデルが用いられたことでも有名である。Framingham 研究は、現在では全米4ヶ所で黒人を対象に含めた16,000人の成人を追跡している、動脈硬化症のリスク要因を探るための Atherosclerosis Risk in Communities Study (ARIC [92])に受け継がれている。

我が国の大規模コホート研究の例としては「コホート研究による発がん要因の評価に関する研究班 (文部省がん特別研究総括班内研究班、班長青木國雄)」が挙げられる。このコホート研究は複数のがんのリスク要因を探索する目的で、10万人の対象者を10年以上追跡することを目標に1988年から研究を開始している。その中では血清の凍結保存を行うなど、5節で述べる新しい研究デザインに対応できるように配慮されている。

本節ではコホート研究を観察研究として紹介したが、曝露をランダムに割り付けることが可能な場合、例えばランダム化臨床試験、もコホート研究の一部として考えることができる。ランダム化臨床試験に関しての統計手法や問題点は広津 [39] のレビューを参照してほしい。ランダム化臨床試験は比較的控制可能な臨床の場で治療を目的として行われるが、地域集団など実際に公衆衛生対策を行う現場で疾病の予防法をランダムに割り付けて比較を行う、

介入研究 (intervention trials) も広い意味でのコホート研究であり、そのデザイン、解析方法などは本節の内容を用いることができる。

コホート研究は集団を前向きに追跡しているため、1つの曝露について広範な健康影響 (複数の疾病) を調べることができる、疾病発生前の生体の生理的・生化学的情報が得られる (疾病の自然史を調べられる)、くり返し測定ができる、などの利点がある。また、曝露、非曝露別に対象者をサンプルできるので、曝露がまれな場合 (職業曝露のように集団中の一部だけが曝露を受けているような状況) でも有効なデザインである。一方、対象としている疾病がまれである場合には、大規模なコホートを長期間追跡しなければ十分な疾病発生が観察できないので、時間とコストがかかるという欠点を持っている (Breslow and Day [8])。

4. ケース・コントロール研究

コホート研究では曝露グループと非曝露グループを一定期間追跡して疾病発生状況を調べたが、同じコホートについて観察期間が終了した時点で、観察期間中に疾病を発生したコホートメンバーからのランダムサンプル (ケースグループ) と観察終了時に疾病を発生していないメンバーからのランダムサンプル (コントロールグループ) について過去の曝露状況を比較することを考えてみよう。これが古典的なケース・コントロール研究 (case-control studies, または後ろ向き研究 retrospective studies) のデザインであるが、通常はケース・コントロール研究の背景に現実のコホートを設定することは難しい。したがって、ケース・コントロール研究ではまず収集可能なケースがあり、そのケースが発生するような “base population” は何か、を考察することにより適切なコントロールグループを見いだす必要がある (Rothman [76]; Wacholder et al. [99-101])。すでに疾病を発生しているケースが利用できるため、ケース・コントロール研究は「まれな疾病」に対して有効な研究デザインである。

4.1 単純な解析

ケース・コントロール研究では疾病発生割合を推定することはできないので、外部からの補助情報がない限り疾病発生割合の差や比を推定することはできない。推定可能なのは曝露オッズ比、

$$OR_E = \frac{P(E=1|D=1)[1-P(E=1|D=0)]}{[1-P(E=1|D=1)]P(E=1|D=0)},$$

であり、この曝露オッズ比は観察打ち切りが曝露とも疾病発生とも独立に起こるときコホート研究から得られる疾病発生オッズ比、

$$OR_D = \frac{P(D=1|E=1)[1-P(D=1|E=0)]}{[1-P(D=1|E=1)]P(D=1|E=0)} = OR_E,$$

に一致する (Cornfield [12]; Greenland and Thomas [36])。したがって対象としている疾病の発生がまれであれば、3.3節に示したように、曝露オッズ比 OR_E を疾病発生割合の比 IPR の良い近似として用いることができる。ケース・コントロール研究は「まれな疾病」に対して有効な研究デザインであるので、「まれな疾病」のもとでの $OR_E \cong IPR$ の関係は疫学的な曝露の効果の解釈に重要である。

ケースとコントロールの曝露割合の差の検定、

$$H_0: P(E=1|D=1) = P(E=1|D=0),$$

は「 $OR_E = 1$ 」の検定と同値であり $OR_E = OR_D$ の関係から、曝露グループと非曝露グループの

疾病発生割合の差の検定「 $P(D=1|E=1)=P(D=1|E=0)$ 」とも同値となる。 OR_E の推定にはケースとコントロールの曝露者数がそれぞれ独立な2項分布にしたがうと仮定した場合の最尤法、合計曝露者数を固定した非心超幾何分布にもとづく条件付き最尤法などを用いる (Gart [20]; Walter and Cook [103])。また、 OR_E の信頼区間はWald信頼区間の他、efficient scoreにもとづいたCornfieldの方法、非心超幾何分布にもとづいた正確な方法を用いて求めることができる (Breslow and Day [7])。

4.2. 層別解析

コホート研究と同様に、交絡が存在する場合にはケース・コントロール研究でも交絡を無視した解析は曝露効果の推定にバイアスが入る。交絡の調整には、層別解析、モデルを用いた調整、の2通りの方法が存在する。層別解析の考え方はコホート研究の場合と全く同じであり、交絡要因のレベルに応じて交絡が無視できるような層に分ける。全ての層でオッズ比の値がほぼ一定であれば、共通なオッズ比の推定値を、最尤法、条件付き最尤法、重み付き最小2乗法、Mantel-Haenszelの方法からそれぞれ求めることができる。ケース・コントロール研究でも層別の方法には、層の数は数個で各層の標本数は十分大きい large-strata、ケース1例に対し交絡要因をマッチさせたコントロールを1ないし数例選択する sparse-data が存在する。最尤法、条件付き最尤法、重み付き最小2乗法による共通オッズ比の推定量は large-strata で一致性を持ちしかも漸近有効である。しかし、最尤推定量と重み付き最小2乗推定量は sparse-data では一致性を持たない (Breslow [4])。条件付き最尤推定量と Mantel-Haenszel 推定量は large-strata でも sparse-data でも共通オッズ比の一致推定量である。特に Mantel-Haenszel 推定量は計算が簡単で、large-strata、sparse-data ともにパラメータの広い範囲で漸近相対効率が90%以上となることから、「全ての層で曝露の効果はない」という帰無仮説の Mantel-Haenszel 検定とともに疫学研究で最も良く使用されている統計手法となっている (Mantel and Haenszel [54]; Gart [20]; Breslow and Day [7]; Rothman [76])。

共通オッズ比の信頼区間としては、large-strata で利用できる Cornfield の方法を拡張したものの (Gart [20])、large-strata、sparse-data の両方で用いることができる Mantel-Haenszel の方法にもとづいた信頼区間 (Robins, et al. [68]; Sato [82]) が比較的計算が簡単で、中程度の標本数でも性質のよい方法である。最近では非心超幾何分布の積にもとづいた正確な信頼区間を高速に計算するアルゴリズムが提案され、統計パッケージにも組み込まれている (Mehta et al. [55]; Vollset et al. [97]; Takagi [89]; Mehta and Walsh [56])。正確な信頼区間も large-strata、sparse-data、どちらでも用いることができる。共通オッズ比を求めるための前提である「各層のオッズ比が共通であるかどうか」を調べるための予備検定として、有名な Breslow-Day 検定を初めとした「オッズ比の均一性検定」が提案されている (Breslow and Day [7]; Liang and Self [48]; Tarone [91]; Yanagawa and Fujii [106])。ただし、均一性検定は一般に検出力が低いことに注意して使用しなければならない (Greenland [24]; Jones et al. [41])。各層でオッズ比が共通であると仮定できない場合には、標準化したオッズ比を曝露効果の指標として用いることもできる (Miettinen [57])。

4.3. ロジスティックモデル

調整すべき交絡要因の数が多い場合、または連続的な交絡変数を調整する場合にはモデルを用いた調整を行わなければならない。疾病発生オッズ比と曝露オッズ比が等しいことから、

$$\begin{aligned}\exp(\beta_L) &= \frac{P(D=1|E=1, Z)[1-P(D=1|E=0, Z)]}{[1-P(D=1|E=1, Z)]P(D=1|E=0, Z)} \\ &= \frac{P(E=1, Z|D=1)[1-P(E=1, Z|D=0)]}{[1-P(E=1, Z|D=1)]P(E=1, Z|D=0)},\end{aligned}$$

が成立し、コホート研究でのロジスティックモデルとほぼ同じモデル、

$$\log \frac{P(E, Z|D=1)}{1-P(E, Z|D=1)} = \alpha_L^* + \beta_L E + \gamma_L Z,$$

がケース・コントロール研究からも得られる。 β_L の推定はケース・コントロールデータがあたかもコホート研究から得られたかのように扱うことで実行できる (Mantel [53])。

マッチングを行ったケース・コントロールデータの解析には、条件付きロジスティックモデルを用いて β_L を推定する必要がある (Breslow and Day [7])。条件付きロジスティックモデルは計算量が多いため、Mantel-Haenszelの方法を拡張した推定方程式を用いるロジスティックパラメータの推定方法が提案されている (Liang [47])。その一方で、条件付きロジスティックモデルで仮説検定を行うためのコンピュータ・インテンシブな正確な方法も提案され、漸近的な検定方法との比較がなされている (Hirji [38])。

4.4 ケース・コントロール研究の例

ケース・コントロール研究の有名な例として、diethylstilbestrol (DES) と膣がんとの関連を紹介しよう。思春期の女兒に非常にまれな膣がんが同一の病院で8例続けて発生し、膣がんのケース1例に対し誕生日に近いコントロールを4例マッチさせたケース・コントロール研究が実施された (Herbst et al. [37])。対象者の初経年齢、既往歴、母親の喫煙歴、妊娠時のX線被曝などさまざまな要因が比較されたが、どれも有意ではなかった。唯一有意な関連が見られたのは、妊娠初期のエストロジェン (DES) 投与であり、マッチした8組中7組はケースの母親が妊娠中にDESを投与されておりコントロールの母親は全てDES投与なし、残りの1組はケース、コントロールの母親ともにDES投与はなかった。このデータからオッズ比の推定値と上側信頼限界は無限大であるが、正確な下側97.5%信頼限界は7.5であり (Mehta and Walsh [56])、強い関連が観察された。

このようにケース・コントロール研究は、コホート研究に比べ時間もコストもかからない、まれな疾病に適している、対象としている疾病の原因と考えられる要因を複数調べることができる、という利点がある。その反面、ケース・コントロール研究では曝露と交絡要因に関する情報を過去にさかのぼって調べなければならないので情報が不正確になりがちである。コホート研究と異なり、まれな曝露が原因であるような疾病には大量のケースとコントロールを必要として有効なデザインとはならない。また、ケースは過去に原因として思い当たることをよく記憶しているが、疾病を発生していないコントロールは同じ曝露を受けていても記憶していない、という思い出しバイアスなどコホート研究では起こらないさまざまなバイアスが入りがちである (Kelsey et al. [42])、という欠点がある。

5. 新しい研究デザイン

3節、4節で述べたコホート研究とケース・コントロール研究の長所、短所は相互に補完的である (Mantel [53])。このためコホート研究を実施しながらケース・コントロール研究を行うというコホート内ケース・コントロール研究の手法が提案されている。ケース・コントロー

ル研究で問題となる曝露情報に関する不正確さは対象者を前向きにフォローアップしていることで回避できるし、詳細な情報収集は全コホートの一部についてのみでいいので曝露効果の推定効率をあまり落とさずにコホート研究で問題となるコストを減少させることができる。本節では最近になって提案された研究デザインである、ネステッド・ケース・コントロール研究 (nested case-control studies)、ケース・コホート研究 (case-cohort studies)、2段階ケース・コントロール研究 (two stage case-control studies)、を紹介する。

5.1. ネステッド・ケース・コントロール研究

ネステッド・ケース・コントロール (シンセティック・ケース・コントロールとも呼ばれる) では対象者が疾病を発生しケースとなった時点でコントロールを選択する。最も一般的なコントロール選択の方法は「ケースが発生した時点のリスク集団から1人またはそれ以上のコントロールをランダムに選択する」時点マッチングである (Liddell et al. [50]; Miettinen [58])。曝露や交絡要因に関する情報はケースと選ばれたコントロールについてのみ収集すればよいので、コストの節約が期待できる。このデザインは、閉じたコホート、ダイナミックコホートどちらでも用いることができる。ケース発生時点でコントロールをマッチさせてデータを収集するため、データの解析には4節で述べたマッチングを考慮した解析である Mantel-Haenszel の方法や条件付きロジスティックモデル、を用いてオッズ比を推定し信頼区間を求めればよい。このデザインから得られるオッズ比は正確に疾病発生率の比 *IRR* を推定しており、古典的なケース・コントロール研究と異なってオッズ比の解釈に「まれな疾病の仮定」を必要としないという利点がある (Liddell et al. [50]; Greenland and Thomas [36])。

コントロールの選択には、将来疾病を発生する対象者やケース本人もコントロールとして選ばれる可能性のあることに注意しなければならない。ケースが自分自身のコントロールとして選ばれるかもしれない、ということは一見奇妙に感じるが生存時間解析の観点から正当化することができる。コホート研究で曝露グループと非曝露グループに対し生存時間解析、Cox 回帰を実施すると想定しよう。ただし、疾病発生の直前のリスク集団 (生存時間解析では「リスクセット」と呼ばれる) 全体についての部分尤度は大変な計算となるので、リスク集団からのランダムサンプルをとり、そのサンプルについてだけ部分尤度を計算することにしよう。Cox 回帰の部分尤度と条件付きロジスティックモデルの条件付き尤度は等しいので、これはネステッド・ケース・コントロールデザインにほかならない (Thomas [93]; Prentice [63])。したがって、時点マッチングを行ってもリスク集団全体からのランダムサンプルではなく、例えば、観察期間を通じて疾病発生がなかった者 (古典的なケース・コントロール研究でのコントロールの候補) のみからコントロールを選択すると、疾病発生率の比の推定にバイアスが入る (Lubin and Gail [51])。

「ケースが発生した時点のリスク集団から1人またはそれ以上のコントロールをランダムに選択する」方法は、一度コントロールとして選ばれた人が別の時点でコントロールとして選ばれる可能性が何度でもある (リスク集団を全部用いた生存時間解析では疾病発生や打ち切りが起こるまで、全ての時点でコントロールとして選ばれていることになる)。このことからコントロール選択のオプションとして「一度コントロールとして選ばれた人は、それ以降はコントロールに選ばない(ただし、ケースにはなり得る)」、「一度コントロールとして選ばれた人は、それ以降は疾病発生や打ち切りが起こるまで常にコントロールとして用いる」などが提案され *IRR* の推定効率を上げることが期待されたが、複雑なコントロールの選択を行ってもそれほどの効率の増加は見られないようである (Robins et al. [69, 72]; Prentice [63]; Langholz and Thomas [45])。また、ネステッド・ケース・コントロール研究では交絡要因などに関する詳細な情報はケースとコントロールについてしか収集していないが、曝露と疾病発生に関してはコ

ホート全体のデータが利用できる場合がある。このような場合、コホート全体のポアソン回帰の尤度とケースとコントロールの条件付き尤度が直交することから、コホート全体の情報を加えて疾病発生率の比を精度良く推定する方法が提案されている(Thomas et al. [94])。

5.2. ケース・コホート研究

ネステッド・ケース・コントロール研究ではケースにマッチさせてコントロールを選択するため、1つの曝露と2つの疾病との関係を調べたい場合、コントロールグループも2つ必要となる。コホート研究では曝露が複数の疾病に影響するかどうかを調べることができたが、コホート研究の枠のなかでケース・コントロール研究を実施しているのだから、複数の疾病との関連を調べる場合でも1つのコントロールグループですむはずである。ケース・コホート研究はこのような考え方から提案されたデザインで、コントロールグループ(サブコホートと呼ぶ)は観察開始時のリスク集団(全コホートメンバー)からあらかじめランダムに選択する(Kupper et al. [44]; Miettinen [59]; Prentice [62])。ケースグループとしては全てのケースを用いても、全ケースからのランダムサンプルを用いてもよい。ケース・コホート研究でも、曝露と交絡要因に関する詳細な情報はケースと全コホートの一部であるサブコホートについてのみ集めればよいので、コホート研究に比べてコストの節約が期待できる。

ケース・コホートデータでは、閉じたコホートで観察打ち切りが曝露とも疾病発生とも独立な場合、サブコホートからコホート全体の曝露割合 $P(E)$ を推定できる。したがって、曝露オッズ比は、

$$OR_{\#}^* = \frac{P(E=1|D=1)[1-P(E)]}{[1-P(E=1|D=1)]P(E)} = IPR,$$

と正確に疾病発生割合の比に一致し、これには「まれな疾病の仮定」を必要としない。 $OR_{\#}^*$ の最尤推定量とその近似的な分散は closed-form で与えられている(Sato [85])。また層別解析で交絡要因を調整する場合、各層に共通な $OR_{\#}^*$ は large-strata では最尤法や重み付き最小2乗法、large-strata, sparse-data 両方で利用できる Mantel-Haenszel の方法から推定できるし、効果の修飾がある場合には標準化した疾病発生割合の比を推定することができる(Greenland [25]; Sato [84])。これらの方法は競合危険(competing risk)による観察の打ち切り(研究対象ではない疾病の発生や死亡によりリスク集団から除かれること)が効果の推定にあまり影響を与えない場合の推定方法であるが、競合危険の影響を補正する方法も提案されている(Flanders et al. [17])。

ネステッド・ケース・コントロール研究と同様に生存時間解析の考え方から、ケースとケースが発生した時点のサブコホート中のリスク集団とを比較することで疾病発生率の比を推定することができる。この場合にも「まれな疾病の仮定」は必要とせずに疾病発生率の比を推定できるが、推定方法は複雑であり現在のところ統計パッケージには組み込まれていない(Prentice [62])。

5.3. 2段階ケース・コントロール研究

2段階ケース・コントロール研究は、第1段階で対象者全体に対し曝露と疾病に関してスクリーニングを行い、第2段階でその対象者の一部について交絡要因などの詳細な情報を収集する研究デザインである。ネステッド・ケース・コントロール研究やケース・コホート研究も、最初に全コホートメンバーに関する曝露情報を収集して疾病発生をフォローアップし、曝露や交絡要因などに関する詳細な情報は全コホートの一部からしか収集していないので、広い意味では2段階デザインと考えることができる(Flanders and Greenland [18]; Zhao and Lipsitz [107])。典型的な2段階ケース・コントロール研究は、第1段階で「曝露を受けたケース」、 「曝

露を受けた非ケース]、「曝露を受けていないケース]、「曝露を受けていない非ケース]をスクリーニング的に同定する。続いて第2段階では「曝露を受けたケース]にウェートを置いて数多くサンプルし詳細な情報を得ることで、曝露も疾病の発生もともにまれである場合に有効なデザインとすることができる (Miettinen [59]; White [104])。コホート研究は「曝露がまれである場合]、ケース・コントロール研究は「疾病発生がまれである場合]に有効なデザインで、それぞれ逆の場合には不適切であったが、2段階ケース・コントロール研究はこの欠点を解消したデザインである。

第2段階のサンプリングが古典的なケース・コントロール研究のようになされた場合、ロジスティックモデルによる曝露効果(オッズ比)の推定が可能である。ロジスティックパラメータの推測に関しては最尤法 (Breslow and Cain [6]) が提案されており、通常のロジスティックモデルに第2段階での対象者のサンプル割合を定数項 (offset) として加えることで、既存の統計パッケージを使ってバイアスのない推定値、分散共分散行列を得ることができる。このほかにも pseudo-likelihood にもとづく方法 (Flanders and Greenland [18])、推定方程式にもとづく方法 (Zhao and Lipsitz [107]) が提案されており、第2段階のデータをネステッド・ケース・コントロールのように時点マッチングを行ってサンプルした場合、pseudo-likelihood にもとづく推定法はポアソン回帰や Cox 回帰モデルのパラメータ (疾病発生率の比) 推定に容易に拡張できる。

ネステッド・ケース・コントロール、ケース・コホート、第2段階ケース・コントロールといったデザインは、コホートのメンバーに関する曝露情報、交絡要因に関する情報がすでにデータベースに登録されている場合、簡単に実施することができる。また、血中の微量成分の測定が必要な場合、特に測定のコストが高ければ、血液 (または血清) を凍結保存し選ばれたケースとコントロールについてのみ解凍して測定を行えばよいので、コホート全体を測定する場合に比べてかなりのコストを節約できるだろう (Prentice [62]; Breslow and Day [8])。実際に、16,000 人のフォローアップを実施している ARIC [92] ではコストを考慮して解析にはネステッド・ケース・コントロールデザインを用いているし、デンマークの病院外来登録を用いて女性を対象に仕事のストレスと妊娠への影響を調べたケース・コホート研究も実施されている (Brandt and Nielsen [2])。

6. 疫学研究における因果推論

疫学研究の多くが「ある要因に曝露されることが疾病発生の原因であるかどうか、という因果関係に関する研究」であるため、因果関係の判定をめぐる古くからさまざまな議論がなされている。本節では最初に疫学研究における因果関係の考え方についての非統計的な議論を簡単に述べ、ついで因果関係に関するモデルとその適用例として交絡の定義について述べる。

6.1. 疫学と因果推論

因果関係についての疫学的な議論は誤解されている部分が多い。その代表的なものが「因果関係判定のための条件」として紹介されている様々な判断基準である。例えば Surgeon General の5条件 (U. S. Department of Health, Education and Welfare [95])、Hill の9条件 (Rothman [76])、Evans の12条件 (Last [46]) などが有名である。「これらの判定条件を満たしているのが因果関係が証明された」と結論する論文を見かけるが、明らかに観察研究から因果関係を証明することはできない。仮説的な因果関係を示唆する結果をたくさん集めれば、研究者が仮説的な因果関係を真実であると信じる割合は増すかもしれない。しかし因果推論で重要なことは、個人が仮説を信じるかどうかではなく、仮説が真実であるかどうかである。

因果関係を証明することはできないが、もし仮説的な因果関係から演繹的に導かれる予測を

否定するような観察が行われれば、仮説が誤っていることを示すことができる。因果関係の仮説に対して批判的な反証 (refutation) を繰り返し試みることによって、それでも否定しきれない仮説は真実である可能性が高い (corroborable) と考える Popper の立場から科学的な疫学研究を行おうという試みがなされている (Buck [9]; Maclure [52])。しかし、疫学は科学の一分野であると同時に公衆衛生活動を伴う実践的な学問である。したがって、因果関係を証明することはできなくとも、公衆衛生対策をとるための十分な証拠が集まった時点で迅速な対応をする必要がある。「因果関係判定のための条件」として紹介されている判断基準は、決して「それを満たせば因果関係があると判断してよい条件」ではなく、正しくは「その一部でも満たせば公衆衛生対策を考えなければならない条件」である。実践科学としての疫学の立場からは、観察研究では因果関係を証明できないことを理由に公衆衛生対策が遅れるようなことがあってはならない (Susser [88])。

Popperian の主張する科学としての疫学と実践家の立場からの公衆衛生としての疫学との間の因果推論についての議論は Rothman [77] を参照してほしい。

6.2. Counterfactual モデル

これまでの節と同様に曝露 E 、疾病 D の発生ともにあり、なし $(1, 0)$ に2分類されているとしよう。さらに、曝露は人為的に操作可能なものとする(性、年齢など操作可能でないものは個人の特性であり曝露とは考えない)。因果関係を考えるモデルとして次のような決定論的なモデルを考えよう。ある個人 i について曝露状況 E に応じて疾病を発生するかどうかを、

$$\begin{aligned} D(i, E) &= 1: \text{曝露状況が } E \text{ のとき疾病を発生する,} \\ &= 0: \text{曝露状況が } E \text{ のとき疾病を発生しない,} \end{aligned}$$

で表す。決定論的には、個人 i について a) 曝露を受けたか受けないかにかかわらず疾病を発生する、b) 曝露を受けたときは疾病を発生するが、曝露を受けないときは疾病を発生しない、c) 曝露を受けたとき疾病を発生しないが、曝露を受けないときは疾病を発生する、d) 曝露を受けたか受けないかにかかわらず疾病を発生しない、の4つの状態が存在する(ただし、我々は個人 i が a-d のどの状態であるかを知らない)。個人レベルの因果関係は、

$$D(i, E=1) - D(i, E=0),$$

で表すことができ、個人 i に限ってはこの値が 0 ならば曝露の効果なし、1 ならば因果的な効果あり、-1 ならば予防的な効果あり、と判断できる。

集団レベルの因果推論に関しては、集団内の $D(i, E)$ の割合の差である因果リスク差 (causal risk difference),

$$CRD = P[D(i, E=1)=1] - P[D(i, E=0)=1],$$

を用いることができる。因果リスク差は「集団全体が曝露を受けた場合の疾病発生割合と、同じ集団全体が曝露を受けなかった場合の疾病発生割合の差」である。因果リスク差を因果パラメータとして考えることは、ランダム化臨床試験で正当化される。一方、観察研究では集団全体についてではなく曝露グループに限った、曝露グループの因果リスク差 (causal risk difference in the exposed),

$$CRD_E = P[D(i, E=1)=1|E=1] - P[D(i, E=0)=1|E=1],$$

つまり「曝露グループの疾病発生割合と、曝露グループがもし曝露を受けていなかったら観察されたであろう疾病発生割合の差」を曝露効果の指標として用いる。

明らかに、CRD の場合、 $E=1$ の個人では $E=0$ のときの疾病発生状況、 $E=0$ の個人では $E=1$ のときの疾病発生状況を観察することはできない。また CRD_E では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 $P[D(i, E=0)=1|E=1]$ を観察することはできない。このモデルは観察できない仮想的な状況にもとづいているので counterfactual モデルと呼ばれ、いくつかの異なった観点から提案されている。しかし、counterfactual モデルの基本的な考え方はランダム化と関連して Fisher, Neyman にまでさかのぼることができる (Rubin [78, 79]; Copas [11]; Holland [40])。もし曝露の割り付けがランダムに行われていれば、我々は CRD や CRD_E のバイアスのない推定値を得ることができる。

6.3. 交絡の定義

3 節、4 節では定義ぬきで交絡を導入したが、ここでは counterfactual モデルを使って交絡を定義しよう。交絡の定義として一般的に用いられてきたのは「(因果連鎖の中間にない) 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ならば交絡あり」とする効果の指標の変化にもとづく定義 (change in estimate criterion) である。この定義では効果の指標として疾病発生割合の差を用いるか、疾病発生割合の比を用いるか、あるいはオッズ比を用いるかで、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という欠点がある。

Counterfactual モデルでは同一個人で $E=1$ と $E=0$ の疾病発生が観察できることを要求しているが、現実の世界ではどちらかの結果しか観察できない。我々は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 $P[D(i, E=0)=1|E=1]$ の代わりに非曝露グループの疾病発生割合 $P[D(j, E=0)=1|E=0]$ を観察し、疾病発生割合の差、

$$IPD = P[D(i, E=1)=1|E=1] - P[D(j, E=0)=1|E=0],$$

により曝露の効果を推定する。曝露グループの理想のコントロールグループである「曝露グループが曝露を受けなかった場合」と現実のコントロールグループである「非曝露グループ」の疾病発生割合が異なる場合、IPD は曝露の効果としてバイアスの入ったものになってしまう。したがって、

$$P[D(i, E=0)=1|E=1] \neq P[D(j, E=0)=1|E=0],$$

である場合「交絡あり」と定義すると、効果の指標によらないと定義となり「コントロールグループの十分性にもとづく定義 (adequacy of the control group criterion)」と呼ばれる (Greenland and Robins [35]; Wickramaratne and Holford [105])。この定義は観察不可能な量を含んでいるが、経験的に導かれた交絡要因の必要条件 (Rothman [76])、観察されたデータだけから交絡がないことは示せない、など交絡の性質を演繹的に導くことができる。

Counterfactual モデルはランダム化の概念と密接に結びついており (Rubin [80]; Greenland [28])、ランダム化臨床試験で Fisher の正確な p 値を用いることの正当性、時間に依存した共変量や、曝露の結果であり疾病の前駆症状であるような中間変数の取り扱いなどに応用されている (Greenland [31]; Robins [64, 66])。また、決定論的な疾病発生モデルだけではなく、確率的な疾病発生モデルや生存時間解析などにも拡張され、観察研究で推定可能な因果パラメータに関する議論もなされている (Robins and Greenland [70]; Robins [67])。ランダム化の仮定にもとづかないでバイアスのない因果パラメータの推定を行うための方法もいくつか提案されている。例えば propensity score を用いる方法では、2 節や 3 節で述べた回帰モデルのように疾病発生リスクを曝露変数と交絡変数で直接モデル化する代わりに、交絡変数が与えられたときの曝露の期待値 $E(E=1|Z)$ である propensity score をデータから推定する。この propensity score が等しいサブグループ内では曝露はランダムに割り付けられているとみ

なせるので、曝露効果のバイアスのない推定が可能となる (Rosenbaum and Rubin [73])。回帰モデルと propensity score を組み合わせて、より効率の高い効果の推定を行う試みもなされている (Robins et al. [71])。

7. その他の話題

本節では、これまでの話題に付随した話題ではあるが、なおかつ疫学研究で重要な問題である「曝露効果推定に与える誤分類 (misclassification) あるいは measurement error の影響」, 「Ecological bias」, 「経時観察研究 (longitudinal studies)」について簡単な文献紹介を行う。

7.1 誤分類, measurement error の影響

現実の研究では曝露変数, 交絡変数, 疾病分類などを全て完全に測定することは不可能であり, 測定には誤分類ないし誤差を伴う。誤分類や measurement error が存在すると, もはや曝露効果を正しく推定することはできない。例えばもっとも単純な, 曝露変数の誤分類が疾病の誤分類には影響しない「独立」な誤分類で, かつ本当に曝露を受けた人が本当に曝露を受けなかった人よりも疾病の誤分類が起りやすい (または, その逆) ということがない「nondifferential」な誤分類を考えてみよう。曝露と疾病が2値変数であって, 独立で nondifferential な誤分類が起している場合, 効果の指標は必ず「効果がない」という方向に薄められてしまう。ただし, 2つ以上の変数や2値以上のカテゴリを含む変数になると独立で nondifferential な誤分類であってもその影響は単純ではなく, 特別な条件下では効果が強くなる方向にバイアスがかかることが起りうる (Dosemeci et al. [14])。Measurement error に関しても, 曝露変数の nondifferential measurement error は線形回帰モデルの場合と同様にロジスティックモデル, Cox 回帰などで「効果がない」という方向にバイアスがかかった効果の推定を与える。Byar and Gail [10] には measurement error に関するシンポジウムの記録が収められており, バイアスの方向や大きさ, differential な誤分類, measurement error, 対象者の一部に対し正確な測定または繰り返し測定を実施することでバイアスを補正する方法 (validation substudies) などが議論されている。

7.2. Ecological bias

疾病が流行しているかどうか調べるためには, 疾病の時間的, 地域的な集積性 (disease clustering) をチェックする必要がある (Tango [90]; Rothenberg et al. [75] は地域集積性に関するシンポジウムの記録)。例えば疾病の地域集積性が観察されたとき, 果たして疾病の発生にその地域の特性 (気温, 降水量など) や固有の生活習慣などが関係しているのかどうか, という疑問が起きてくる。その際, コストのかかるコホート研究やケース・コントロール研究などの個人レベルの疫学研究を計画する前に, 地域別の疾病発生率と各地域の特性を表す変数との相関を調べる, 地図にプロットして比較する, といった ecological studies が要因のスクリーニングとしてよく行われる。パーソナルコンピュータ, 特にグラフィックツールの発達にともなって, ecological studies は簡単に実施することができるようになった。しかし, 地域単位での関連を見る ecological studies では個人レベルの疫学研究とは異なったバイアスが入ることが分かってきた。Ecological studies では, 効果の修飾があると効果の推定にバイアスが入ってしまう, 2値の曝露に nondifferential な誤分類があると効果を強める方向にバイアスがかかる, といった個人レベルの疫学研究では起らない ecological bias が存在する (Greenland and Morgenstern [33]; Brenner et al. [3])。したがって, ecological studies の結果の解釈には十分な注意を払わなければならない (Rosenbaum and Rubin [74]; Greenland [30])。

7.3. 経時観察研究

コホート研究のメリットとして, くり返し測定ができることを挙げたが, くり返し測定する

変数が曝露変数（曝露が1時点ではなく生活習慣など長期にわたる場合やコホートに登録されてから曝露を受ける可能性がある場合）、交絡変数である場合には時間に依存した共変量 (time dependent covariates) として Cox 回帰などのモデルにとり込むことができる。また曝露と疾病の中間変数については counterfactual モデルを用いた解析方法が提案されている。これまでの例では疾病発生は非可逆的な事象と考えていたが、ぜん息発作のようにくり返しのある事象や疾病発生の surrogate measure として中間変数を用いる場合、結果変数もくり返し測定できる場合がある。疾病発生（またはその surrogate）に関してデータをくり返し測定する研究を経時観察研究と呼び、疾病発生が個人内では独立ではないので特別な扱いを必要とする。Selwyn [86] には経時観察研究からのデータ解析に関するシンポジウムの記録がおさめられており、Liang-Zeger タイプの推定方程式 (Liang et al. [49]) を初めとする最近の研究成果が報告されている。また Ashby et al. [1] には継続調査だけでなく、cluster sampling, familial correlation といった、相関のあるデータ解析全般に関する文献がまとめられている。

8. 終わりに

本論文では疫学研究で用いられる生物統計手法のレビューを行った。今後の方向として第1に挙げられるのは、従来のコホート研究、ケース・コントロール研究の長所を取り入れた、より効率の高い研究デザイン（5節）を現実の大規模コホート研究に応用することであろう。理論的にはいかに優れた研究デザインであっても、実際の疫学研究でその実施可能性をチェックし、有効であることを示さなければ価値がない。この点に関しては新しいデザインを用いた疫学研究がいくつか動き出しているので、改良すべき点などが次第に明らかになると考えている。第2には、疫学的観察研究から因果推論をどのように行うか、またそれをサポートする統計手法としてどのようなものがあるかを整理し、やはり実際の研究に活かしていくことが挙げられる。こちらはまだ理論的な枠組み (counterfactual モデル) が明らかになった段階であり、もう少し問題の整理に時間がかかるかもしれない。

少し技術的な問題としては、疾病発生割合の差や比、疾病発生率の差に関する回帰モデルの開発が、特に sparse-data について、望まれる。ロジスティックモデル、Cox 回帰といった疫学研究で比較よく使われるモデルに関しては、重要な交絡要因を見落としてしまった場合の影響、measurement error の影響、個人差の導入 (変量効果モデル) など、より現実のデータ構造を反映した展開がなされている。しかし、疾病発生割合の差や比、疾病発生率の差では回帰モデル自体の研究がほとんどなされていない。一方で、仮定したモデルが現実のデータと合わない (model misspecification) 場合の影響なども研究されているが、ロジスティックモデルや Cox 回帰に代わるモデルとして疾病発生割合の差や比、疾病発生率の差をモデル化することは自然なので、これらの回帰モデルにもとづいたパラメータの推定方法の開発が必要である。

理論的な研究成果が応用に、しかも健康という我々にとって最も重要な問題の解決に直接結びつくところが、疫学-生物統計学のすばらしさだと筆者は信じている。

謝辞：本論文を読んで有益なコメントをいただいた本田純久先生、柳本武美先生、山本精一郎先生、レフェリーの方々に深く感謝します。

なおこの研究の一部は、統計数理研究所共同研究（4-共研-74）の補助を受けた。

参 考 文 献

- [1] Ashby, M, Neuhaus, J. M., Hauck, W. W., Bacchetti, P., Heibron, D. C., Jewell, N. P., Segal, M. R., and Fusaro, R. E. (1992). An annotated bibliography of methods for analysing correlated categori-

- cal data, *Statistics in Medicine*, **11**, 67-99.
- [2] Brandt, L. P. A. and Nielsen, C. V. (1992). Job stress and adverse outcome of pregnancy : A causal link or recall bias? *American Journal of Epidemiology*, **135**, 302-311.
- [3] Brenner, H., Savitz, D. A., Jockel, K.-H., and Greenland, S. (1992). Effects of nondifferential exposure misclassification in ecologic studies, *American Journal of Epidemiology*, **135**, 85-95.
- [4] Breslow, N. E. (1981). Odds ratio estimators when the data are sparse, *Biometrika*, **68**, 73-84.
- [5] Breslow, N. E. (1991). Statistical design and analysis of epidemiologic studies : Some directions of current research, *Handbook of Statistics, Vol. 8* (C. R. Rao and R. Chakraborty, eds.), Elsevier Science Publishers, 125-144.
- [6] Breslow, N. E. and Cain, K. C. (1988). Logistic regression for two-stage case-control data, *Biometrika*, **75**, 11-20.
- [7] Breslow, N. E. and Day, N. E. (1980). *Statistical Methods in Cancer Research, Vol-I. The Analysis of Case-Control Studies*, Oxford University Press.
- [8] Breslow, N. E. and Day, N. E. (1987). *Statistical Methods in Cancer Research, Vol-II. The Design and Analysis of Cohort Studies*, Oxford University Press.
- [9] Buck, C. (1975). Popper's philosophy for epidemiologists, *International Journal of Epidemiology*, **4**, 159-168.
- [10] Byar, D. P. and Gail, M. H., eds. (1989). Introduction to workshop on errors-in-variables, October 7-9, 1987, *Statistics in Medicine*, **8**, 1027-1029.
- [11] Copas, J. B. (1973). Randomization models for the matched and unmatched 2×2 tables, *Biometrika*, **60**, 467-476.
- [12] Cornfield, J. (1951). A method of estimating comparative rates from clinical data, *Journal of the National Cancer Institute*, **11**, 1269-1275.
- [13] Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B*, **49**, 1-39.
- [14] Dosemeci, M., Wacholder, S., and Lubin, J. H. (1990). Does nondifferential misclassification of exposure always bias a true effect towards the null value? *American Journal of Epidemiology*, **132**, 746-748.
- [15] Elandt-Johnson, R. C. (1975). Definition of rates : Some remarks on their use and misuse, *American Journal of Epidemiology*, **102**, 267-271.
- [16] Eysenck, H. J. (1991). Were we really wrong? *American Journal of Epidemiology*, **133**, 429-433.
- [17] Flanders, W. D., DerSimonian, R., and Rhodes, P. (1990). Estimation of risk ratios in case-base studies with competing risks, *Statistics in Medicine*, **9**, 423-435.
- [18] Flanders, W. D. and Greenland, S. (1991). Analytic methods for two-stage case-control studies and other stratified designs, *Statistics in Medicine*, **10**, 739-747.
- [19] Gail, M. H. (1991). A bibliography and comments on the use of statistical models in epidemiology in the 1980s, *Statistics in Medicine*, **10**, 1819-1885.
- [20] Gart, J. J. (1971). The comparison of proportions : A review of significance tests, confidence intervals, and adjustments for stratification, *Review of the International Statistical Institute*, **39**, 148-169.
- [21] Gart, J. J. and Nam, J. (1988). Approximate interval estimation of the ratio of binomial parameters : A review and corrections for skewness, *Biometrics*, **44**, 323-338.
- [22] Gart, J. J. and Nam, J. (1990). Approximate interval estimation of the difference in binomial parameters : Correction for skewness and extension to multiple tables, *Biometrics*, **46**, 637-643.
- [23] Greenland, S. (1982). Interpretation and estimation of summary ratios under heterogeneity, *Statistics in Medicine*, **1**, 217-227.
- [24] Greenland, S. (1983). Tests for interaction in epidemiologic studies : A review and a study of power, *Statistics in Medicine*, **2**, 243-251.
- [25] Greenland, S. (1986). Adjustment of risk ratios in case-base studies (hybrid epidemiologic designs). *Statistics in Medicine*, **5**, 579-584.
- [26] Greenland, S. (1987). Interpretation and choice of effect measures in epidemiologic analysis, *American Journal of Epidemiology*, **125**, 761-768.
- [27] Greenland, S. (1990). Re : "Those who were wrong", *American Journal of Epidemiology*, **132**, 585

-586.

- [28] Greenland, S. (1990). Randomization, statistics, and causal inference, *Epidemiology*, **1**, 421-429.
- [29] Greenland, S. (1991). Estimating standardized parameters from generalized linear models, *Statistics in Medicine*, **10**, 1069-1074.
- [30] Greenland, (1991). Divergent biases in ecologic and individual-level studies, *Statistics in Medicine*, **11**, 1209-1223.
- [31] Greenland, S. (1991). On the logical justification of conditional tests for two-by-two contingency tables. *The American Statistician*, **45**, 248-251.
- [32] Greenland, S. and Holland, P. (1991). Estimating standardized risk differences from odds ratio, *Biometrics*, **47**, 319-322.
- [33] Greenland, S. and Morgenstern, H. (1989). Ecological bias, confounding, and effect modification, *International Journal of Epidemiology*, **18**, 269-274.
- [34] Greenland, S. and Robins, J. M. (1985). Estimation of a common effect parameter from sparse follow-up data. *Biometrics*, **41**, 55-68.
- [35] Greenland, S. and Robins, J. M. (1986). Identifiability, exchangeability, and epidemiologic confounding. *International Journal of Epidemiology*, **15**, 413-419.
- [36] Greenland, S. and Thomas, D. C. (1982). On the need for the rare disease assumption in case-control studies, *American Journal of Epidemiology*, **116**, 547-553.
- [37] Herbst, A. L., Ulfelder, H., and Poskanzer, D.C. (1971). Adenocarcinoma of the vagina: Association of maternal stilbestrol therapy with tumor appearance in young women, *New England Journal of Medicine*, **284**, 878-881.
- [38] Hirji, K. F. (1991). A comparison of exact, mid-p, and score tests for matched case-control studies, *Biometrics*, **47**, 487-496.
- [39] 広津千尋 (1993). 臨床試験の統計学的側面, 本号所収.
- [40] Holland, P. W. (1986). Statistics and causal inference (with discussion), *Journal of the American Statistical Association*, **81**, 945-970.
- [41] Jones, M. P., O'Gorman, T. W., Lemke, J. H., and Woolson, R. F. (1989). A Monte Carlo investigation of homogeneity tests of the odds ratio under various sample size configurations, *Biometrics*, **45**, 171-181.
- [42] Kelsey, J. L., Thompson, W. D., and Evans, A. S. (1986). *Methods in Observational Epidemiology*. Oxford University Press.
- [43] Kleinbaum, D. G., Kupper, L. L., and Morgenstern, H. (1982). *Epidemiologic Research*, Van Nostrand Reinhold.
- [44] Kupper, L. L., McMichael, A. J., and Spirtas, R. (1975). A hybrid epidemiologic study design useful in estimating relative risk, *Journal of the American Statistical Association*, **70**, 524-528.
- [45] Langholz, B. and Thomas, D. C. (1991). Efficiency of cohort sampling designs: Some surprising results, *Biometrics*, **47**, 1563-1571.
- [46] Last, J. M., ed. (1988). *A Dictionary of Epidemiology* (2nd Ed.). Oxford University Press.
- [47] Liang, K.-Y. (1987). Extended Mantel-Haenszel estimating procedure for multivariate logistic regression models, *Biometrics*, **43**, 289-299.
- [48] Liang, K.-Y. and Self, S. G. (1985). Tests for homogeneity of odds ratio when the data are sparse, *Biometrika*, **72**, 353-358.
- [49] Liang, K.-Y., Zeger, S. L., and Qaqish, B. (1992). Multivariate regression analysis for categorical data (with discussion), *Journal of the Royal Statistical Society, Series B*, **54**, 3-40.
- [50] Liddell, F. D. K., McDonald, J. C., and Thomas, D. C. (1977). Methods for cohort analysis: Appraisal by application to asbestos mining, *Journal of the Royal Statistical Society, Series A*, **140**, 469-490.
- [51] Lubin, J. H. and Gail, M. H. (1984). Biased selection of controls for case-control analysis of cohort studies, *Biometrics*, **40**, 63-75.
- [52] Maclure, M. (1985). Popperian refutation in epidemiology, *American Journal of Epidemiology*, **121**, 343-350.
- [53] Mantel, N. (1973). Synthetic retrospective studies and related topics, *Biometrics*, **29**, 479-486.
- [54] Mantel, N. and Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective

- studies of disease, *Journal of the National Cancer Institute*, **22**, 719-748.
- [55] Mehta, C. R., Patel, N. R., and Gray, R. (1985). Computing an exact confidence interval for the common odds ratio in several 2×2 contingency tables, *Journal of the American Statistical Association*, **80**, 969-973.
- [56] Mehta, C. R. and Walsh, S. J. (1992). Comparison of exact, mid-p, and Mantel-Haenszel confidence intervals for the common odds ratio across several 2×2 contingency tables, *The American Statistician*, **46**, 146-150.
- [57] Miettinen, O. S. (1972). Components of the crude risk ratio, *American Journal of Epidemiology*, **96**, 168-172.
- [58] Miettinen, O. S. (1976). Estimability and estimation in case-referent studies, *American Journal of Epidemiology*, **103**, 226-253.
- [59] Miettinen, O. S. (1982). Design options in epidemiologic research: An update, *Scandinavian Journal of Work and Environmental Health*, **8** (Supplement 1), 7-14.
- [60] 丸井英二 (1991). 疫学の歴史, 疫学 (豊川裕之編), メヂカルフレンド社, 7-104.
- [61] National Heart, Lung, and Blood Institute (1987). *The Framingham Study: an epidemiological study of cardiovascular disease*, U. S. Department of Health and Human Services (NIH Publication no. 87-2703).
- [62] Prentice, R. L. (1986). A case-cohort design for epidemiologic cohort studies and disease prevention trials, *Biometrika*, **73**, 1-11.
- [63] Prentice, R. L. (1986). On the design of synthetic case-control studies, *Biometrics*, **42**, 301-310.
- [64] Robins, J. M. (1987). A graphical approach to the identification and estimation of causal parameters in mortality studies with sustained exposure periods, *Journal of Chronic Diseases*, **40** (Supplement 2), 139S-161S.
- [65] Robins, J. M. (1988). Confidence intervals for causal parameters, *Statistics in Medicine*, **7**, 773-785.
- [66] Robins, J. M. (1989). The control of confounding by intermediate variables, *Statistics in Medicine*, **8**, 679-691.
- [67] Robins, J. M. (1992). Estimation of the time-dependent accelerated failure time model in the presence of confounding factors, *Biometrika*, **79**, 321-334.
- [68] Robins, J. M., Breslow, N. E., and Greenland, S. (1986). Estimators of the Mantel-Haenszel variance consistent in both sparse data and large-strata limiting models, *Biometrics*, **42**, 311-323.
- [69] Robins, J. M., Gail, M. H., and Lubin, J.H. (1986). More on "Biased selection of controls for case-control analysis of cohort studies", *Biometrics*, **42**, 292-299.
- [70] Robins, J. M. and Greenland, S. (1989). The probability of causation under a stochastic model for individual risk, *Biometrics*, **45**, 1125-1138.
- [71] Robins, J. M., Mark, S. D., and Newey, W. K. (1992). Estimating exposure effects by modelling the expectation of exposure conditional on confounders, *Biometrics*, **48**, 479-495.
- [72] Robins, J. M., Prentice, R. L., and Blevins, D. (1989). Designs for synthetic case-control studies in open cohorts, *Biometrics*, **45**, 1103-1116.
- [73] Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects, *Biometrika*, **70**, 41-55.
- [74] Rosenbaum, P. R. and Rubin, D. B. (1984). Difficulties with regression analyses of age-adjusted rates, *Biometrics*, **40**, 437-443.
- [75] Rothenberg, R. B., Steinberg, K. K., and Thacker, S. B. (1990). Introduction: The public health importance of clusters: A note from the Centers for Disease Control, *American Journal of Epidemiology*, **132**, Supplement, S3-S5.
- [76] Rothman, K. J. (1986). *Modern Epidemiology*, Little, Brown and Company.
- [77] Rothman, K. J., ed. (1988). *Causal Inference*, Epidemiology Resources Inc.
- [78] Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization, *Annals of Statistics*, **6**, 34-58.
- [79] Rubin, D. B. (1990). Neyman (1923) and causal inference in experiments and observational studies, *Statistical Science*, **5**, 472-480.
- [80] Rubin, D. B. (1991). Practical implications of modes of statistical inference for causal effects and the critical role of the assignment mechanism, *Biometrics*, **47**, 1213-1234.

- [81] Sato, T. (1990). Confidence limits for the common odds ratio based on the asymptotic distribution of the Mantel-Haenszel estimator, *Biometrics*, **46**, 71-80.
- [82] Sato, T. (1990). Confidence intervals for effect parameters common in cancer epidemiology, *Environmental Health Perspectives*, **87**, 95-101.
- [83] Sato, T. (1992). On inconsistency of the common rate difference estimators from sparse follow-up data, *Annals of the Institute of Statistical Mathematics*, **44**, 529-535.
- [84] Sato, T. (1992). Estimation of a common risk ratio in stratified case-cohort studies, *Statistics in Medicine*, **11**, 1599-1605.
- [85] Sato, T. (1992). Maximum likelihood estimation of the risk ratio in case-cohort studies, *Biometrics*, **48**, 1215-1221.
- [86] Selwyn, M. R. (1992). Introduction to Symposium, *Statistics in Medicine*, **11**, 1797.
- [87] Stolley, P. D. (1991). When genius errs: R. A. Fisher and the lung cancer controversy, *American Journal of Epidemiology*, **133**, 416-425.
- [88] Susser, M. (1991). What is a cause and how do we know one? A grammar for pragmatic epidemiology, *American Journal of Epidemiology*, **133**, 635-648.
- [89] Takagi, H. (1991). Approximate and exact methods for computing the confidence limits for the common odds ratio in a series of 2×2 tables, *Bulletin of the Biometric Society of Japan*, **12**, 57-65.
- [90] Tango, T. (1984). The detection of disease clustering, *Biometrics*, **40**, 15-26.
- [91] Tarone, R. E. (1985). On homogeneity tests based on efficient scores, *Biometrika*, **72**, 91-95.
- [92] The ARIC Investigators. (1989). The atherosclerosis risk in communities (ARIC) study: Design and objectives, *American Journal of Epidemiology*, **129**, 687-702.
- [93] Thomas, D. C. (1981). General relative risk functions for survival time and matched case-control analysis, *Biometrics*, **37**, 673-686.
- [94] Thomas, D. C., Blettner, M., and Day, N. E. (1992). Use of external rates in nested case-control studies with application to the international radiation study of cervical cancer patients, *Biometrics*, **48**, 781-794.
- [95] U. S. Department of Health, Education and Welfare. (1964). *Smoking and Health*, Public Health Service Publication No. 1103.
- [96] Vandenbroucke, J. P. (1989). Those who were wrong, *American Journal of Epidemiology*, **130**, 3-5.
- [97] Vollset, S. E., Hirji, K. F. and Elashoff, R. M. (1991). Fast computation of exact confidence limits for the common odds ratio in a series of 2×2 tables, *Journal of the American Statistical Association*, **86**, 404-409.
- [98] Wacholder, S. (1985). Binomial regression in GLIM: Estimating risk ratios and risk differences, *American Journal of Epidemiology*, **123**, 174-184.
- [99] Wacholder, S., McLaughlin, J. K., Silverman, D. T., and Mandel, J. S. (1992). Selection of controls in case-control studies, I. Principles, *American Journal of Epidemiology*, **135**, 1019-1028.
- [100] Wacholder, S., Silverman, D. T., McLaughlin, J. K., and Mandel, J. S. (1992). Selection of controls in case-control studies, II. Types of controls, *American Journal of Epidemiology*, **135**, 1029-1041.
- [101] Wacholder, S., Silverman, D. T., McLaughlin, J. K., and Mandel, J. S. (1992). Selection of controls in case-control studies, III. Design options, *American Journal of Epidemiology*, **135**, 1042-1050.
- [102] Walker, A. M. (1991). *Observation and Inference*, Epidemiology Resources Inc.
- [103] Walter, S. D. and Cook, R. J. (1991). A comparison of several point estimators of the odds ratio in a single 2×2 contingency table, *Biometrics*, **47**, 795-811.
- [104] White, J. E. (1982). A two stage design for the study of the relationship between a rare exposure and a rare disease, *American Journal of Epidemiology*, **115**, 119-128.
- [105] Wickramaratne, P. J. and Holford, T. R. (1987). Confounding in epidemiologic studies: The adequacy of control group as a measure of confounding, *Biometrics*, **43**, 751-765.
- [106] Yanagawa, T. and Fujii, Y. (1990). Homogeneity test with a generalized Mantel-Haenszel estimator for $L \times K$ contingency tables, *Journal of the American Statistical Association*, **85**, 744-748.
- [107] Zhao, L. P. and Lipsitz, S. (1992). Designs and analysis of two-stage studies, *Statistics in Medicine*, **11**, 769-782.