原　　　　　著

医学における統計学の応用について

（第７編）疫学研究の指標と評価方法

新潟大学医学部衛生学教室（主任：山本正治教授）
遠藤　和男・山崎　　理

An Application of Statistical Methods to Medical Science (Part Ⅶ) The Indeces of Epidemiologic Studies and The Methods for Evaluation

Kazuo ENDOH and Osamu YAMAZAKI

Department of Hygiene and Preventive Medicine
Niigata University School of Medicine
(Director: Prof. Masaharu YAMAMOTO)

The causes of a disease have not been clarified sufficiently. For example, there may be no one who denies the relationship between tabacco smoking and lung cancer. Besides no one can assert that the former should be only one cause of the latter, because genetic factors may be involved.

The factors which should elevate the risk for the occurrence of a disease, are called as risk factors before they have been identified as the cause of it. In present paper, authors intend to give some interpretations to the methods of epidemiologic studies and the indeces for evaluation of risk factors by using some samples. Both of them may help young researchers to realize the basis of Evidence-based Medicine.

Key words: Relative Risk, Odds Ratio, case-control study, metched pair, logistic model
相対危険度，オッズ比，患者－対照研究，マッチド・ペア，ロジスティック・モデル

Reprint requests to: Kazuo ENDOH, Department of Hygiene & Preventive Medicine, Niigata University School of Medicine, Niigata City, 951-8510, JAPAN.

別刷請求先：　〒951-8510　新潟市旭町通１番町
新潟大学医学部衛生学教室　　　　遠藤和男

は じ め に

疾病の原因は，必ずしも明らかにされてきたわけではない．例えば，喫煙と肺がんの関連性[1]を否定する人はいないし，また喫煙が唯一の原因であるとも言い切れない．遺伝も関与している可能性があろう．原因として確定されたわけではないものの，疾病にかかる危険性を高めると考えられる因子を危険因子（risk factor）と呼ぶ．本編では，疫学調査の手法と危険因子の評価方法について，実例をあげて解説を加えた．若手研究者がEvidence－based Medicine の基礎を理解する上で，何らかの参考となれば幸いである．

1. 横断研究（cross-sectional study）

断面調査とも言い，最も典型的な例としてアンケート調査[2]があげられる．**表1**に示したように，調査の結果は2×2分割表[3]にまとめられる．属性 B の有無が疾病や症状の有無になる場合が多い．統計学的にはカイ二乗検定によって評価される．属性 A の有無によって疾病や症状を有する割合（母数 n が集団を代表する場合にのみ有病率または有症率といってよい）を検討する場合には，表中の a～d の記号を用いて $\frac{a}{a+b}$ と $\frac{c}{c+d}$ とを比較することになる．ただし，横断研究では N＝a＋b＋c＋d を分母として，4つの比 $\frac{a}{N}$, $\frac{b}{N}$, $\frac{c}{N}$, $\frac{d}{N}$ が平等であるかどうかを見ているに過ぎない[4]．すなわち，横に見て a：b＝c：d または縦に見て a：c＝b：d が成立するかどうかをみている．数学的な条件としてはいずれも ad＝bc となる．この場合カイ二乗式において

$$\chi_0{}^2=\frac{(ad-bc)^2\cdot N}{(a+b)(c+d)(a+c)(b+d)}=0 \quad \cdots\cdots \mathrm{I})$$

が成立することになる．

表1　横断研究での2×2表

		属性 B		計
		あ り	な し	
属性A	あ り	a	b	a＋b
	な し	c	d	c＋d
計		a＋c	b＋d	N

N＝a＋b＋c＋d に対するそれぞれの比

$\frac{a}{N}$, $\frac{b}{N}$, $\frac{c}{N}$, $\frac{d}{N}$ を斜めに見ている．

a：b＝c：d または a：c＝b：d

2. コホート研究（cohort study，表2）

コホートとは前進するローマ軍の1単位に由来する．コホート研究では，危険因子の曝露の有無によって最初に2群に分けて，これから疾病が発生するかどうかを追跡する．したがって，時間的な経過からみて前向き研究（prospective study）とも呼ばれる．**表1**での属性 A が危険因子への曝露，属性 B は疾病の発生となる．発生率は曝露群では $\frac{a}{a+b}$，非曝露群では $\frac{c}{c+d}$ となる．両者の比較では（1）前者が後者の何倍にあたるか

表2　追跡研究での2つの見方

患者－対照研究

			疾病		計	
			あ り	な し		
コホート研究	曝露	あ り	a	b	a ＋ b	横に見る ⇨
		な し	c	d	c ＋ d	
	計		a ＋ c	b ＋ d	N	

相対危険度 RR

$$=\frac{\frac{a}{a+b}}{\frac{c}{c+d}}$$

縦に ⇩ 見る

$\frac{a}{a+c}$ と $\frac{b}{b+d}$ との比較，オッズ比 OR $=\frac{ad}{bc}$

表3　喫煙習慣別肺がん死亡率（千人対）

年齢階級	非喫煙	軽　度	中等度	大量
35 ～ 54	0.00	0.09	0.17	0.26
55 ～ 64	0.00	0.32	0.92	3.10
65 ～ 74	0.00	1.35	3.34	4.80
75歳以上	1.17	2.78	2.70	4.16
総　数	0.07	0.47	0.86	1.66

軽度：1～14，中等度：15～24，大量：25～mg/日[5)]

と，(2) 前者は後者よりどれだけ多いかの2とおりが考えられる．(1) 曝露群の疾病発生率が非曝露群の何倍にあたるかをみた指標を相対危険度（RR：Relative Risk）と言い，

$$RR=\frac{a}{a+b}\Big/\frac{c}{c+d} \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \text{II)}$$

RR＝1の時分母を払って，a (c+d) ＝c (a+b) から，I）と同様に ad＝bc が成立するため，RR の有意性もカイ二乗検定[3)]による．

(2) の指標を寄与危険度（AR：Attributable Risk）と言い，曝露群と非曝露群との差，すなわち曝露によって過剰になった発生率を示す．さらにこれを曝露群の発生率で割ったものを寄与（危険）割合（ARP：AR Percent）と呼び，疾病発生に危険因子が寄与する割合，逆に言えば曝露を避ければ，疾病を予防できる割合を示している．

$$ARP=1-\frac{c}{c+d}\Big/\frac{a}{a+b}=1-\frac{1}{RR} \quad \cdots\cdots\cdots \text{III)}$$

【例1】喫煙と肺がん発生と関連

表3は喫煙と肺がん死亡との関係を初めて客観的に指摘した Doll & Hill[5)] のデータである．軽度，中等度及び大量喫煙者の相対危険度はそれぞれ，6.7，12.3及び23.7と，喫煙量が多いほど危険度も高い（量－反応関係）．また，大量喫煙者の寄与（危険）割合は95.8％となり，喫煙以外の原因の関与は極めて少ないと考えられる．

3.　患者一対照研究［case-control (referent) study］

症例－対照研究とも呼ばれる．調査時点で患者と対照となる健常人の2群に分けて，過去における危険因子への曝露の有無を調査する．時間的な向きがコホート研究とは逆となり，後ろ向き研究（retrospective study）に分類される（図1参照）．

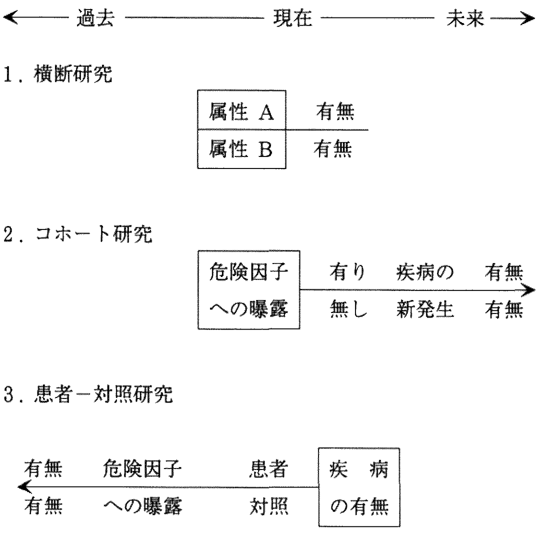

図1　疫学研究における時間の流れ

患者群，対照群における曝露ありの割合は，それぞれ $\frac{a}{a+c}$，$\frac{b}{b+d}$ から算出されるため，コホート研究とは異なり，**表2**を縦⇩方向に見る必要がある．

ところで，患者及び対照が地域母集団から抽出された確率をそれぞれ $\frac{1}{k}$，$\frac{1}{\ell}$ と仮定すると，**表2**での患者群 a，c 及び対照群 b，d をそれぞれ k，ℓ 倍すれば，地域母集団における過去の曝露の有無と，現在の疾患の有無との関係を**表4**のように推定できる．曝露群，非曝露群からの発生率はそれぞれ $\frac{ka}{ka+\ell b}$，$\frac{kc}{kc+\ell d}$ となる．

高血圧や糖尿病など頻度の高い疾病は別として，地域母集団における患者の数は，対照者数に比べて非常に小さいと考えられ，ka＜＜ℓb，kc＜＜ℓd であるために前者をそれぞれ無視した場合，前述の発生率は，$\frac{ka}{\ell d}$，$\frac{kc}{\ell d}$ に近似できる．両者の比は $\frac{a}{b}\Big/\frac{c}{d}=\frac{ad}{bc}$ となる．コホート研究における相対危険度そのものではないものの，近似値としてオッズ比（OR：Odds Ratio）と呼ばれる．オッズとは，競馬の勝ち馬などの予想配当倍率をいう．また，$OR=\frac{ad}{bc}=1$ ……………………IV）の条件はI）式と同じであり，カイ二乗検定[3)]によっ

表 4　母集団中における予想頻度

本　来

			現在の疾病		
			あ り	な し	計
過去の露曝	有	あ り	k a	ℓ b	ka ＋ ℓb
	無	な し	k c	ℓ d	kc ＋ ℓd
計			k（a＋c）	ℓ（b＋d）	N

縦に ⇩ 見る

横に ⇨ 見る

$$[相対危険度]\ \frac{\dfrac{ka}{ka+\ell b}}{\dfrac{kc}{kc+\ell d}} \fallingdotseq \frac{\dfrac{ka}{\ell b}}{\dfrac{kc}{\ell d}} = [オッズ比]\ \frac{ad}{bc}$$

表 5　コホート及び患者－対照研究の比較

研究の型	利　点	欠　点
コホート研究（前向き調査）	①仮説の証明が容易である． ②バイアスの混入が少ない． ③相対危険度だけでなく，罹患率を計算できる． ④他疾患との関連の可能性も追求できる．	①多数の被験者が必要である． ②稀な疾患では追跡期間が長くなり，脱落が問題となる． ③調査費用がかさむ． ④診断基準が変わる恐れあり．
患者－対照研究（後向き調査）	①被験者数が少なくてよい． ②過去の曝露の有無についての調査であり，期間が短い． ③調査費用がかからない． ④稀な疾患にも対応できる．	①仮説の証明が困難である． ②種々のバイアス，特に記憶上のバイアスが混入し易い． ③オッズ比のみで，寄与危険度や罹患率は計算できない．

て評価される．

4.　コホート研究と患者－対照研究の比較

コホート研究では，疾患の発生まで追跡する必要があるため，長期にわたる観察が必要であるが，相対危険度による仮説の証明は容易である．一方，患者－対照研究では比較的容易に結果が得られるものの，過去の曝露について調査するため，さまざまなバイアス（bias）[1]が入り込みやすい．それぞれの利点及び欠点をまとめると，**表 5** のようになる．危険因子について検討する場合，まず患者－対照研究で可能性を確かめた上で，次にコホート研究を行って因果関係[1]について論じるのが理想であるが，前者だけで後者を伴わない研究も多い．

5.　95%信頼区間（CI：confidence interval）

A．相対危険度の95%信頼区間

相対危険度は罹患率の比として求められ，このまま評価されることが多い．ただし，自然対数を用いた Katz の方法[6]が知られており，次式から指数で戻せばよい．

$$\ln(\mathrm{RR}) \pm 1.96\sqrt{\frac{b/a}{a+b}+\frac{d/c}{c+d}}$$

B．オッズ比の95%信頼区間

一方，オッズ比は罹患率の比ではなく，近似値に過ぎないため，必ず95%信頼区間を求める．その下限が 1 より大きい場合には危険因子，その上限が 1 より小さい場合には防御因子，1 をはさむ場合には有意ではないと判断される．95%信頼区間を求める方法として，Woolf の方法[7]と Miettinen の test based method[8]に

表6　【例2】での2×2表

		患　者	対　照	計
因子	あ　り	30	20	50
	な　し	20	30	50
計		50	50	100

$$OR=\frac{30\cdot 30}{20\cdot 20}=2.25\ (N.S.)$$

$$\chi_0{}^2=4.00,\ \chi_Y{}^2=3.24$$

ついて，比較検討してみる．

【例2】オッズ比の95%信頼区間の算出

患者50例，対照50例について，ある因子の有無を調べたところ，**表6**のような結果を得た．

(1) オッズ比は $OR=\dfrac{30\times 30}{20\times 20}=2.25$ である．

(2) 対数オッズ比 ln（OR）の分散は

$$\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}$$ である．

(3) Woolf の方法[7]では，以下の数値を指数で戻してやる．

$$\ln(OR)\pm 1.96\sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}} \quad \cdots V)$$

$\sqrt{\dfrac{1}{a}+\dfrac{1}{b}+\dfrac{1}{c}+\dfrac{1}{d}}=0.408$ から，

95%信頼区間は，1.01～5.01 となる．

(4) ところで，カイ二乗検定で元式による値を $\chi_0{}^2$，イェーツの修正項－N/2を用いた値を $\chi_Y{}^2$ とすると，

$$\chi_Y{}^2=\frac{(|ad-bc|-N/2)^2\cdot N}{(a+b)(c+d)(a+c)(c+d)} \quad \cdots\cdots VI)$$

I），VI）式から $\chi^2{}_0=4.00$，$\chi^2{}_Y=3.24<3.841$ となり，後者によるとオッズ比は有意ではない．すなわち，（3）の信頼区間との間にズレが生じてしまう．

(5) Miettinen の test based method[8] では，検定に用いたカイ二乗の値をそのまま用いることができる．

$$\exp[(1\pm 1.96/\sqrt{\chi^2})\cdot \ln(OR)] \quad \cdots\cdots\cdots VII)$$

から求められる．カイ二乗の値として $\chi^2{}_0=4.00$ を用いると 1.02～4.98，$\chi^2{}_Y=3.24$ によると 0.93～5.44 となる．前者は（3）の区間に近く，後者では検定の結果と一致し，1をはさんでいるので有意ではない．

(6) Miettinen の方法が絶対的かというと，①2×2表の度数 a～d が小さいときには不正確である．②オッズ比が1から遠ざかるほど，本来の区間より狭くなる．という2点が指摘されており，①では Cornfield の方法[9]，②の場合には Gart & Thomas 法[10]を用いることが推奨されている．詳細については専門書[11]を参照されたい．

(7) a～d の出現度数のいずれかに0が認められた場合には，オッズ比及び95%信頼区間は計算不能となる．したがって，IV）～VI）式の a～d にすべて0.5を加えて計算すればよい[11]．

表7　マッチド・ペアの場合

患者＼対照		因子への曝露 +	因子への曝露 −
因子	＋	a	b
	－	c	d

a～d は患者－対照の組数

$$OR=\frac{b}{c},\ \chi^2{}_{MN}=\frac{(|b-c|-1)^2}{b+c}$$

6. マッチしたデータによる分析[1)11)]

患者群と対照群の性・年齢の構成が異なる場合，危険因子の判断を誤る恐れがあるため，予め患者と対照の性別や年齢階級（±3または5歳以内）をマッチさせて対照を選ぶ方法がある．マッチさせた項目については危険因子として検討できない．患者及び対照の数が多い場合にはそのままIV）～VII）式を適用できる．ただし，特に数が少ない場合には，患者と対照をペア（matched pair）とし，1組として評価した方が有利である．一度ペアを作った以上は，ペアを崩したり，組み換えたりできない．したがって，2×2表及びカイ二乗検定も特殊となる．

(1) **表7**に示したように，2×2表の表頭は患者について，表側は対照についての因子への曝露の有無となる．

(2) 表中 a は患者が＋で対照も＋，d は両者とも－－と符号が一致するため無視される．すなわち，患者のみ＋の頻度は対照のみ＋の頻度の何倍にあたるかを見る．

$$OR=\frac{b\ (患者+対照-)}{c\ (患者-対照+)},\ 期待値=\frac{b+c}{2} \quad \cdots\cdots VIII)$$

(3) カイ二乗の元式 $\Sigma\dfrac{(|観察値-期待値|-1/2)^2}{期待値}$ に代入すると，McNemar の検定式[12]が得られる．

表8　【例3】でのペアの表

		対照の因子		計
		＋	－	
患者の因子	＋	30	24	54
	－	11	35	46
計		41	59	100

$$OR=\frac{24}{11}=2.18\ [1.02\sim4.75]$$

Ⅸ）式から $\chi^2_{MN}=4.114\ (p<0.05)$

Ⅵ）式なら $\chi_Y^2=9.015\ (p<0.01)$

$$\chi^2_{MN}=\frac{(|b-c|-1)^2}{b+c} \quad \cdots\cdots Ⅸ)$$

(4) 95%信頼区間については，Ⅴ）式の a，d を無視して

$$\ln(OR)\pm1.96\sqrt{\frac{1}{b}+\frac{1}{c}} \quad \cdots\cdots Ⅹ)$$

となることが理解できるであろう．ただし，Fleiss[13] は $n=b+c$ とおいて，$p=\frac{b}{n}$ の下限 p_L 及び上限 p_U を計算し，

$$\frac{p}{1-p} \quad \cdots\cdots Ⅺ)$$

によって95%区間へと変換した．

正規分布で $z(0.05/2)=1.96$ を z，$q=1-p$ とおくと，

$$p_L=\frac{(2b+z^2-1)-z\sqrt{z^2-(2+1/n)+4p(nq+1)}}{2(n+z^2)}$$

$$P_L=\frac{(2b+z^2+1)-z\sqrt{z^2+(2-1/n)+4p(nq-1)}}{2(n+z^2)} \quad \cdots\cdots Ⅻ)$$

(5) なお，患者の例数が少ない割に対照の方が得られやすいため，1：m にマッチさせる場合や，さらに脱落が生じて，対照の数が異なる場合でもオッズ比を計算できる．詳しくは専門書[11]を参照されたい．

【例3】1対1のマッチド・ペアの例

拙著[3]及び第1編[14]に掲載した例の数値を多少変更して，**表8**に示した結果が得られたとする．

(1) オッズ比 $OR=24/11=2.18$

(2) $\chi^2_{MN}=(|24-11|-1)^2/35=4.114>3.841$

したがって，(1) の OR は有意 ($p<0.05$) である．

(3) Ⅹ）式による95%信頼区間は1.07～4.45である．

表9　【例3】でペアを崩した表

		患　者	対　照	計
因子	あ　り	54	41	95
	な　し	46	59	105
計		100	100	200

※網　をかけた数値は**表8**と一致する

$$OR=\frac{54\cdot59}{41\cdot46}=1.69\ (N.S.)$$

Ⅵ）式から $\chi_Y^2=2.887$

表10　【例4】でマッチさせた表

患者＼対照		作業歴の有無 ＋	作業歴の有無 －	計
作業歴	＋	a	b	30
	－	c	d	20
計		10	40	50

※網　をかけた数値は問題文による数値．

一方，上記の $p_L=0.506$，$p_U=0.826$ をⅪ）式によって変換すると，1.02～4.75となる．すなわち，前者の方が狭くなっていることがわかる．

(4) Ⅸ）式の代わりにⅥ）式を用いると，$\chi_Y^2=9.015$ から誤って $p<0.01$ と判断される（(2) では $p<0.05$）．

(5) **表8**からペアを崩すと患者＋は30＋24＝54人，患者－は11＋35＝46人，同様に対照＋は41人，対照－は59人となる．したがって，それぞれが**表3**の a～d に該当すると考えて，Ⅵ）式によって検定すると，$\chi_Y^2=2.887$ であるから，有意差なしと判断されてしまう（**表9**）．以上から，少数例ではペアを作った方が，有利であることが理解できるであろう．

【例4】第89回医師国家試験問題 A13[15] から

ある企業で，腰痛患者50人と年齢をマッチさせた50人とについて，重量物運搬作業の業務歴を調査した結果，患者では30人，対照者では50人であった．この作業の腰痛に対するオッズ（odds）比を求めなさい．

(1) マッチングしたというので，**表7**の形式に基づいて**表10**を作成してみると，b 及び c の数値を特定できない．業務歴ありが組数ではなく，それぞれの人数で

示されていることに気づくであろう．

(2) そこで，マッチさせない場合の**表11**を作成する．オッズ比は，ad/bc=30・40/10/20=6となる．検定及び信頼区間については省略する．

(3) このように1対1のペアを作成しない場合もある．年齢については，危険因子として検討できない代わりに，バイアスとして働いている可能性を否定できる．

7. 量－反応関係（dose-response relationship）[1]

例1では軽度，中等度及び大量喫煙者の相対危険度は6.7, 12.3, 23.7と上昇し，量－反応関係が認められた．マンテル－エクステンション（Mantel-extension）法[16]による検定について，実例にそって解説を加える．

【例5】胆嚢がんと赤色唐辛子の摂取

表12には，チリの女性胆嚢がんの患者－対照研究[17]のうち，赤色唐辛子の成績を示した．ほとんど食べないに対するオッズ比は，1日1回未満 2.49（1.11～5.55），一回以上 5.54（1.96～15.70）とそれぞれ有意であった．

表11　【例4】でマッチさせない表

		患者	対照	計
作業歴	あり	30	10	40
	なし	20	40	60
計		50	50	100

※網　をかけた数値は**表10**と一致する

$$\mathrm{OR}=\frac{30\cdot 40}{10\cdot 20}=6.00\ [2.43\sim 14.84]$$

$$\chi_Y{}^2=15.042\ (p<0.001)$$

(1) 得点 R_j は計 M_j と総数 N から計算する．

$$R_0=\frac{1}{N}\left(\frac{M_0+1}{2}\right)=\frac{1}{148}\left(\frac{77+1}{2}\right)=0.264$$

一般形は $R_k=\frac{1}{N}\left(M_0+\cdots+M_{k-1}+\frac{M_k+1}{2}\right)$

であり，$R_1=0.679$, $R_2=0.919$ となる．

(2) 患者数 a_j と得点 R_j とを乗じた総和を O とする．すなわち $O=\Sigma a_j\cdot R_j=43.186$

(3) 観察値 O に対する期待値 E と，(4) 分散 V を計算する．

$$E=\frac{n_1}{N}\Sigma M_j\cdot R_j=\frac{74}{148}\cdot 74.537=37.269$$

$$(4)\ V=\frac{n_1\cdot n_2}{N^2\ (N-1)}\left[N\Sigma M_j\cdot R_j{}^2-(\Sigma M_j\cdot R_j)^2\right]$$

$$=74\cdot 74\ (148\cdot 47.689-74.537^2)\ /148^2/147=2.555$$

(5) $\chi_0{}^2=\frac{(O-E)^2}{V}$ は自由度1のカイ二乗分布に従う．

$\chi_0{}^2=(43.186-37.269)^2/2.555=13.703>10.8276$

したがって，量－反応関係は有意（$p<0.001$）である．

8. ロジスティック・モデル（logistic model）[18]

XI）式の対数をとって，$z=\ln\frac{p}{1-p}$ に変換することをロジット（logistic unit, logit）変換と呼ぶ．さらに，$e^{-z}=\frac{1}{p}-1$ から $p=\frac{1}{1+e^{-z}}$ へと変形できる．z がxの一次回帰式として，$z=\alpha+\beta x$ と書き表せるなら，

$$y=\frac{1}{1+\exp\left[-(\alpha+\beta x)\right]}\quad\cdots\cdots\cdots\cdots\text{XII)}$$

をロジスティック関数，回帰式はモデル的なのでロジスティック・モデルと呼ぶ．もともとの関数 $y=\frac{1}{1+e^{-x}}$

表12　チリ女性における赤色唐辛子の摂取頻度[17]

摂取頻度	患者 a_j	対照	計 M_j	オッズ比 OR	(1) 得点 R_j*	(2) $a_j\cdot R_j$	(3) $M_j\cdot R_j$	(4) $M_j\cdot R_j{}^2$
ほとんど食べない	28	49	77	1.00	0.264	7.392	20.328	5.367
1日1回未満	27	19	46	2.49	0.679	18.333	31.234	21.208
1日1回以上	19	6	25	5.54	0.919	17.461	22.975	21.114
計	74	74	148	3.22	1.000	43.186	74.537	47.689
	n_1	n_2	N			O		

*$R_k=\frac{1}{N}\left(M_0+\cdots+M_{k-1}+\frac{M_k+1}{2}\right)$，(3) 期待値 E，(4) 分散 V の計算は本文を参照．

は，**図2**に示したように0～1の間の値をとり，正規分布の累積確率密度曲線である．x を時間軸 t とすれば，死亡率曲線となる．x を用量，y を百分率とすれば量－反応曲線であり，50%致死量（LD_{50}）が求められる．

また，因子が複数ある場合には重回帰モデルのように，

$z=\alpha+\beta_1 x_1+\beta_2 x_2+\cdots$ ……………ⅩⅣ)

をⅫ）式に入れ，多重ロジスティック・モデルと呼ぶ．

ところで，x が1つだけの場合，x＝0及び1とおくと $z_0=\alpha$，$z_1=\alpha+\beta$ から $\beta=z_1-z_0$ である．

$$\exp(\beta)=\frac{e^{z1}}{e^{z0}}=\frac{p_1/1-p_1}{p_0/1-p_0}=\frac{ad}{bc}$$ ………ⅩⅤ)

表2から $p_1=\dfrac{a}{a+b}$，$p_0=\dfrac{c}{c+d}$ と，それぞれ曝露群及び非曝露群の陽性率と考えられるため，exp（β）はオッズ比そのものとなっている．

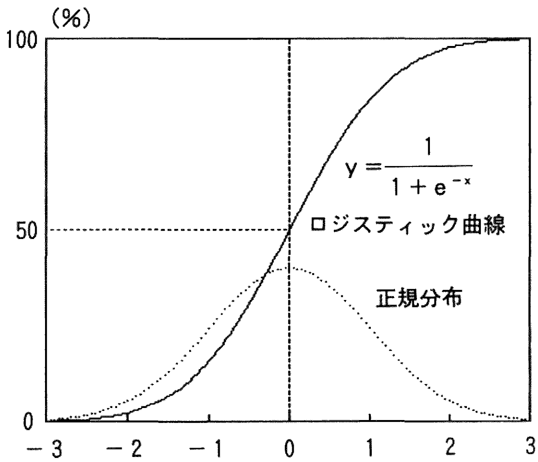

図2　ロジスティック曲線と量－反応関係

また，2つの因子 β_1，β_2 に同時に曝露した場合のオッズ比は，exp（$\beta_1+\beta_2$）＝$OR_1 \cdot OR_2$ として，和ではなく積で示されることに注意されたい[11)18)]．

【例5】多重ロジスティック・モデルの例

表13は冠動脈性心疾患のリスクを12年間にわたって追跡した，有名なコホート研究（Framingham Study）[19)] の成績の一部〔40～49歳男性〕である．

(1) 喫煙は順序尺度[3)]により，0＝吸わないに対する1箱未満のオッズ比は exp（0.4223）＝1.53 であるのに対して，3＝1箱以上のオッズ比は exp（0.4223×3）＝3.55 となっている．表中の z 値から，この場合には量－反応関係[1)]が認められることが予想できる．なお，喫煙者全体のオッズ比を求めるには，1～3を1に変換して計算しなおす必要がある．

(2) 年齢48歳，総コレステロール値 220 mg/dl，収縮期血圧 130 mmHg，相対体重 100 %，ヘモグロビン Sahli 値 120 %，喫煙，心電図異常ともになし＝0の場合を考えると，

$\hat{\beta}_0+\hat{\beta}_1 x_1+\cdots\cdots+\hat{\beta}_7 x_7$

$=-13.2573+0.1216\cdot 48+\cdots+0=-2.5465$

Ⅻ）式に代入して，p＝1/[1＋exp（2.5465)]＝0.0727 となる．すなわち，前述の特性を持つ男性1000人のうち，12年間で72.7人が冠動脈性心疾患を発生すると予想できる．ただし，コホート研究でなく患者－対照研究の場合には，このような予測はできないので注意されたい．

(3) 心電図異常は名義尺度であるものの，有り無しの二

表13　多重ロジスティック・モデルの例〔40～49歳男性[19)]〕

変　　数	(1) 推定値 $\hat{\beta}$	(2) 標準誤差	z 値 (1)/(2)	有意性
X_0 定　　数（$\hat{\alpha}$）	−13.2573	—	—	—
X_1 年　　齢（歳）	0.1216	0.0437	2.783	$p<0.01$
X_2 総コレステロール（mg/dl）	0.0070	0.0025	2.800	$p<0.01$
X_3 収縮期血圧（mmHg）	0.0068	0.0060	1.133	
X_4 相対体重（%）	0.0257	0.0091	2.824	$p<0.01$
X_5 血色素（ザーリ%）	−0.0010	0.0098	−0.833	
X_6 喫煙（0～3＊1）	0.4223	0.1031	4.096	$p<0.001$
X_7 心電図（0，1＊2）	0.7206	0.4009	1.797	

＊1　非喫煙：0，1日1箱未満：1，1箱：2，＞1箱：3

＊2　異常なし：0，異常あり：1

項的な（binomial）データは問題がない．ヨーロッパ人，アジア人，アフリカ人などでは，3種類で＋，－とする場合もある．アジア人（1, 0），アフリカ人（0, 1），ヨーロッパ人が（0, 0）のダミー変数なら，項目数＝種類－1でよい．詳しくは専門書[11)15)]を参照されたい．

参 考 文 献

1) **山本正治, 他監訳**：疫学テキスト，第 2 版，西村書店（新潟），1986.

2) **内藤雅子, 酒井亮二, 訳**：アンケート調査，廣川書店（東京），1992.

3) **遠藤和男, 山本正治**：医統計テキスト，西村書店（新潟），1992.

4) **山本正治**：第 3 章疫学．シンプル衛生公衆衛生学，第 9 版，pp81～101, 2000.

5) **Doll, R., Hill, A.B.**: Lung cancer and other causes of death in relation to smoking. Brit. Med. J., **2**: 1071～1081, 1956.

6) **津崎晃一, 監訳**：数学いらずの医科統計学, pp77, メディカル・サイエンス・インターナショナル（東京），1997.

7) **Woolf, B.**: On estimating the relation between blood group and disease. Ann. Hum. Genet., **19**: 251～253, 1955.

8) **Miettinen, O.S.**: Estimability and estimation in case-referent studies. Am. J. Epidemiol., **103**: 226～235, 1976.

9) **Cohnfield, J.**: A statistical problem arising from retrospective studies. Proceedings of the Third Berkeley Symposium, Vol. Ⅳ (Neyman, ed.) Unv. Calf. Press (Barkeley), pp135～148, 1956.

10) **Gart, J.J., Thomas, D.G.**: The performance of three approximate confidence limits methods for the odds ratio. Am. J. Epidemiol., **115**: 453～470, 1982.

11) **重松逸造, 監訳**：疫学・臨床医学のための患者対照研究．ソフトサイエンス社（東京），1985.

12) **McNemar, Q.**: Note on the sampling error of the differences between correlated proportions or percentages. Psychometrika, **12**: 523～534, 1947.

13) **Fleiss, J.L.**: Statistical methods for rates and proportions, 2 nd ed., John & Wiley (New York), 1981.

14) **遠藤和男**：医学における統計学の応用について（第 1 編）2 × 2 表を中心として．新潟医誌, **102**: 147～154, 1988.

15) **日本医事新報社, 編**：第89回医師国家試験問題及び模範解答．日本医事新報, **No.3701**: 104, 1987.

16) **Mantel, N.**: Chi-square tests with one degree of freedom; extension of the Mantel-haenszel procedure. J. Am. Stat. Assoc., **58**: 690～700, 1963.

17) **遠藤和男, 他**：チリ女性の胆嚢がんの危険因子について．日本公衛誌, **44**: 113～122, 1997.

18) **橋本修二**：14ロジスティックモデルに基づくデータ解析，新しい疫学．pp 225～239, 日本公衆衛生協会（東京），1991.

19) **Truett, J., Cornfield, J., Kannel, W.**: A multivariate analysis of the risk of coronary heart disease in Fremingham. J. Chron. Dis., **20**: 511～524, 1967.

（平成12年 5 月26日受付）