

# 疫学研究における因果モデル作成上の 新しい手法の開発とその応用について

高 木 廣 文

## I. はじめに

疫学研究においては因果関係の解明を最終的な目標とするが、データから得られるのは2つの変数間の相関関係のみであり、これから直接に2変数間の因果関係を立証することはできないとされている<sup>1)</sup>。しかしながら、因果関係の存在する場合、確実に2変数間に相関関係は存在するため、因果関係の存在の前提として、少なくとも相関関係の存在が必要である。この点から、疫学研究では変数間の相関関係を調べることがよく行なわれる。

一般に2変数間の相関の程度を示す指標として相関係数が用いられている。通常、疫学研究では1つの結果（死亡など）に対して、複数の原因が考えられることがある。このような場合、相関係数をもとにして多変量解析の手法を用いることが多い。しかし、このような解析法とは別に、S. Wrightによって考案されたパス解析 path analysis<sup>2)</sup>を用いることもできる。パス解析は主に社会学の分野で多く用いられてきており、変数間の因果関係と相関係数との関係については、Blalockの因果律がよく知られている<sup>3)</sup>。

しかし、従来のパス解析の方法は研究開始の時点で、変数間の因果関係を仮定した特定のモデルについて、その関係の強さを相関係数などから算出することに重点が置かれてきた。しかしながら、多くの疫学研究においては、その研究の初期において、そのような因果モデル causal model を仮定することは、ほとんど不可能である。逆に、調査などによって得られたデータからモデルを作成することを必要とする場合が大部分であろう。

データから因果モデルを作成するために、Takagiは1つのアルゴリズム algorithm を与えており<sup>4)</sup>、今回はその手法の簡単な説明と応用例を示す。例には、食道がんの国別訂正死亡率と食物供給量のデータを用い、因果モデルを作成する。さらに、この結果と従来から行なわれている重回帰分析 multiple regression analysis<sup>1)</sup>の結果を比較し、若干の考察を加えた。

## II. 分析方法

### 1. Takagi による因果連鎖モデル<sup>4)</sup>

各変数間の相関関係から因果関係を示すモデルを作成するために、Takagi は以下の3点を仮定している。

- (a) 相関は因果を示す。
- (b) 各変数の効果 effect は特定の結果と考えられる方向にのみ向かう。
- (c) モデルには効果がフィードバックするようなループは存在しない（すなわち、ある変数が他の変数に影響を及ぼすと、順にその他の変数にも影響し、最終的に初めの変数にその影響が波及するような経路は存在しないものとする）。

上記の仮定とともに、以下に述べる相関係数と因果モデルとの関係は重要である。

#### (i) 因果連鎖と相関<sup>3)</sup>

図1は3変数X, Y, およびZの連鎖モデルを示したものである。Xは他の変数を通じてYに影響を及ぼす。これはXからYに向う矢印により表わすことにする。同様に、YはZに影響を及ぼしている。この場合、X, Y, およびZの3変数の間の相関を考えると、各相関係数の間には、

$$\cdots \rightarrow X \rightarrow \cdots \rightarrow Y \rightarrow \cdots \rightarrow Z \rightarrow \cdots$$

Fig.1 Correlation coefficients between pairs of 3 set variables on causal chain model.

$$\rho_{XZ} \leq \rho_{XZ}, \quad \rho_{XZ} \leq \rho_{XY} \quad (1)$$

が成り立つ。ここで  $\rho_{XZ}$  はXとZの相関係数を示す。

この関係は、図1のような因果連鎖を形づくる変数間では、連鎖上で近い位置にある場合の相関は、遠くにあるものより必ず大きいことを示すものである。この関係は4変数以上についても同様に成り立つことが

簡単に示せる<sup>3)</sup>。

逆に(1)式の関係が成り立つとき、図1のような因果連鎖が必ず存在する保証はない。しかし、仮定の(b)と(c)が正しい場合には、連鎖モデルを設定することは妥当であると考えられる。

## (ii) 直接効果

図2は変数XとZが因果連鎖上にあるとき、新たに変数YがXに影響を与えるものとして追加された状況を示すものである。この場合には、YのZへの影響は、変数Xを経由するものと、直接的にZへ影響するものの2通りを考えることができる。前者は間接効果 indirect effect, 後者は直接効果 direct effect とよばれる。この直接効果の大きさをDとすると、

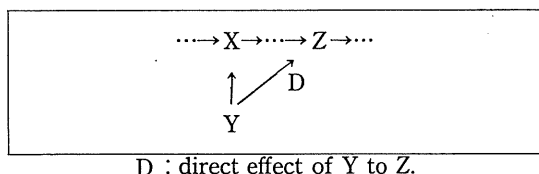


Fig.2 Direct effect of Y to Z on causal model.

$$D = \frac{\rho_{YZ} - \rho_{XY}\rho_{XZ}}{1 - \rho_{XY}^2} \quad (2)$$

によって求めることができる<sup>4)</sup>。

Dは2変数X, Yを独立変数, Zを従属変数とした場合, Yの標準偏回帰係数 standard partial regression coefficient に一致する。Dの有意性を統計的に検定し, 有意ならば図2のように, YからZへのパス path の存在を仮定する。有意でなければ, パスを省略する。

以上述べた(i)と(ii)の考えを用いて, 因果モデルを作成するアルゴリズムを以下に簡単に記述する。

## (iii) 因果モデルのためのアルゴリズム

ここでは, 原因と考えられる変数を $X_i$  ( $i = 1, 2, \dots, p$ )とし, 結果と考えられる変数をYとする。

(a) Yと最大の相関をもつ変数 $X_1$ を選ぶ。

この操作は, (i)で述べた因果連鎖と相関の関係から直接導入できる。

(b) Yまたは既に選択された変数 $X_1$ と最大の相関をもつ変数 $X_2$ を選択する。 $X_1$ の選択に際し,  $X_2$ が $X_1$ との相関によって選ばれた場合, (2)式によって,  $X_2$ のYへの直接効果を計算する。直接効果が事前に決められた統計的有意水準(例えば5%水準)を満たすならば,  $X_2$ とYとをパスによって結びつける。有意でなければ, このパスは存在しないものとする。

この操作は上述の(i)の因果連鎖と相関の関係, および(ii)の直接効果の考えを用いることで導入された。

(c) 上記(b)の操作をくり返し行なう。2変数間の相関係数が有意でなくなった段階で, 変数の選択を終了する。

(d) 選択された変数を用いて, 因果モデルを図に表わし, 各パス係数 path coefficient を数学的に解き, 相関係数の値を代入することで, パス係数の値を推定する。

これは最終的に因果モデルを作り, 各変数の効果を推定するためである。この場合, パス係数の算出にあたり, 重回帰分析を行ない, 標準偏回帰係数をパス係数として用いることが多い。この方法は必ずしも得られたモデルのパス係数の正確な推定値に一致するものではないが, 近似値としては十分であることが多い。

求めたパス係数の有意性を検討し, かりに有意でないのなら, そのパスを取り除き, 再度パス係数の値を求め直す。この場合, 当然モデルを変更する必要がある。

この過程をくり返し行ない, 研究上の仮説設定に十分なモデルとなることを吟味する必要がある。

上述した(a)から(d)までの段階を図示したものを, 図3に示した。

以上がTakagi(1980)による因果モデル作成のアルゴリズムである。従来, 主観的にモデルを仮定していたものが, 幾つかの仮定を置くことにより, コンピュータなどの使用を通して, 自動的にある程度の客観性をもって行なうことが, これによって可能となった。

本報告では, F値が2以上のものを統計的に有意なものとして扱っている。また, Takagiのモデルは, 因果連鎖を基礎として導入されているので, ここでは因果連鎖モデル causal chain model とよぶことにする。

## 2. 重回帰分析

重回帰分析は特定の変数YをP個の変数,  $X_1, X_2, \dots, X_P$ を用いて予測, もしくは推定するために用いられる。変数Yは,

$$\hat{Y} = \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_P X_P + C \quad (3)$$

によって推定され, その値は $\hat{Y}$ となる。ここでCは定数項である。通常は, 全変数を平均0, 分散1に標準化した場合の, 標準偏回帰係数  $\beta_i$  ( $i = 1, 2, \dots, P$ )を求め, この大きさにより, 各変数のYへの効果の大きさを推定する。また, 全変数が標準化された場合, 定数項Cは0となる。

重回帰分析においても, Yへの効果を有すると考えられる変数を選択する必要があるが, 本報告ではこれには前進選択法を用い, 自由度再調整済み重相関係数

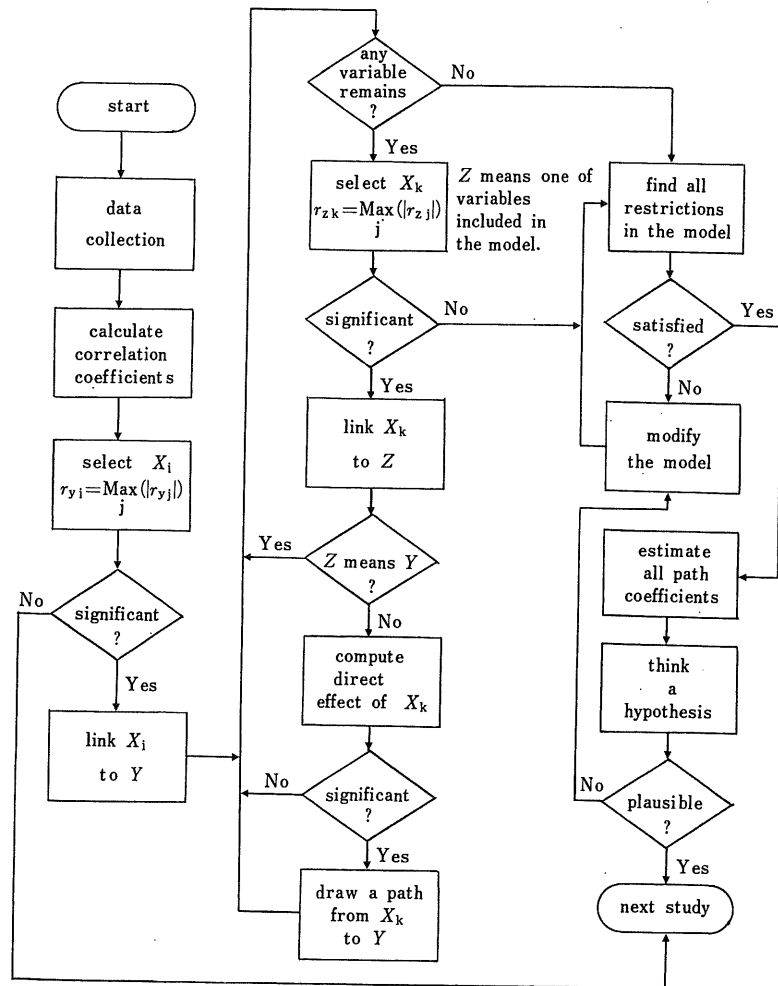


Fig.3 Algorithm for causal modeling

の最大のもののまでを採用した<sup>5)</sup>。

### III. 資料

分析のために用いた資料は, Segi(1979)による1974年の国別癌訂正死亡率から, 男女別の食道癌のデータを<sup>6)</sup>, またこれとは別にFAO(1976)の資料から, 国別の1人1日あたりの食物供給量のデータ(1961年から1965年の平均)を用いた<sup>7)</sup>。

上記の2資料の完備する46ヶ国が分析対象となった。また食物は, Total Energy(熱量), Cereal(穀類), Root(根菜植物類), Sugar(砂糖), Pulse(豆類), Tree Nuts(木の実類), Vegetable(野菜), Fruit(果実類), Meat(肉類), Egg(卵類), Fish(魚類), Milk Products(乳および乳製品), Vegetable Oil(植物油), Animal Oil(動物性油脂), Stimulant(コー

ヒー, 紅茶など), およびAlcohol(アルコール類)の16項目を用いた。

実際の計算には東京大学大型計算機センターのHITAC M-200Hを用い, 重回帰分析にはSPSS<sup>8)</sup>を使用し, 因果連鎖モデルにはFORTRAN 77を用いた自作のプログラムを使用した。

### IV. 結果と考察

男女別の食道癌の訂正死亡率ならびに各食物間の相関係数を表1に示した。因果連鎖モデルも重回帰分析も基本的には, この相関係数行列から分析が始まる。

#### 1. 因果連鎖モデルによる結果

前述したアルゴリズムに従って, 食道癌と食物間の連鎖の状態を求めたものを表2に示した。表2中の

Effect は、各食物と食道癌の間では標準偏回帰係数 に一致し、食物間では相関係数に一致する。

|                     |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
|---------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1)Esophagus:males   | 1.000 |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 2)Esophagus:females | .444  | 1.000 |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 3)Total Energy      | .107  | .081  | 1.000 |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 4)Cereal            | -.059 | -.167 | -.025 | 1.000 |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 5)Root              | .135  | -.007 | .480  | -.279 | 1.000 |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 6)Sugar             | -.004 | .333  | .406  | -.607 | .083  | 1.000 |       |       |       |       |       |       |       |       |       |       |       |       |
| 7)Pulse             | -.227 | -.208 | -.665 | .190  | -.436 | -.309 | 1.000 |       |       |       |       |       |       |       |       |       |       |       |
| 8)Tree Nuts         | .206  | -.088 | -.059 | .259  | -.242 | -.304 | -.023 | 1.000 |       |       |       |       |       |       |       |       |       |       |
| 9)Vegetable         | .337  | -.069 | .352  | .308  | .195  | -.348 | -.213 | .368  | 1.000 |       |       |       |       |       |       |       |       |       |
| 10)Fruit            | -.269 | -.328 | -.375 | -.331 | -.071 | -.165 | .393  | -.056 | -.080 | 1.000 |       |       |       |       |       |       |       |       |
| 11)Meat             | .259  | .249  | .766  | -.440 | .354  | .590  | -.660 | -.168 | .132  | -.253 | 1.000 |       |       |       |       |       |       |       |
| 12)Egg              | .062  | .007  | .718  | -.291 | .335  | .424  | -.635 | .213  | .353  | -.200 | .702  | 1.000 |       |       |       |       |       |       |
| 13)Fish             | .159  | .391  | -.029 | -.193 | .065  | .030  | -.265 | .313  | -.070 | -.184 | -.056 | -.062 | 1.000 |       |       |       |       |       |
| 14)Milk Products    | -.038 | .336  | .702  | -.474 | .375  | .676  | -.629 | -.314 | -.138 | -.227 | .713  | .527  | .244  | 1.000 |       |       |       |       |
| 15)Vegetable Oil    | -.058 | -.212 | .315  | -.119 | .144  | .035  | -.107 | .181  | .193  | .032  | .051  | .299  | .163  | .163  | 1.000 |       |       |       |
| 16)Animal Oil       | .046  | .042  | .750  | -.309 | .515  | .493  | -.623 | -.257 | .097  | -.262 | .756  | .675  | -.148 | .624  | -.087 | 1.000 |       |       |
| 17)Stimulant        | .028  | -.043 | .491  | -.514 | .280  | .533  | -.449 | -.089 | -.050 | .083  | .519  | .523  | .017  | .566  | .337  | .474  | 1.000 |       |
| 18)Alcohol          | .343  | -.115 | .618  | -.048 | .535  | -.007 | -.365 | -.036 | .572  | -.073 | .433  | .381  | -.056 | .200  | .302  | .417  | .291  | 1.000 |
|                     | 1)    | 2)    | 3)    | 4)    | 5)    | 6)    | 7)    | 8)    | 9)    | 10)   | 11)   | 12)   | 13)   | 14)   | 15)   | 16)   | 17)   | 18)   |

Table1 Correlation coefficients of esophageal cancer with food items.

Table2 Variable selection and effect to other variables in causal modeling for esophageal cancer in males.

| From             | → | To            | Effect | F-value(df) | P     |
|------------------|---|---------------|--------|-------------|-------|
| 1.Alcohol        | → | Esophagus     | 0.343  | 5.852(44)   | 0.020 |
| 2.Meat           | → | Esophagus     | 0.429  | 3.581(43)   | 0.065 |
|                  | → | Total Energy  | 0.766  | 62.339(44)  | 0.000 |
| 3.Milk Products  | → | Esophagus     | -0.453 | 5.199(43)   | 0.028 |
|                  | → | Meat          | 0.713  | 45.625(44)  | 0.000 |
| 4.Animal Oil     | → | Esophagus     | -0.348 | 2.534(43)   | 0.119 |
|                  | → | Meat          | 0.756  | 58.596(44)  | 0.000 |
| 5.Total Energy   | → | Alcohol       | 0.618  | 27.213(44)  | 0.000 |
| 6.Cereal         | → | Sugar         | -0.607 | 25.735(44)  | 0.000 |
| 7.Root           | → | Alcohol       | 0.535  | 17.616(44)  | 0.000 |
| 8.Sugar          | → | Milk Products | 0.676  | 37.005(44)  | 0.000 |
| 9.Pulse          | → | Total Energy  | -0.665 | 34.949(44)  | 0.000 |
| 10.Tree Nuts     | → | Vegetable     | 0.368  | 6.902(44)   | 0.012 |
| 11.Vegetable     | → | Alcohol       | 0.572  | 21.343(44)  | 0.000 |
| 12.Fruit         | → | Pulse         | 0.393  | 8.046(44)   | 0.007 |
| 13.Egg           | → | Total Energy  | 0.718  | 46.784(44)  | 0.000 |
| 14.Fish          | → | Tree Nuts     | 0.313  | 4.791(44)   | 0.034 |
| 15.Vegetable Oil | → | Stimulant     | 0.337  | 5.624(44)   | 0.022 |
| 16.Stimulant     | → | Milk Products | 0.566  | 20.757(44)  | 0.000 |

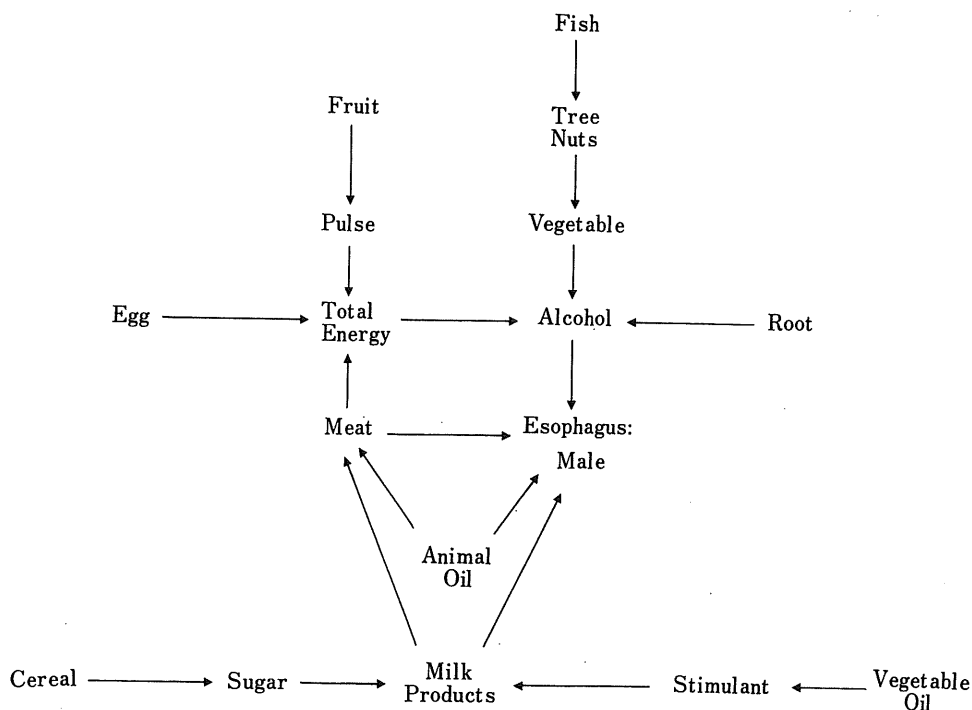


Fig.4 1st step of causal modeling for esophageal cancer in males by the use of sixteen food items.

表2のままでは、変数間相互の関係についての理解が困難であるので、これを図示したものが図4である。図4より、男の食道癌に直接効果のあるものは、Alcohol, Meat, Animal Oil, および Milk Products の4変数である。他の変数には直接効果はなく、上記4変数を通じて間接効果のみを及ぼすものとして取り扱うことにする。ただし、図4からも明らかなように、

Total Energy は Meat の影響を受け、さらにそれが Alcohol に影響し、最終的に男の食道癌に効果を及ぼすという連鎖を構成しており、因果モデル上で重要な役割を果している。

女の食道癌についても同様の手順により、図5に示したような因果モデルが作成できる。図5から明らかに、直接効果のある食物は、Fruit, Fish, Animal Oil,

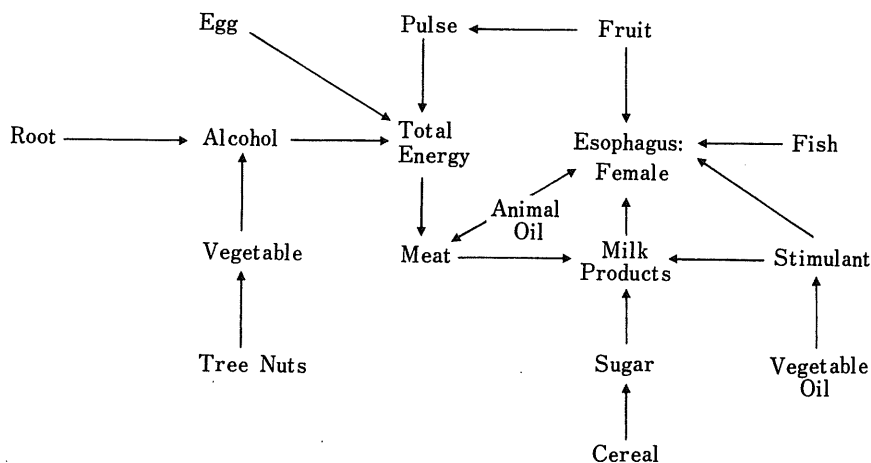


Fig.5 1st step of causal modeling for esophageal cancer in females by the use of sixteen food items.

Milk Products, および Stimulant の5変数であり, 他の変数は間接効果のみと考えられる。また, Pulse, Total Energy, および Meat の3変数は男の食道癌の場合と同様に, 因果連鎖をなし, モデル上意義のあるものと考えられる。

図4, 図5のモデルは全ての変数を含み, この点が

らモデルとしての簡潔さを失っている。モデルは簡単なほど直観的な理解がよく, かつ研究上にも有用と考えられるので, これらモデルの中心部のみを取り出し, より簡単なモデルを図6, 図7のように作成した。図中の $U_1$ から $U_5$ は, 各変数の誤差要因(未知の変数による影響)を示している。

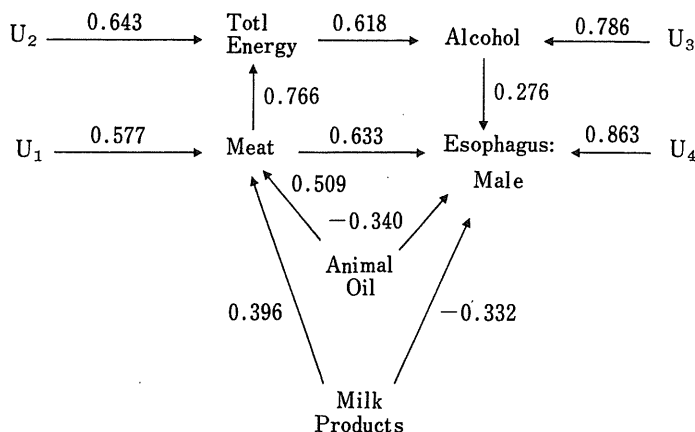


Fig.6 2nd step for causal modeling for esophageal cancer in males by the use of five food items.

\*Numerical values are standard partial regression coefficients.

図6の男の食道癌の場合, 5変数がモデル中に取り入れられた。これらの変数の効果は各パス上の数値(パス係数)により示されている。ここでは, パス係数の推定値として, 重回帰分析による各変数の標準偏回帰係を用いている。

図6の男の食道癌では, Meatのパス係数が0.633と最も正に高く, 次いでAlcoholが0.276と正の値をもつ。すなわち, これらの供給量の増加が, 男の食道癌

の訂正死亡率の増加に直接的に関係するものと考えられる。逆に, Animal OilとMilk Productsでは負のパス係数を持ち, 直接的には訂正死亡率を減少させる傾向をもつことを示すものと考えられる。また, Animal OilとMeat, MeatとTotal Energy, Total EnergyとAlcoholの各間には, 高い正相関があり, 各食物供給量の間では関係が強いことを示している。

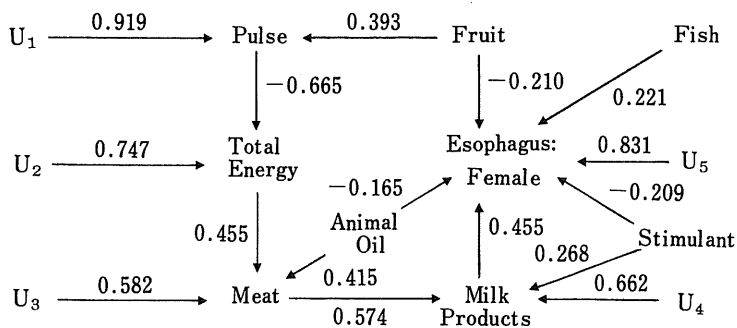


Fig.7 2nd step of causal modeling for esophageal cancer in females by the use of eight food items.

\*Numerical values are standard partial regression coefficients.

図 7 の女の食道癌では Milk Products が正のパス係数 (0.455) をもち、男の場合とは直接効果が逆の関係になっている。また Fish も正のパス係数をもち、これらの供給量の増加が訂正死亡率の増加に関係することを示すものと考えられる。逆に、Fruit, Animal Oil,

および Stimulant などの直接効果は負であり、これらの変数は抑制的に働くことになる。

## 2. 重回帰分析との比較

上記の結果と比較するために、重回帰分析を試みた。

Table3 Variable Selection by Forward Selection Procedure in the case of Multiple Regression Analysis

| Step             | Food          | simple<br>cor.coef. | R <sup>2</sup> | $\hat{R}^2$ | $\hat{R}^2$  | F -<br>value |
|------------------|---------------|---------------------|----------------|-------------|--------------|--------------|
| Esophagus:Male   |               |                     |                |             |              |              |
| 1.               | Alcohol       | 0.343               | 0.117          | 0.097       | 0.078        | 5.852        |
| 2.               | Fruit         | -0.270              | 0.178          | 0.139       | 0.103        | 3.143        |
| 3.               | Total Energy  | 0.107               | 0.246          | 0.192       | 0.140        | 3.800        |
| 4.               | Meat          | 0.259               | 0.355          | 0.292       | 0.232        | 6.965        |
| 5.               | Tree Nuts     | 0.206               | 0.418          | 0.345       | 0.276        | 4.308        |
| 6.               | Egg           | 0.062               | 0.454          | 0.370       | 0.290        | 2.579        |
| 7.               | Vegetable     | 0.337               | 0.482          | 0.386       | <u>0.295</u> | 2.019        |
| 8.               | Cereal        | -0.059              | 0.498          | 0.390       | 0.286        | 1.239        |
| Esophagus:Female |               |                     |                |             |              |              |
| 1.               | Fish          | 0.391               | 0.153          | 0.134       | 0.115        | 7.950        |
| 2.               | Sugar         | 0.333               | 0.256          | 0.222       | 0.189        | 5.983        |
| 3.               | Vegetable Oil | -0.212              | 0.340          | 0.293       | 0.248        | 5.324        |
| 4.               | Fruit         | -0.328              | 0.377          | 0.316       | 0.258        | 2.405        |
| 5.               | Animal Oil    | 0.042               | 0.399          | 0.324       | 0.252        | 1.487        |
| 6.               | Meat          | 0.249               | 0.450          | 0.366       | <u>0.285</u> | 3.641        |
| 7.               | Tree Nuts     | -0.088              | 0.467          | 0.368       | 0.274        | 1.151        |
| 8.               | Vegetable     | -0.069              | 0.493          | 0.383       | 0.278        | 1.913        |
| 9.               | Alcohol       | -0.115              | 0.505          | 0.381       | 0.263        | 0.880        |

R<sup>2</sup>: Squared Multiple Correlation.

$\hat{R}^2$ : Squared Multiple Correlation Adjusted for the Degree of Freedom.

$\hat{R}^2$ : Squared Multiple Correlation Readjusted for the Degree of Freedom.

表 3 に変数選択の過程を示した。男については、自由度再調整済み重相関係数の 2 乗が 0.295 と最も大きい 7 変数、すなわち Alcohol, Fruit, Total Energy, Meat, Tree Nuts, Egg, ならびに Vegetable を選択した。同様にして、女については Fish, Sugar, Vegetable Oil, Fruit, Animal Oil および Meat の 6 変数を選択した。ただし、女では自由度再調整済み重相関係数は Animal Oil を含めることで一度減少するが、これと極めて相関の高い Meat (0.756) を含める

ことで再上昇している。

これらの変数の標準偏回帰係数は表 4 の Beta の欄に示した。この場合、標準偏回帰係数はパス係数と一致するものである。

男の食道癌では、標準偏回帰係数は Meat が 0.803, Alcohol が 0.411 と、前述の因果連鎖モデルの結果と同様、大きな効果をもつことがわかる。しかし、重回帰分析では因果連鎖モデルに選択されなかった変数がいくつか含まれている。それらは Fruit, Tree Nuts,

Egg, および Vegetable の4変数である。

Table4 Partial Regression Coefficients of Selected Variables

| Food                    | Beta   | B        | SE(B)   | F      | P     |
|-------------------------|--------|----------|---------|--------|-------|
| <u>Esophagus:Male</u>   |        |          |         |        |       |
| 1.Alcohol               | 0.411  | 0.0194   | 0.00858 | 5.088  | 0.030 |
| 2.Fruit                 | -0.332 | -0.0180  | 0.00716 | 6.329  | 0.016 |
| 3.Total Energy          | -0.676 | -0.0047  | 0.00174 | 7.487  | 0.009 |
| 4.Meat                  | 0.803  | -0.0144  | 0.00381 | 14.356 | 0.001 |
| 5.Tree Nuts             | 0.292  | 0.0401   | 0.01995 | 4.035  | 0.052 |
| 6.Egg                   | -0.384 | -0.0729  | 0.03964 | 3.384  | 0.074 |
| 7.Vegetable             | 0.236  | 0.0370   | 0.02606 | 2.019  | 0.163 |
| (Constant)              |        | (13.181) |         |        |       |
| <u>Esophagus:Female</u> |        |          |         |        |       |
| 1.Fish                  | 0.356  | 0.0120   | 0.00422 | 8.025  | 0.007 |
| 2.Sugar                 | 0.278  | 0.0027   | 0.00146 | 3.524  | 0.068 |
| 3.Vegetable Oil         | -0.328 | -0.0034  | 0.00128 | 7.164  | 0.011 |
| 4.Fruit                 | -0.219 | -0.0040  | 0.00229 | 2.962  | 0.093 |
| 5.Animal Oil            | -0.415 | -0.0035  | 0.00160 | 4.823  | 0.034 |
| 6.Meat                  | 0.380  | 0.0023   | 0.00119 | 3.641  | 0.064 |
| (Constant)              |        | (1.0097) |         |        |       |

Beta: Standard regression coefficient.

B: Regression coefficient.

SE(B): Standard error of B.

Takagi による因果連鎖モデルは、クラスタ分析 cluster analysis の最近隣法 nearest neighborhood method によく似た手順をとることが判っている<sup>4)</sup>。図4の結果から、Tree Nats, Vegetable および Alcohol の3変数は1つの因果連鎖上にある。いいかえれば、これら3変数は1つのクラスタを構成している。このうち、因果連鎖モデルでは、食道癌に最も近い Alcohol が採用され、重回帰分析では上記3変数全てを選択したことになる。しかし、重回帰分析でも、これら3変数のうち男の食道癌に最も近い Alcohol のF値が、最も大きくなっている。

上記のことは同様に、Egg と Fruit にもあてはまる。この2変数は図4では Total Energy と連鎖を作り、因果連鎖モデルでは Total Energy が、これらのクラスタを代表するものとして採用されたと考えることができる。

重回帰分析における問題としては、Total Energy と Meat が0.766と極めて高い相関をもつことである。このため、両者ともに男の食道癌との相関は各々0.107、

0.259とそれほど高くはない正相関をもつにすぎないものが、変数間の内部相関の影響により、標準偏回帰係数が極めて大きなものとなっている。この点は、多重共線性 multicollinearity<sup>9)</sup>の問題とともに、解釈上の点でも考慮が必要であろう。すなわち、Total Energy は男の食道癌と0.107の正相関をもつにすぎないが、Meat と0.766という高い正相関をもつために、標準偏回帰係数は-0.676と極めて大きな負の値をもつことになったのである。

女の食道癌についての重回帰分析の結果は、図7の因果連鎖モデルに比べ、Animal Oil が負の比較的大きな係数をもつが、Fish と Fruit では数値上大差はない。しかし、重回帰分析では Meat が直接効果のあるものとして扱われており、図7のモデルでは Milk Products を通じての間接効果を与えるものとされているのと異なっている。また、男の食道癌の場合と同様、Vegetable Oil は因果モデルでは Stimulant により、その所属するクラスタを代表されている。同様のことは、Sugar と Milk Products の間にもいえる。



上述のように因果連鎖モデルと重回帰分析では共通点もあるが、相違点もかなりある。しかし、この相違点も因果連鎖モデルの作成過程を検討することで、ある程度説明が可能であり、さらに重回帰分析だけでは得られない情報も与えることができると考えられる。両者を同時に使用することにより、一方だけの使用による解釈上の誤りなども避け得るのではないかと考えられる。

今回は、因果連鎖モデルと重回帰分析の結果のみを比較したが、この他の多変量解析、とくに因子分析法は近年医学分野でも用いられるようになり、いくつかの報告もある<sup>10-12)</sup>。これらの分析方法との比較も、実際のデータを通じて検討することは、極めて意義のあることと考えられるので、今後この点についても比較を行なうつもりである。

最後に本報告で使ったデータについて述べる必要がある。一般に、疫学研究において因果関係の解明に迫るには、地域を単位とした地域相関（間接相関）ではなく、個人を単位とした個人相関（直接相関）を用いる方が優れている<sup>13)</sup>。これは、地域単位での結果が必ずしも個人に還元されないことから明らかであろう。しかし、疫学研究において地域相関や地理分布などを検討することも、また必要なことである。この理由の一つとしては、各疾患の地域差が存在するという事実、また一般に地域により食生活なども異なるということが考えられる。これらの点から、広範囲にわたって個人のデータが得られない場合や、研究の初期において仮説を立てる場合などに地域相関を用いることは有用であると考えられる。しかしながら、今後個人単位のデータに基づく分析も試みる必要はある。

本報告において、因果連鎖モデルの応用について述べてきたが、研究上のモデルが必ずしも実態をそのまま表わすものではないという認識が必要かもしれない。そのモデルが最終的なものでない限り、あくまでも仮説的なものであり、次の研究に有用な示唆を与えらるという点に価値があるものと考ええる。しかし、上述のような手法を重回帰分析などと同時に使用すること

により、疫学的な因果関係の解明のための強力な方法になり得ると考えられる。

## V. まとめ

Takagi による因果連鎖モデルの説明とその応用について、重回帰分析との比較を通じて行なった。資料には食道癌の男女別46ヶ国の訂正死亡率データと、食物供給量のデータを用いた。得られた結果は以下の通りである。

1. 男の食道癌について、因果連鎖モデルでは Alcohol, Meat, Animal Oil, Milk Products, および Total Energy の5変数が選ばれた。このうち Alcohol と Meat は正の直接効果を、Animal Oil と Milk Products は負の直接効果をもつ。重回帰分析では、Alcohol, Fruit, Total Energy, Meat, Tree Nuts, Egg および Vegetable の7変数が選ばれた。このうち Alcohol, Meat, Tree Nuts, および Vegetable は正の直接効果を、Fruit, Total Energy, Egg は負の直接効果をもつ。

2. 女の食道癌について、因果連鎖モデルでは Fish, Fruit, Pulse, Total Energy, Meat, Animal Oil, Milk Products, および Stimulant の8変数が選ばれた。このうち、Fish と Milk Products は正の直接効果を、Fruit, Animal Oil, Stimulant は負の直接効果をもつ。

重回帰分析では Fish, Sugar, Vegetable Oil, Fruit, Animal Oil, および Meat の6変数が選ばれた。このうち、Fish, Sugar, Meat は正の直接効果を、Vegetable Oil, Fruit, Animal Oil は負の直接効果をもつ。

3. 上記結果に考察を加え、因果連鎖モデルと重回帰分析の共通点および相違点を指摘し、あわせて疫学研究における因果連鎖モデルの有用性を指摘した。

なお本報告の要旨は第9回日本行動計量学会総会で発表した<sup>14)</sup>。

## 参考文献

- 1) Cohen, J. and Cohen, P.: Applied multiple regression/Correlation analysis for behavioral sciences, John Wiley & Sons, New York, 1975
- 2) Wright, S.: Correlation and causation, Journal of Agricultural Research, 20: 557-585, 1921
- 3) Blalock, H.M. Jr.: Causal inferences in nonexperimental research, the Univ. of North Carolina, Chapel Hill, 1964
- 4) Takagi, H.: An algorithm for causal modeling in epidemiological studies, Behaviormetrika, 8: 41-55, 1980
- 5) 柳井晴夫, 高根芳雄: 多変量解析法, 朝倉書店, 東京, P57, 1977
- 6) Segi, M.: Age-adjusted death rates for cancer for selected sites (A-classification) in 51 countries in 1974, 1979
- 7) Food Agriculture Organization of the United Nations: Food supply-calories per caput per day, Monthly Bulletin of Agricultural Economics and Statistics, 25(4), 1976
- 8) Nie, N.H., Hull, C.H., et al: Statistical Package for the

- Social Sciences, 2nd edit., McGraw-Hill, New York, 1975
- 9) Gnanadesikan: Methods for Statistical Data Analysis of Multivariate Observations, John Wiley & Sons, 1977, 丘本正, 磯貝恭史訳, 統計的多変量データ解析, 日科技連, 東京, P 20, 1979
- 10) Howell, M.A.: Factor analysis of international cancer mortality data and per capita food consumption, Br.J. Cancer, 29: 328-336, 1974
- 11) Yanai, H., Inaba, Y., Takagi, H., and Yamamoto, S.: An epidemiological study on mortality rates of various cancer sites during 1958 to 1971 by means of factor analysis, Behaviormetrika, 5: 55-74, 1978
- 12) Yanai, H., Inaba, Y., Takagi, H., and Yamamoto, S.: Multivariate analysis of cancer mortalities for selected sites in 24 countries in the world, Environmental Health Perspectives, 32: 83-101, 1979
- 13) 山本俊一, 稲葉裕, 柳井晴夫, 高木廣文: がんの疫学統計の方法に関する研究, 厚生省「がんの疫学」班分担研究報告, 1980
- 14) 高木廣文, 稲葉裕, 柳井晴夫: がん死亡率と食物消費量の相関分析—因果連鎖モデルによる解析—, 第9回日本行動計量学会総会発表論文抄録集, 152-155, 1981