研究論文

比例ハザード・モデルを用いた長期低線量 放射線被曝による健康影響の評価

服部芳明^{*1,*}, 前田和甫^{*1}

(1981年12月8日受理)

(1982年3月3日再受理)

An Application of the Proportional Hazard Model to the Assessment of Health Effects by Long Term and Low Level Radiation Exposure

Yoshiaki HATTORI^{*1,*} and Kazuho MAEDA^{*1}

As a method for analysing epidemiological observations from populations occupationally exposed to radiation, the proportional hazard model was applied to the simulated observations on the mortality assuming 40 years follow-up. The risk factor of cancer death for compiling the simulation is based on the value proposed by ICRP. The applicability of this method is discussed.

Using the proportional hazard model, the estimated least population size for detecting the excess cancer deaths is 50 to 150 thousand. Two alternative conditions are also investigated using simulated observations, *i.e.*, the risk factor is assumed 10 times greater, and the length of follow-up study period is shortened to 30 years.

This method was compared to the multiple logistic model applied to deceased samples. The former method was more adoptive from the viewpoint of reducing the effects from time trend factors.

Therefore, the proportional hazard model is applicable to assess the risk level of cancer death from radiation exposures based on the follow-up studies of human populations.

KEY WORDS : proportional hazard model, health effects, low level radiation exposure, simulated epidemiological observation, follow-up study, cancer mortality

I 序 論

放射線がヒトに与える影響に関する研究は高線量被曝による影響よりも低線量被曝とくに長期にわたる低線量率での被曝による影響を見出すことに重点が移ってい

る¹⁾。長期の低線量被曝は、動物実験からの外挿がどこまで可能か今日いまだに定式化されておらず、したがってヒトの集団を長期間追跡して得られたデータの疫学的観察に基づくリスクの推定が重要となる。

近年においてはアメリカの Hanford の原子力施設の放射線被曝作業者について追跡調査した結果の研究が多数報告されている²⁻⁵⁾。著者ら⁶⁾はこれらの報告がそれぞれ異なる結論に至っていることに着目し、この違いは因果関係を検討するためのデータ解析の際の「みかけの関連」を補正する方法の違いによるものと考え、多変量モデルによる解析が現象に比較的近い結論を導くのに役立つと論じ、多重ロジスティック・モデル (multiple logistic model, 以後 MLM と略記) の有効性についても同

^{*1} 東京大学医学部保健学科疫学教室; 東京都文京区本郷 7-3-1 (〒113)

Department of Epidemiology, School of Health Sciences, Faculty of Medicine, University of Tokyo; 7-3-1 Hongo, Bunkyo-ku, Tokyo.

* 現在: 日本原子力事業株式会社; 神奈川県川崎市川崎区浮島町 4-1 (〒210)

Present Address: Nippon Atomic Industry Group Co., Ltd.; 4-1 Ukishima-cho, Kawasaki-ku, Kawasaki-shi, Kanagawa-ken.

時に検討を行った。

長期低線量被曝の影響をみるための疫学的データはその性質上、追跡調査により得られたデータであることが望ましく、それを解析するための多変量モデルは時間で変化する要因を考慮したものがより適切である。そこで今回は Cox⁷⁾ が生命表の回帰モデルとして提唱した比例ハザード・モデル (proportional hazard model, 以後 PHM と略記) を取り上げた。

わが国では過去において、疫学的データを得るための長期放射線被曝者の追跡調査は、各方面でその必要性についての様々な議論はされているが、今回は関連事業者からの記録の閲覧への協力が得られず、解析データの入手はできなかった。著者らの懇請に対して目的への理解と関心は示されても協力は得られなかったものである。そこで本研究ではシミュレーションによって仮想的に得られたデータへの適用で手法の検討を行うこととした。

PHM は不可逆的变化の個体における発生について影響をおよぼす多くの要因を同時に見出すために用いられるモデルで、ここで取扱う追跡調査からの放射線のリスクの推定に適していよう。本研究では職業上の放射線被曝を受けるヒトの集団での癌による死亡を追跡した調査を仮定したシミュレーションからのデータに対して PHM による解析を適用し、この手法による検出力、妥当性および有効性を検討したものである。

II 方 法

1. シミュレーションの方法

職業上の放射線被曝を受ける集団を 40 年間追跡することを仮定したシミュレーションを行った。性別はすべて男とした。

(1) 年齢構成

追跡開始時を第 0 年次とし、その時点の年齢構成は 25~34 歳の離散一様分布によって定めた。

(2) 死亡率

死因を、1. 白血病、2. 固形癌、3. その他の死因、の 3

とおりとした。各々の死因について、年齢 a 、追跡開始より t 年次における死因 i ($i=1, 2, 3$) についての死亡率 $M_i(a, t)$ を次の関数で定めた。

$$M_i(a, t) = 10^{-5} \exp(\alpha_{i0} + \alpha_{i1}a + \alpha_{i2}t + \alpha_{i3}at) \quad (1)$$

$\alpha_{i0}, \alpha_{i1}, \alpha_{i2}, \alpha_{i3}$ は定数であり、これらは各死因ごとに次のように決定した。日本の 1975~80 年の毎年の 5 歳階級男子死亡率^{8~13)} の 20~79 歳の分の自然対数に対して、年齢、年次から 1980 をひいた値、およびこの 2 つの積の 3 変数を用いた重回帰を行い、その回帰係数の値を $\alpha_{i0} \sim \alpha_{i3}$ とした。この $\alpha_{i0} \sim \alpha_{i3}$ の値と回帰の際の重相関係数の値とを **Table 1** に示した。(1) 式の $M_i(a, t)$ から前項で定めた年齢構成の 10 万人に対する観察期間 40 年次までの死亡数期待値を死因別に **Fig. 1** に示した。

(3) 放射線による過剰の癌死亡のリスク

低線量放射線による過剰の癌死亡のリスク係数を IC RP Pub. 26¹⁵⁾ の (60) 項では約 10^{-2}Sv^{-1} としている。就労年齢を通じての被曝で全リスクの 80% が現れると

Fig. 1 Expected number of naturally occurring deaths per 100,000 for each cause.

Table 1 Parameters for the death rate functions.

Cause of death	Parameter				R^*
	α_{i0}	α_{i1}	α_{i2}	α_{i3}	
1 Leukemia	-0.24013	0.039931	-0.089578	0.00216566	0.9907
2 Solid cancers	-0.31796	0.103380	-0.014865	0.00022024	0.9959
3 Others	2.13925	0.084585	-0.058282	0.00035869	0.9854

R^* Correlation coefficients of the regression.

Death rate function: $M_i(a, t) = 10^{-5} \exp(\alpha_{i0} + \alpha_{i1}a + \alpha_{i2}t + \alpha_{i3}at)$, where a is age, t is year.

ICRP Pub. 27¹³⁾の(39)項に推定されており、他の死因による死亡が存在しない場合の放射線による癌死亡のリスクとして $1.25 \times 10^{-5} \text{ mSv}^{-1}$ をシミュレーションに用いた。このうち 1/5 の $0.25 \times 10^{-5} \text{ mSv}^{-1}$ を白血病について、4/5 の $1.00 \times 10^{-5} \text{ mSv}^{-1}$ を他の固形癌についてのリスク係数とした。

(4) 被曝から癌死亡までの潜伏期間

癌死亡の原因となる放射線被曝から死亡までの期間の分布を ICRP Pub. 27 の(12)項¹³⁾、BEIR III¹⁴⁾の p. 137 に記載されているものを参考とし、白血病は 2~22 年の幅で中央値 12 年、固形癌は 10~30 年の幅で中央値 20 年の整数値に丸めた三角形分布となるように定めた。両者の幅が等しくなるようにしたため固形癌の潜伏期間はやや短くなった。

(5) 職業上の被曝線量と被曝期間

個人の年間の職業上の被曝線量の集団内での分布は、ICRP Pub. 26¹⁵⁾の(100)項および UNSCEAR '77¹⁶⁾、Annex E の第 30 項で、約 5 mSv を平均とした対数正規分布にはほぼ従うといわれている。本研究では、各個人についての年間被曝線量の期待値の分布を、平均 1.2、分散 1 の正規分布の指数の値 (単位 mSv) をとり、33 mSv を上限としてそれ以上は 33 mSv としたもので定めた。各個人において年次間の分布が前記の期待値の 0.5~1.5 倍の一様分布となるように個人の年間の被曝線量を決定した。個人の年間被曝線量の期待値の分布、および、10 万人の 1 年間の被曝をシミュレートして得られた線量の 1 mSv 階級の分布を Fig. 2 に示した。後者の分布の平均値は約 5.36 mSv、分散は約 42 mSv² で

ある。

職業上の被曝を受ける期間は追跡開始時より 55 歳までとし、それ以前とそれ以後は職業上の被曝はないものとした。

(6) シミュレーションの実行

追跡調査の対象となる標本の人口を 5,000 人、10,000 人、15,000 人とした各々の場合、および人口が 10,000 人で放射線の癌死亡に対するリスク係数が(3)項で定めたものの 10 倍の大きさである場合の 4 とおりについて、各々 10 回ずつシミュレーションを実行した。

2. シミュレーションからのデータの解析

シミュレーションで生成した追跡調査のデータをすべて PHM^{7,18)}を用いて解析した。この方法自体の説明は付録に述べた。

今回の解析のハザード関数に対するモデルは次式のように定めた。

$$\begin{aligned}\lambda(t; \mathbf{z}_t) &= \lambda_0(t) \exp(\mathbf{z}_t' \boldsymbol{\beta}) \\ &= \lambda_0(t) \exp(\beta_c c_t + \beta_a a) \quad (2)\end{aligned}$$

ただし、 $\mathbf{z}_t = (c_t, a)'$ 、 $\boldsymbol{\beta} = (\beta_c, \beta_a)'$ 、 $\mathbf{z}_t'$ などは $\mathbf{z}_t$ などの転置である。説明変数ベクトル $\mathbf{z}_t$ の第 1 成分 c_t は職業上の被曝の線量当量を第 t 年次 ($t=0, 1, \dots, 39$) まで累積したもの (以下、集積線量と記す)、第 2 成分 a は第 0 年次における年齢である。パラメータベクトル $\boldsymbol{\beta}$ の成分の β_c と β_a は各々 c_t と a の係数である。 $\lambda(t; \mathbf{z}_t)$ は個体の観測値ベクトル $\mathbf{z}_t$ のもとでの第 t 年次の癌 (白血病または固形癌) による死亡率、 $\lambda_0(t)$ は予測の必要ない t についてのある関数である。本研究では放射線による過剰の癌死亡を推定するため、他の死因による死

Fig. 2 Distribution of annual exposure dose: expected personal dose and a trial of simulation with population of 100,000.

亡はすべて打ち切り標本として解析をすすめた。

(2) 式のモデルでパラメータベクトル β の最尤推定値¹⁷⁾ $\hat{\beta} = (\hat{\beta}_c, \hat{\beta}_a)'$ を離散時刻の観測の場合の近似尤度¹⁸⁾から求めた。解は NEWTON-RAPHSON 法¹⁹⁾によって求めた。推定値の分散を情報行列^{17,18)} $I(\beta)$ に推定値 $\hat{\beta}$ を代入してその逆行列の対角成分によって推定した。

この方法による解析結果から検出力を検討した。集積線量の係数 β_c とその推定値の標準誤差が少なくとも同じぐらいの値にならなければ放射線の影響として過剰の癌死亡があったことを検出できない。 $\hat{\beta}_c$ の標準誤差は情報行列から求める $\hat{\beta}_c$ の分散の推定値の 10 回の平均の平方根をとることで推定した。パラメータ推定値の分散は標本数の逆数に比例する¹⁷⁾ため、解析結果を用いてパラメータ推定値と標本数の関係が推定できる。ただし、PHM のパラメータ β の推定は偏尤度 (partial likelihood)^{18,20)}を用いて行うためパラメータ推定値の分散は観測された死亡数の逆数に比例するが、ここでは死亡数は標本数に比例するため標本数との関係として表した。

一方、 β_c の値、すなわち $\hat{\beta}_c$ の期待値は 10 回の試行の平均では β_c 自身が小さいために十分な推定ができない。このため、放射線の癌死亡へのリスク係数を 10 倍に仮定したシミュレーションからのデータを PHM によって解析した結果を、リスク係数の値と検出力について検討するとともに次のように利用した。単位線量当りの癌死亡に対するリスク係数、 $1.25 \times 10^{-5} \text{ mSv}^{-1}$ (白血病と固形癌についての和) を r とする。このとき β_c の値を $\beta_c^{(r)}$ 、リスク係数が 10 倍であった場合の β_c の値を $\beta_c^{(10r)}$ と表すと、ハザード関数の集積線量による増分の値が 10 倍となることで近似的に表せる。

$$\left[\frac{\partial}{\partial c_i} \lambda(t; z_i) \right]_{\beta_c = \beta_c^{(10r)}} \cong 10 \left[\frac{\partial}{\partial c_i} \lambda(t; z_i) \right]_{\beta_c = \beta_c^{(r)}} \quad (3)$$

これに、

$$\partial \lambda(t; z_i) / \partial c_i = \beta_c \lambda_0(t) \exp(z_i' \beta) \quad (4)$$

を代入し両辺に共通なものを消去すると、

$$\beta_c^{(10r)} \exp(\beta_c^{(10r)} c_i) \cong 10 \beta_c^{(r)} \exp(\beta_c^{(r)} c_i) \quad (5)$$

となり、 $\beta_c^{(r)}$ 、 $\beta_c^{(10r)}$ もともに非常に小さい値をとるため

$$\exp(\beta_c^{(10r)} c_i) \cong \exp(\beta_c^{(r)} c_i) \cong 1 \quad (6)$$

となる。ゆえに (5) 式はさらに

$$\beta_c^{(10r)} \cong 10 \beta_c^{(r)} \quad (7)$$

と近似され、単位線量当りのリスクと係数 β_c の値とが β_c が十分に小さければ比例することが仮定できる。推定値の絶対誤差が等しければ $\hat{\beta}_c^{(10r)}$ の方が $\hat{\beta}_c^{(r)}$ より

も相対誤差が 1/10 となることが期待され、(3)、(6) 式からの近似による誤差を考慮しても β_c の値の推定値として $\hat{\beta}_c^{(10r)}/10$ の方が $\hat{\beta}_c^{(r)}$ よりも精度が高いといえ、これを採用できる。

追跡調査期間の長さは推定の精度に大きく影響する。標本数 15,000 人のシミュレーションからのデータを用いて 30 年間で追跡を終えたとして PHM による解析を行い、40 年間追跡した場合と比較した。30 年間の癌死亡数の期待値は標本数の 3.0%、40 年間では 8.6% であるから、この解析モデルが正確に成立っていればパラメータ推定値の分散は約 2.9 倍となるはずである。

PHM との比較のために、MLM (多重ロジスティック・モデル)²¹⁾を用いた解析を標本数 15,000 人のシミュレーションからのデータに適用した。MLM は今回は次式で表した。

$$\begin{aligned} \log p(x) / \{1 - p(x)\} &= x' \theta \\ &= \theta_0 + \theta_c c + \theta_a a + \theta_y y \end{aligned} \quad (8)$$

ただし、 $x = (1, c, a, y)'$ 、 $\theta = (\theta_0, \theta_c, \theta_a, \theta_y)'$ 、 $\log$ は自然対数である。 $p(x)$ は死亡者の死因が癌である確率とする。説明変数ベクトル x の第 1 成分は定数 1、 c は死亡時までの集積線量、 a は第 0 年次における年齢、 y は死亡年次、パラメータベクトル θ の成分 $\theta_0, \theta_c, \theta_a, \theta_y$ はそれぞれ説明変数の係数である。この解析は全標本ではなく、死亡例だけを用いて行った。 θ の最尤推定値 $\hat{\theta} = (\hat{\theta}_0, \hat{\theta}_c, \hat{\theta}_a, \hat{\theta}_y)'$ を NEWTON-RAPHSON 法によって求めた。 $\hat{\theta}$ の分散を情報行列の逆行列から推定した。

なお、計算はすべて東京大学大型計算機センターの HITAC M200H により行い、プログラム言語は VOS3 最適化 FORTRAN 77 を用いた。

III 結 果

1. 標本数を変化させた場合

標本数を 5,000, 10,000, 15,000 とした各 10 回のシミュレーションからのデータを PHM により解析した。その結果得られたパラメータ推定値とその分散の推定値の 10 回についての平均値と標準偏差 (不偏分散より求めた) を、Table 2 に示した。

集積線量の係数の推定値 $\hat{\beta}_c$ の分散の平均値と標準偏差を標本数の逆数に対してプロットし、Fig. 3 に示した。プロットの標準誤差は図の標準偏差の 1/10 となる。標本数 5,000, 10,000, 15,000 に対して $\hat{\beta}_c$ の分散は各々 8.07×10^{-8} , 4.08×10^{-8} , $2.71 \times 10^{-8} \text{ mSv}^{-2}$ となり、ほぼ原点を通る直線上にのった。

$\hat{\beta}_c$ の平均値は $10^{-5} \sim 10^{-4} \text{ mSv}^{-1}$ でありばらつきが大きく、 $\hat{\beta}_a$ 平均値は $0.10 \sim 0.12 \text{ year}^{-1}$ でありばらつき

Table 2 Results of the proportional hazard model applied to the simulations using 3 different sample sizes.

Size	5,000		10,000		15,000	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
$\hat{\beta}_c$	0.918×10^{-5}	27.550×10^{-5}	4.335×10^{-5}	31.831×10^{-5}	5.569×10^{-5}	17.391×10^{-5}
var ($\hat{\beta}_c$)	8.072×10^{-8}	1.314×10^{-8}	4.082×10^{-8}	0.541×10^{-8}	2.713×10^{-8}	0.204×10^{-8}
$\hat{\beta}_a$	10.252×10^{-2}	1.759×10^{-2}	10.603×10^{-2}	0.966×10^{-2}	11.778×10^{-2}	1.102×10^{-2}
var ($\hat{\beta}_a$)	29.032×10^{-5}	1.185×10^{-5}	14.096×10^{-5}	0.427×10^{-5}	10.084×10^{-5}	0.299×10^{-5}

Mean, S.D.: means and standard deviations of 10 trials for each class.

$\hat{\beta}_c$: Estimated coefficient for cumulative dose (mSv^{-1}).

$\hat{\beta}_a$: Estimated coefficient for age at 0th year (year^{-1}).

Fig. 3 Relationship between sample size and variance of estimated coefficient for cumulative dose using proportional hazard model.

Note: Plots indicate mean values and standard deviations in analyses of the simulated observations.

は小さかった。

2. リスクを10倍とした場合

標本数 10,000 で放射線の癌死亡に対するリスク係数を10倍としたシミュレーションからのデータを PHM により解析した。その結果得られたパラメータ推定値およびその分散の推定値の10回の試行についての平均値と標準偏差を **Table 3** に示した。

集積線量の係数の推定値 $\hat{\beta}_c$ の平均値は 6.82×10^{-4}

Table 3 Results of the proportional hazard model applied to the simulations using a 10 time greater risk factor.

	Mean	S.D.
$\hat{\beta}_c$	68.189×10^{-5}	12.225×10^{-5}
var ($\hat{\beta}_c$)	2.841×10^{-8}	0.175×10^{-8}
$\hat{\beta}_a$	8.839×10^{-2}	0.891×10^{-2}
var ($\hat{\beta}_a$)	13.125×10^{-5}	0.290×10^{-5}

See the note of **Table 2**.

mSv^{-1} となり、その分散の推定値の平均値は $2.84 \times 10^{-8} \text{mSv}^{-2}$ であった。標準誤差になおすと $1.69 \times 10^{-4} \text{mSv}^{-1}$ で、 $\hat{\beta}_c$ の平均値との比は約 0.25 倍であった。この結果から (7) 式でリスク係数が10倍でない場合の β_c の値を1標準誤差の範囲で推定すると $5.13 \sim 8.51 \times 10^{-5} \text{mSv}^{-1}$ となった。これは前節の結果とも一致した。

また、この値の2乗 ($2.63 \sim 7.24 \times 10^{-9} \text{mSv}^{-2}$) が **Fig. 3** に示した β_c の推定値の分散と等しくなるような標本数は、この図の直線が、 $\hat{\beta}_c$ の分散と標本数との積がおおよそ $4.0 \times 10^{-4} \text{person} \cdot \text{mSv}^{-2}$ となることから約 5~15 万人と推定された。すなわち、 β_c の値とその標準誤差がほぼ等しくなる標本数がこの人数となった。

3. 追跡期間を短縮した場合

標本数が 15,000 の場合のシミュレーションからのデータで追跡期間を 3/4 の 30 年間として PHM により解析した。得られたパラメータの推定値とその分散の推定値の10回の平均値と標準偏差を **Table 4** に示した。

集積線量の係数の推定値 $\hat{\beta}_c$ の平均値は 1.993×10^{-4}

Table 4 Results of the proportional hazard model applied to the simulations for a follow-up period shortened by 3/4.

	Mean	S.D.
$\hat{\beta}_c$	19.925×10^{-5}	20.024×10^{-5}
var ($\hat{\beta}_c$)	10.873×10^{-8}	0.764×10^{-8}
$\hat{\beta}_a$	12.073×10^{-2}	1.510×10^{-2}
var ($\hat{\beta}_a$)	28.523×10^{-5}	1.353×10^{-5}

See the note of **Table 2**.

mSv⁻¹ であった。 $\hat{\beta}_c$ の分散推定値の平均値は 1.087×10^{-7} mSv⁻² で第 1 節の 40 年の場合の 2.713×10^{-8} mSv⁻² に対して約 4 倍となった。第 0 年次の年齢の係数の推定値 $\hat{\beta}_a$ の分散 (2.852×10^{-4}) は 40 年の場合 (1.008×10^{-4}) とくらべ約 2.8 倍となり, これは理論値 (II 章 2 節, 2.9 倍) とほぼ一致した。

4. MLM による解析

標本数 15,000 のシミュレーションの **Table 5** に示した数の死亡例のデータ 10 組に対して MLM を用いて解析した。その結果得られたパラメータの推定値とその分散の推定値の 10 回の平均値と標準偏差を **Table 6** に示した。

集積線量の係数の推定値 $\hat{\theta}_c$ の平均値は 1.86×10^{-4} mSv⁻¹, 分散推定値の平均値を標準誤差になおすと 2.25×10^{-4} mSv⁻¹ で, 標準誤差は θ_c の値の約 1.2 倍と推定された。第 0 年次の年齢の係数の推定値 $\hat{\theta}_a$ の平均値は 2.22×10^{-2} year⁻¹, $\hat{\theta}_a$ の分散の平均値を標準誤差にして 1.36×10^{-2} year⁻¹ となり標準誤差と θ_a の比は約 0.6 倍と推定された。PHM の同じシミュレーションからの解析による β_a の推定では, 標準誤差と期待値の比が約

Table 5 Decreased samples observed from the simulations applied to population of 15,000.

Total	Cancer	Non-cancer
2963	1301	1662
2909	1258	1651
2912	1251	1661
3050	1333	1717
2984	1220	1764
3070	1308	1762
3014	1316	1698
2993	1266	1727
2966	1269	1697
2960	1268	1692

Table 6 Results of the multiple logistic model applied to the diseased samples from simulations.

	Mean	S.D.
$\hat{\theta}_0$	-2.531	0.437
var ($\hat{\theta}_0$)	1.912×10^{-1}	0.046×10^{-1}
$\hat{\theta}_c$	1.861×10^{-4}	2.744×10^{-4}
var ($\hat{\theta}_c$)	5.073×10^{-8}	0.380×10^{-8}
$\hat{\theta}_a$	2.223×10^{-2}	1.298×10^{-2}
var ($\hat{\theta}_a$)	1.853×10^{-4}	0.041×10^{-4}
$\hat{\theta}_y$	5.300×10^{-2}	0.313×10^{-2}
var ($\hat{\theta}_y$)	1.788×10^{-5}	0.056×10^{-5}

mean, S.D.: Means and standard deviations of 10 trials.

$\hat{\theta}_0$, constant; $\hat{\theta}_c$, coefficient for cumulative dose (mSv⁻¹); $\hat{\theta}_a$, coefficient for age of 0th year (year⁻¹); $\hat{\theta}_y$, coefficient for year of death. (year⁻¹).

0.09 であったのにくらべ, 推定の誤差は年齢の係数についてみた場合, 大きかった。

IV 考 察

1. 対象の仮定について

今回のシミュレーションでは追跡の対象として職業上の被曝を受ける作業者の集団を仮定した。

長期低線量被曝の影響についての疫学的研究を行うための資料としては, 一般の人口を対象として自然放射線の量と疾病・死亡統計などを用いたものや, 医療の目的での被曝を受けた集団を対象とした調査なども考えられる。しかし, 被曝線量の記録の得やすさ, 対象の均一性などの点から, 長期低線量被曝の影響の調査対象としては職業上の被曝を受ける集団が最も適切である。また, 現在では被曝線量の多さからも, 問題として取り上げるべき対象は一般人よりも職業上の被曝を受ける集団であろう。

本研究では放射線のヒトの健康への影響として癌による死亡を取上げた。遺伝的影響や催奇形性も低線量被曝の影響として問題とされるが, 癌の誘発についての研究は疫学的なデータを用いた方法としては最も取り組みやすいと思われる。とくに, 遺伝的影響を疫学調査によって研究するならば, 何世代にもわたる追跡が必要となり, 実用上の困難が多いと思われる。

癌が誘発されても必ずしも死亡にまで至るとは限らない。実際には発症のデータを用いるのが望ましいが, 大標本を追跡する調査においては一般に癌の発生を追跡す

ることは、各地の癌登録の実態を考えれば難しく、死亡のデータの方が癌の発生に関する情報として今日では精度が高いと思われる。発症数と死亡数の違いは当然、癌の部位によって異なるであろうが、多くの場合、発症と死亡の間に強い関連があることは確かである。とくに、比較的低線量の反復被曝者における白血病では、致命率の高い慢性骨髄性白血病が一般の白血病患者に比べて多いといわれており^{22,23)}、このような点からは放射線の長期被曝の影響として考える場合の発症と死亡の差は小さいとして扱えると思われる。

本研究では、癌を白血病とその他の固形癌とに分けた。白血病は放射線のヒトへの影響として特異性が高いとされていること、および他の癌にくらべて潜伏期間が短いとされていることからシミュレーションでは分けて扱った。その他の癌については単純化のため固形癌としてまとめて全身に均等に被曝すると仮定して扱った。

2. PHM による検出力

シミュレーションからのデータの PHM による解析結果から以下のことが考えられる。

放射線の影響として過剰の癌死亡を検出するのに必要となる標本数は、本研究で仮定したような職業上の被曝を受ける集団を対象とした 40 年間の追跡結果を PHM により解析した場合には、少なくとも数万から十数万人となるであろう。追跡の期間を短くするほど推定の誤差が大きくなり、30 年間の追跡では 40 年間の数倍の誤差が見込まれる。これは、追跡期間の短縮により観測される癌死亡数が減少することから起こるのに加え、被曝から癌死亡までの潜伏期間があるためである。追跡期間が短くとも検出力を大きくするためには過去において職業上の被曝を受けた高齢者を対象とする方法が考えられる。

放射線被曝によるリスクを検出し得る標本数は、本研究ではモデルのパラメータ値とその標準誤差の等しくなる標本数とした。統計的仮説検定の立場から考えると、これは、「放射線による影響はない」という帰無仮説の片側検定において、第 1 種の誤差（帰無仮説が正しいときにこれが棄却される確率）および第 2 種の誤差（帰無仮説が誤っているときにこれを採択する確率）を例えば次のように定めたときの標本数となる²⁴⁾。第 1 種の誤差 16%，第 2 種の誤差 50%，または、第 1 種の誤差 20%，第 2 種の誤差 46%，あるいは、第 1、第 2 種の誤差ともに 31%。放射線被曝の影響については過大評価が原則となっているから、第 1 種の誤差は他の場合よりも大きく定めるべきであろう。

本研究で仮定した放射線の癌死亡に対するリスク係数は、むしろ上限とみなされるのが通常であるが、これよ

りも実際のリスクが大きければ上記より少ない標本数でも検出が可能である。リスク係数をこの 10 倍と仮定した場合には相対誤差のみならず絶対誤差の減少もみられ、数千例の標本でも過剰の癌死亡の検出が可能であろうことが示唆された。したがって、標本数が多くない場合でも、PHM による解析では、現在認められている放射線の癌死亡に対するリスクについての知見が上限として妥当であるか否かについての検討が可能であろう。

3. PHM と MLM の比較と PHM の有効性

PHM と MLM による解析の結果を比較すると、第 0 年次の年齢による癌死亡率の差は PHM の方が、集積線量によるその差は MLM の方が検出されやすかった。追跡年次と集積線量は相関が高く、死亡年次と第 0 年次の年齢との相関はそれにくらべて低いことから、MLM が年次変化による影響を補正する程度が PHM よりも小さいと考えられる。PHM では年次変化についての仮定が少なく、推定すべきパラメータの数も形の上では 2 つ少なくすむ。

MLM では、死亡例だけをデータとして集めれば解析ができ、最終的な集積被曝線量を調査すれば比較的手軽に適用できるという点で多くの用途があろう。

MLM によるリスクの推定に関しての既に著者らの行った研究^{6,25)}では、多変量モデルの導入による見かけの関連の補正の必要性について検討した。すなわち、年次を追っての対象者の加齢による変化や癌死亡が全死亡中に占める割合の変化などが、個人の集積被曝線量とその性格上増加することとの間に見かけの関連をつくり、これをまず補正すべきであるというのが論点であった。今回用いた PHM は年次変化に対して MLM よりも柔軟に対応することができた。

PHM はハザード関数についてその値が説明変数の線形結合の指数に比例することだけを仮定し、特定の関数を定めない⁷⁾。この点でワイブル分布を生存期間の分布に仮定した指数ワイブル・モデル (exponential-Weibull model)²⁶⁾などよりも適用しやすい。

SANDERS⁴⁾ は Hanford の放射線作業者のデータについて平均余命の算出による検討を行い、また、BUNGER²⁷⁾らは職業上の被曝のリスク評価のために生命表分析の利用を提唱している。PHM は死亡に影響する多くの要因についての情報が得られているときには生命表分析による解析よりも有効であろう。

4. シミュレーションのモデルと解析のモデル

本研究でシミュレーションに用いたモデルは BEIR III¹⁴⁾ の p. 169 で示されている、被曝線量と過剰の癌死亡数が比例する絶対リスク・モデル (absolute risk

model) であるが、一方、解析のモデルとして用いた PHM は被曝線量が過剰の癌死亡の全癌死亡に対する割合に比例する相対リスク・モデル (relative risk model) である。MLM もまた相対リスク・モデルに近い。シミュレーションと解析との用いたモデルの違いは解析時に放射線の影響を検出しにくくしていると考えられる。どちらのモデルが実際に起こる現象に近いかは今後の研究の成果を待たなければならないであろうが、相対リスク・モデルの方が事実に近いければ、追跡調査に必要な標本数は本研究で推測したより少なくてもよいことになる。

5. 各種の要因について

本研究でシミュレーションに用いた死亡率は 1975～80 年の日本の男子死亡率の変化を外挿したものである。この外挿は予測としてはきわめて不正確なものであろうが、ここでは単に死亡率が年次を追って変化するひとつのモデルを仮定したもので予測のためのものではない。この変化のしかたで解析結果は異なるであろうが、全体の傾向が同じであれば細かな違いは解析結果を大きく変えることはないと考ええる。

今回用いた被曝線量の指標は線量当量を累積した集積線量である。職業上の被曝はほとんどガンマ線によるものであろうし、この方式は現在広く用いられているものであり、これを用いるのが妥当と考える。しかし、放射線被曝から癌死亡までの潜伏期間の分布が明らかになれば、PHM の説明変数としては線量当量を年次別に重み付きで合計した方が大きな検出力を与えるであろう。

シミュレーションでは被曝時の年齢による影響の違いを考慮しなかった。広島・長崎の原爆被曝者に関する調査でも、白血病の潜伏期間は若年者において短く、70 歳以降の被曝者に出現する白血病は他の年齢層よりも急性型のものが多いという報告²⁹⁾などがある。しかし、本研究ではシミュレーションを単純化するためと、職業上の被曝を受ける年齢層はそれほど幅が広くないと思われるため、被曝時の年齢の影響についてはモデル化しなかった。

本研究では、放射線の集積線量の他には追跡開始時の年齢だけを PHM の説明変数として用いたが、この他にも重要と思われる要因は多いであろう。たとえば、追跡を全対象について同時に開始できるとは限らないことから追跡を開始した暦年や、癌との関連があるといわれる喫煙など、あるいは各個人の職種などのカテゴリー・データも用いることが必要であろう。

6. 線量登録制度の利用について

現在の日本の線量登録制度では²⁹⁾被曝線量を追跡することはできるが疾病の発生や死亡などは追跡できない。

登録制度の成立の経緯からしても離職者の継続的な追跡を事業目的の一環に加えることは無理であろう。しかし、この制度の対象となる人数は発足前の昭和 51 年に 8 万人に達すると推定されている²⁹⁾ことから、今日では 10 万人を上廻るのではと思われる。これだけの多くの人数の記録が統一した内容で保管されていることはいわば宝の山といっても過言ではないであろう。将来、登録されている人々の追跡が企画されたとき、それを実現すべく、企画の目的と内容や企画者の資格などについて今日から検討しておくことを関係者に強く期待したい。その際、被登録者の生死、死因などの基本的な情報は放射線管理関係とは別の行政主体において管理されているため、その方面からの情報を入手するための方途の検討も重要な事項となるであろう。NEWCOMBE³⁰⁾は各国の線量登録制度から、将来の疫学的データとしての利用のための登録項目についての提案を行っている。この項目の中には毎年の全身、局部の被曝線量などの他に、社会保険番号などの識別番号や出生地、父母の名まで含まれている。このような提起も参考となるであろう。

V 結 論

放射線に職業上被曝する作業者についての線量と死亡の追跡調査のシミュレーションを行って得られた疫学的データに比例ハザード・モデル (PHM) を適用して解析した場合の検出力、妥当性を検討し以下の結論を得た。

PHM を 40 年間の追跡調査の 3 とおりの標本数によるシミュレーションからのデータに適用した結果、放射線が過剰の癌死亡と関連があることを検出するのに必要な最小の標本数、すなわち問題とする係数値とその推定の標準誤差が等しくなるような標本数は約 5～15 万人と推定された。

現在推定されている放射線の癌死亡に対するリスク係数より 10 倍高いリスク係数を仮定したシミュレーションからのデータを PHM により解析した結果、係数の推定値の相対誤差だけでなく絶対誤差についても減少がみられ、数千人の標本でリスクの検出が可能であろうと推定された。

シミュレーションからのデータで 30 年までの追跡データが得られた場合について PHM を用いて解析した結果、線量の係数推定値の標準誤差は 40 年の追跡の場合の約 2 倍となり、潜伏期間の存在から、観測される癌死亡数の減少から予期される以上に誤差が大きくなった。

PHM による解析を多重ロジスティック・モデル (MLM) による解析と比較した結果、PHM の方が MLM よりも年次変化の要因についての補正の程度が向上する

ことが明らかとなった。

以上の点から、放射線の長期低線量被曝者の追跡調査からの疫学的データを解析するための多変量モデルとして PHM は妥当なものと考えられ、データが年次ごとに欠くことなく収集されれば、パラメトリックな方法としては PHM は、現在最も適切な手法であろうと結論される。

付 録

比例ハザード・モデル

比例ハザード・モデルは多くの個体の生存に関与する各種の要因の影響をみるために提起されたモデルである^{7,18)}。

定められた観測開始時刻から時間 t の後まで個体が生存している確率を $F(t)$ と表す。 $1-F(t)$ は生存期間の確率分布関数となり、 $-f(t)=d\{1-F(t)\}/dt$ がその密度関数となる。ハザード関数 $\lambda(t)$ は次式で表される。

$$\lambda(t) = -f(t)/F(t) \quad (9)$$

各個体について生存期間に関係すると思われるいくつかの要因があらかじめ観測されているものとする、ハザード関数は時間 t だけでなく観測値のベクトル（説明変数ベクトル） $\mathbf{z}$ によっても規定されるとみなせる。このハザード関数を $\lambda(t; \mathbf{z})$ としてモデルを仮定することができる。Cox⁷⁾ はこのモデルとして時間 t についてのみの関数 $\lambda_0(t)$ と説明変数の線形結合の指数関数の積を提唱した。

$$\lambda(t; \mathbf{z}) = \lambda_0(t) \exp(\mathbf{z}'\boldsymbol{\beta}) \quad (10)$$

ただし $\mathbf{z} = (z_1, z_2, \dots, z_p)'$, $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)'$ 。

(10) 式の $\boldsymbol{\beta}$ はパラメータのベクトルで、 $\lambda_0(t)$ を定めることなく観測値からこれを最尤推定することができる。 β_i の値が正であることはそれに対応する説明変数 z_i の値が大きいほど生存期間の短縮に寄与し、 β_i が負ならば延長に寄与することになる。

個体の集団を観測によって死亡 (failure) の起こった k 個の時刻 $t_1, t_2, \dots, t_k$ を測定したとする。一般には離散時刻で観測が行われ、1つの時刻 t_j において2個体以上の死亡が観測され得る。このときの偏尤度の数値計算には非常に多くの計算を必要とするため実用的でなく、これに代わる近似的な尤度¹⁷⁾も提唱された。この近似尤度の対数 $L(\boldsymbol{\beta})$ は、

$$L(\boldsymbol{\beta}) = \sum_{j=1}^k \left\{ \mathbf{s}_j' \boldsymbol{\beta} - d_j \log \sum_{l \in R(t_j)} \exp(\mathbf{z}_l' \boldsymbol{\beta}) \right\} \quad (11)$$

となる。 $\mathbf{s}_j$ は時刻 t_j で死亡した全個体の $\mathbf{z}$ の値の和、 d_j はその個体数、 $R(t_j)$ は時刻 t_j の直前まで生存していた個体の集合 (risk set)、 $\mathbf{z}_l$ は任意の $R(t_j)$ の要素

l についての $\mathbf{z}$ の値である。(11)式を最大とするパラメータ・ベクトルの最尤推定値 $\hat{\boldsymbol{\beta}}$ を求めるためには、 $\boldsymbol{\beta}$ についての連立方程式

$$\frac{\partial}{\partial \boldsymbol{\beta}} L(\boldsymbol{\beta}) = \sum_{j=1}^k \left\{ \mathbf{s}_j - d_j \mathbf{A}_j \right\} = \mathbf{0} \quad (12)$$

ただし、

$$\mathbf{A}_j = \frac{\sum_{l \in R(t_j)} \mathbf{z}_l \exp(\mathbf{z}_l' \boldsymbol{\beta})}{\sum_{l \in R(t_j)} \exp(\mathbf{z}_l' \boldsymbol{\beta})}$$

を解けばよい。 $L(\boldsymbol{\beta})$ の2次偏微分は

$$\mathbf{I}(\boldsymbol{\beta}) = -\frac{\partial^2}{\partial \boldsymbol{\beta}^2} L(\boldsymbol{\beta}) = \sum_{j=1}^k d_j \mathbf{C}_j$$

ただし、

$$\mathbf{C}_j = \frac{\sum_{l \in R(t_j)} \mathbf{z}_l \mathbf{z}_l' \exp(\mathbf{z}_l' \boldsymbol{\beta})}{\sum_{l \in R(t_j)} \exp(\mathbf{z}_l' \boldsymbol{\beta})} - \mathbf{A}_j \mathbf{A}_j' \quad (13)$$

となる。 $\mathbf{I}(\boldsymbol{\beta})^{-1}$ は $\hat{\boldsymbol{\beta}}$ の分散共分散行列の近似値となる¹⁷⁾。

説明変数は時刻0（観測開始時）において定まっていた固定されている (fixed covariate)¹⁷⁾ 場合がほとんどである。しかし、これは変化することもある。年齢のように時刻0において既に変化が確定しているもの (defined covariate)¹⁷⁾ は固定している場合と同様に扱うことができる。時間とともに確率的に変化する説明変数、例えば放射線の被曝線量などの変量は、ancillary covariate¹⁷⁾ とよばれ、これについても $\mathbf{z}$ の値を時間とともに変化するものとして $\mathbf{z}_l$ と表して同様に扱える。

参 考 文 献

- 1) International Commission on Radiological Protection; "Problems Involved in Developing an Index of Harm," ICRP Publication 27 (1977), Pergamon Press, Oxford.
- 2) T.F. MANCUSO, A.M. STEWART and G.M. KNEALE; Radiation exposure of Hanford workers dying from cancer causes, *Health Phys.*, **33**, 369 (1977).
- 3) A. BRODSKY; A statistical method for testing epidemiological results, *Health Phys.*, **36**, 611 (1979).
- 4) B.S. SANDERS; Low level radiation and cancer deaths, *Health Phys.*, **34**, 521 (1978).
- 5) E.S. GILBERT and S. MARKS; A analysis of the mortality of workers in a nuclear facility, *Radiat Res.*, **79**, 122 (1979).

- 6) Y. HATTORI and K. MAEDA; An application of the logistic function to the assessment of risk in long-term radiation exposure using simulated observations, *Health Phys.*, **41**, 255 (1981).
- 7) D.R. COX; Regression models and life-tables, *J. R. Statist. Soc.*, **B 34**, 197 (1972).
- 8) 厚生統計協会; 国民衛生の動向, 第7表, 厚生 of 指標, **22** (9), 292 (1975).
- 9) 厚生統計協会; 国民衛生の動向, 第7表, 厚生 of 指標, **23** (9), 332 (1976).
- 10) 厚生統計協会; 国民衛生の動向, 第7表, 厚生 of 指標, **24** (9), 344 (1977).
- 11) 厚生統計協会; 国民衛生の動向, 第7表, 厚生 of 指標, **25** (9), 382 (1978).
- 12) 厚生統計協会; 国民衛生の動向, 第7表, 厚生 of 指標, **26** (9), 400 (1979).
- 13) 厚生統計協会; 国民衛生の動向, 第7表, 厚生 of 指標, **27** (9), 406 (1980).
- 14) Committee on the Biological Effects of Ionizing Radiations, National Research Council; "The Effects on Populations of Exposure to Low Levels of Ionizing Radiation : 1980" (1980), National Academy Press, Washington, D.C.
- 15) International Commission on Radiological Protection; "Recommendations of the International Commission on Radiological Protection", ICRP Publication 26 (1977), Pergamon Press, Oxford.
- 16) United Nations Scientific Committee of the Effects of Atomic Radiation; "Source and Effects of Ionizing Radiation, Annex E, III" (1977), U.N., New York.
- 17) C.R. RAO; "Linear Statistical Inference and Its Applications", 2nd Ed., p. 353 (1973), John Wiley & Sons, New York.
- 18) J.D. KALBFLEISCH and R.L. PRENTICE; "The Statistical Analysis of Failure Time Data", p. 70 (1980), John Wiley & Sons, New York.
- 19) 伊理正人; "数値計算", p. 39 (1981) 朝倉書店, 東京.
- 20) D.R. COX; Partial likelihood, *Biometrika*, **62**, 269 (1975).
- 21) S.H. WALKER and D.B. DUNCAN; Estimation of the probability of an event as a function of several independent variables, *Biometrika*, **54**, 167 (1967).
- 22) 宮田久寿, 榎本英寿, 前田和甫; 放射線取扱者および放射線治療患者における白血病: i) 東日本, 日血会誌, **31**, 784 (1968).
- 23) 脇坂行一, 刈米重夫; 放射線取扱者および放射線治療患者における白血病: ii) 西日本, 日血会誌, **31**, 792 (1968).
- 24) 柳井晴夫, 高木廣文; 臨床医学における標本数の定め方, 医学のあゆみ, **102**, 881 (1977).
- 25) 服部芳明, 前田和甫; 低レベル長期放射線被曝の影響の疫学的検討の試み, 保健物理, **15**, 93 (1980).
- 26) A.R. DYER; An analysis of the relationship of systolic blood pressure, serum cholesterol and smoking to 14-year mortality in the Chicago peoples gas company study: I. total mortality in expotential-Weibul model, *J. Chron. Dis.*, **28**, 565 (1975).
- 27) B.M. BUNGER, J.R. COOK and M.K. BARRICK; Life table methodology for evaluating radiation risk: an application based on occupational exposures, *Health Phys.*, **40**, 439 (1981).
- 28) M. ICHIMARU, T. ISHIMARU and J.L. BELSKY; Incidence of leukemia in atomic bomb survivors belonging to a fixed cohort in Hiroshima and Nagasaki, 1950-71: radiation dose, year after exposure, age at exposure, and type of leukemia, *J. Radiat. Res.*, **19**, 262 (1978).
- 29) 星野雅之; 被曝線量登録制度について, 保健物理, **13**, 21 (1978).
- 30) H.B. NEWCOMBE; Design and future use of national dose resisters for regular control and epidemiology, *Health Phys.*, **39** 783 (1980).