Environmental Health Perspectives
Vol. 87, pp. 63–67, 1990

Use of Tumor Diameter to Estimate the Growth Kinetics of Cancer and Sensitivity of Screening Tests

by Naohito Yamaguchi,* Takashi Yanagawa,† Takesumi Yoshimura,‡ Norio Kohrogi,§ Keiji Tanaka,‖ Yuichi Nakamura,‖ and Toshiteru Okubo*

A statistical method has been developed that is useful for studying the relationship between the growth kinetics of malignant tumors and the detection probability either through symptoms or by screening. Mathematical models that describe the distribution of pathological variables in malignant tumors, detected after various histories of screening, are derived and parameters for detection probabilities and the growth kinetics are then estimated by the maximum likelihood procedure. By this method the probabilities of detection through symptoms as well as by screening can be estimated as functions of pathological variable(s) such as tumor size. The growth rate of tumor can also be estimated from the distribution of pathological variables. The present method was applied to gastric cancer in Japan, where an annual screening program for the disease exists. The detection probability for the indirect X-ray used as the screening test was estimated to be $0.323 \times (\text{diameter})^2/[1 + 0.323 \times (\text{diameter})^2]$. The doubling time of gastric cancer was estimated to be 2.90 months.

Introduction

The natural history of cancer consists of two different stages: the stage of transformation and the stage of growth. In the former stage, a normal cell is transformed into a cancer cell by a series of inheritable genomic alterations; in the stage of growth, the cancer cell multiplies and grows to a mass until it becomes large enough to be clinically diagnosed due to symptoms. In this paper we focus on this latter stage and discuss a new method that is useful in describing the nature of growth in this stage.

A malignant tumor in the stage of growth is primarily characterized by the state of its pathological or morphological variables such as size of tumor, invasion of submucosa, and metastasis to lymph nodes or distant organs. These pathological variables determine whether the tumor manifests any clinical symptoms and, therefore, whether it can be detected through symptoms. These variables also determine the probability of tumor detection by a screening test and the prognosis of the individual after the tumor is detected.

The growth kinetics of a malignant tumor, which can be expressed as change in pathological variables with time, is one of the major questions to be answered to better understand the total picture of the process of carcinogenesis. It also provides important information useful for evaluating the potential impact of cancer screening programs in terms of the degree of improvement in prognosis.

In this paper a mathematical method is presented by which the functional relationships between the pathological variables of a tumor and the probability of detection of the tumor, either through symptoms or by a screening test, can be estimated from data of the distribution of pathological variables among cancers with various histories of screening. The growth rate, expressed as the change in the pathological variables per unit of time, can also be estimated by this method. Though the pathological variables are numerous, we choose here the maximum diameter (diameter, hereafter) of tumor as the quantitative measure of the pathological feature. The relationship between the distri-

*Department of Environmental Epidemiology, Institute of Industrial Ecological Sciences, University of Occupational and Environmental Health, Yahatanishiku, Kitakyushu 807, Japan.
†Department of Mathematics, Kyushu University, Higashiku, Fukuoka 812, Japan.
‡Department of Clinical Epidemiology, Institute of Industrial Ecological Sciences, University of Occupational and Environmental Health, Yahatanishiku, Kitakyushu 807, Japan.
§Fukuoka Red Cross Hospital, Minamiku, Fukuoka 815, Japan.
‖Fukuoka City Medical Association, Sagaraku, Fukuoka 814, Japan.
Address reprint requests to N. Yamaguchi, Dept. of Environmental Epidemiology, Institute of Industrial Ecological Sciences, University of Occupational and Environmental Health, Yahatanishiku, Kitakyushu 807, Japan.

bution of diameter and the mode of detection (symptoms versus screening) as well as the history of screening before tumor detection is analyzed based on mathematical models of growth kinetics.

Assumptions

Detection of a Tumor through Symptoms

We assume that the only reason a patient with a malignant tumor seeks medical aid is because of the symptom(s) manifested due to the tumor. We also assume that the magnitude of the symptoms is related to the pathological state of the tumor.

We use the following conditional probability:

$$r(x) = \lim_{dx \to +0} \frac{\Pr\{x \leq L_D < x + dx \mid L_D \geq x\}}{dx} \quad (1)$$

where x is the diameter of the tumor at a certain point in time and L_D is the diameter at detection through symptoms.

Growth of Tumor

We assume that a tumor grows at a rate proportional to the number of cancer cells in the tumor. This implies that the growth of tumor is exponential.

$$x = \alpha \cdot \exp(\beta \cdot t), \quad (2)$$

where t is the time measured from the start of growth, α is the diameter of a single cancer cell, and β is a parameter related to the growth rate. We assume that all tumors have the same value for β.

Incidence

We assume that the incidence rate of the cancer in the population involved is constant over time. Since the incidence rate may differ with age and other variables, this assumption implies that the distribution of these variables in the population does not change with time.

Detection by Screening Test

We assume that a person's decision to participate in the annual screening program is not affected by the existence of a tumor. We also assume that the ability of the screening test concerned in detecting a tumor is related to the pathological state of the tumor. We use a function $v(x)$ for the probability of detection of a tumor with diameter x by the screening test.

Mathematical Formulation

In this section, the probability density functions of diameter of tumors in various situations are derived based on the assumptions previously mentioned. Situations discussed here are: when tumors are detected through symptoms in a population with no history of screening and when tumors are detected by $(k + 1)$'th screen in persons who have k negative results in consecutive screens in the past ($k = 0,1,2,\ldots$).

Let $f_c(x|k)$ denote the probability density function of diameter x in tumors detected through symptoms in persons who have been exposed to k consecutive screens and $f_s(x|k)$ the probability density function of diameter x in tumors detected by $(k + 1)$'th screen in persons who have been exposed to k consecutive screens.

Detection through Symptoms in a Population without Screening

Suppose that we collect cancer cases of people who were diagnosed in a population during a given period $[a, b]$. If this period is long enough to ensure that the probability of a cancer's being included in the study is independent of the age of tumor measured from the start of growth,

$$f_c(x|0) = r(x) \cdot \exp\left[-\int_0^x r(u)du\right]. \quad (3)$$

Thus, $r(x)$ can be estimated by analyzing the distribution of diameter in tumors detected through symptoms in persons who have never been exposed to a screen.

Detection by Screening

Let us consider the distribution of diameter in tumors detected by $(k + 1)$'th screen in a periodically screened population. The probability that a tumor is detected when the diameter falls into a short range $(x, x + dx)$ by a screen after k negative results in consecutive screens with interval d in the past can be expressed as the product of *a*) the probability that the tumor is not detected through symptoms until the diameter reaches x, *b*) the probability that the tumor is not detected by the k screens in the past, and *c*) the probability that the tumor of diameter x is detected by the last screening test. Let $P(x,k)$ denote this probability,

$$P(x,k) = \exp\left[-\int_0^x r(u)du\right] \cdot \prod_{i=1}^{k} [1 - v(x_i)] \cdot v(x), \quad (4)$$

where x_i denotes the diameter of the tumor at $(k-i+1)$'th screen. $x_i = x \cdot \exp(-\beta \cdot d \cdot i)$.

Based on the Bayes Theorem,

$$\begin{aligned} f_s(x|k) &= \Pr\{x \leq L < x + dx | L_D > x, \\ &\quad \text{not detected by } k \text{ screens}, \\ &\quad \text{detected by } (k + 1)\text{'th screen}\} \\ &= \frac{P(x,k) \cdot h(x)}{\int_{L_1}^{L_2} P(x,k) \cdot h(x)dx}, \end{aligned}$$

where L is the diameter of tumor, L_1 and L_2 are the minimum and maximum values of L and $h(x) = \Pr(x \leq L < x + dx)$. If $h(x)$ is uniformly distributed,

$$f_s(x|k) = \frac{P(x,k)}{\int_{L_1}^{L_2} P(x,k)dx}. \quad (5)$$

If $k = 0$, i.e., in the patients with no history of screening,

$$P(x,0) = \exp\left[- \int_0^x r(u)du \right] \cdot v(x).$$

Thus, $P(x,0)$ is independent of the growth rate of tumor, if $h(x)$ is uniformly distributed.

$$f_s(x|0) = \frac{\exp\left[- \int_0^x r(u)du \right] \cdot v(x)}{\int_{L_1}^{L_2} \exp\left[- \int_0^x r(u)du \right] \cdot v(x)dx}. \quad (6)$$

Application

In this section, the above method is applied to gastric cancer in Japan.

Materials and Methods

Table 1 shows the data of the distributions of diameter in gastric cancers of various histories of screening. The cases of gastric cancers detected through symptoms are from the Center for Adult Disease of Fukuoka City Medical Association (CAD-FCMA). They were diagnosed as having gastric cancer during the period 1978 to 1981. The cases of gastric cancers detected by screening are from CAD-FCMA and the Health Center of Karatsu City. They were diagnosed as having gastric cancer in these institutes after indirect X-ray examinations as the screening test during the period 1969 to 1981.

Since $f_c(x|0)$ is related only to $r(x)$ as shown in Eq. (3), $r(x)$ is directly estimable from the observed distribution (column 1 in Table 1). For $r(x)$, we chose polynomials of varying degrees as follows:

$$r(x) = \eta + \sum_{i=1}^{l} \varphi_i \cdot x^i, \quad (l = 1,2, \ldots). \quad (7)$$

Since $f_s(x|0)$ is related only to $r(x)$ as shown in Eq. (6), $v(x)$ is estimable from the observed distribution (column 2 in Table 1) by substituting $\hat{r}(x)$ in Eq. (6). For $v(x)$, we arbitrarily chose the following function:

$$v(x) = \frac{g(x)}{1 + g(x)}, \; g(x) = \sum_{i=1}^{m} \varphi_i \cdot x^i, \; (m = 1,2, \ldots).$$

The growth parameter β is estimable from the observed distribution of diameter of tumors detected by screening among people who have k negative results in consecutive screenings in the past [Eq. (5)] by substituting $\hat{r}(x)$ and $\hat{v}(x)$ for $r(x)$ and $v(x)$. We chose gastric cancers detected by screening after one negative result in a screen 1 year before (column 3 in Table 1) and those detected by screening after two negative results in screens 1 and 2 years before (column 4 in Table 1). A log likelihood function is formulated by combining these two types of data.

The maximum likelihood estimates (MLEs) of parameters are calculated by the Newton-Raphson method, and the best fit models are chosen from among various models based on the Akaike Information Criterion (AIC) (*1*).

Centimeters and months are used as the units for diameter and time, respectively.

Table 1. Distribution of diameter of gastric cancers which were detected either through symptoms or by screening among people with various histories of past screening.

Diameter, cm	Detection group			
	1[a]	2[b]	3[c]	4[d]
– 1	9	5	4	4
1 – 2	18	20	7	10
2 – 3	11	39	12	9
3 – 4	11	33	12	6
4 – 5	17	37	5	4
5 – 6	16	26	4	2
6 – 7	13	13	4	2
7 – 8	9	9	1	0
8 – 9	5	7	0	0
9 – 10	4	4	1	1
10 –	10	12	0	3
Total	123	205	50	41

[a] Detected through symptoms among people with no history of past screening.
[b] Detected by screening among people with no history of past screening.
[c] Detected by screening among people with history of one negative result in the screening of 1 year before.
[d] Detected by screening among people with history of two negative results in the past screenings of 1 and 2 years before.

Results

***Estimation of* r(x).** The AICs for polynomials of degree 1 and 2 were 599.56 and 602.32, respectively. Therefore, we chose the polynomial of degree one as the best fit model for $r(x)$. The MLEs and estimated covariance matrix of parameters in the polynomial are shown in Table 2.

The 95% confidence intervals of η and φ_1 are (0.0173, 0.1113) and (0.0230, 0.0546), respectively. Figure 1 illustrates the observed distribution function and the distribution function calculated by $\hat{r}(x)$.

***Estimation of* v(x).** To calculate the MLEs of the parameters in $v(x)$, we first estimated the true distri-

Table 2. MLEs and estimated covariance matrix of parameters.

	MLEs	Covariance matrix	
		$\hat{\eta}$	$\hat{\varphi}_1$
$\hat{\eta}$	0.0643	5.77×10^{-4}	-1.45×10^{-4}
$\hat{\varphi}_1$	0.0388	-1.45×10^{-4}	6.48×10^{-4}
$\hat{r}(x) = \hat{\eta} + \hat{\varphi}_1 x$			

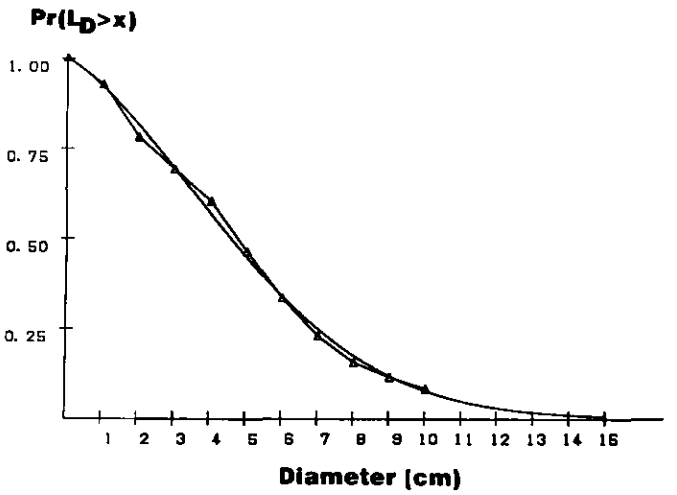

FIGURE 1. Observed (Δ—Δ) and fitted (—) distribution function of diameter of gastric cancers detected through symptoms. L_D denotes the diameter at detection through symptoms.

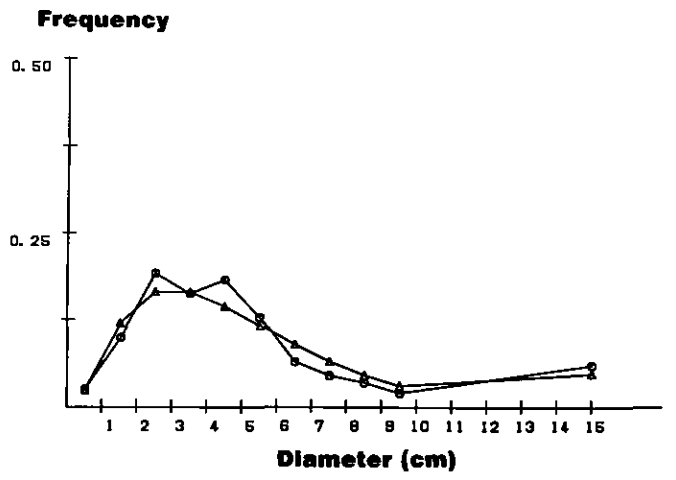

FIGURE 2. Observed (O—O) and fitted (Δ—Δ) distribution of diameter of gastric cancers detected by the first screening.

Table 3. AICs for various models of $v(x)$.

Models [$g(x)$]	AIC
$\delta_1 x$	908.28
$\delta_2 x^2$	899.96
$\delta_3 x^3$	902.29
$\delta_1 x + \delta_2 x^2$	901.94
$\delta_1 x + \delta_2 x^2 + \delta_3 x^3$	901.58

bution of diameter of preclinical tumors that are not yet detected at a certain point in time by using Eq. (6) and then estimating $v(x)$ based on the fact that the difference between the true and the observed relative frequencies can be explained by the function $v(x)$. The AICs for various models of $v(x)$ were as shown in Table 3.

So we chose the function of power 2 (model 2) as the best fit function. The MLE and SE of δ_2 were 0.323 and 0.110, respectively. Figure 2 illustrates the observed distribution function of diameter in gastric cancers detected by first screenings and the distribution function

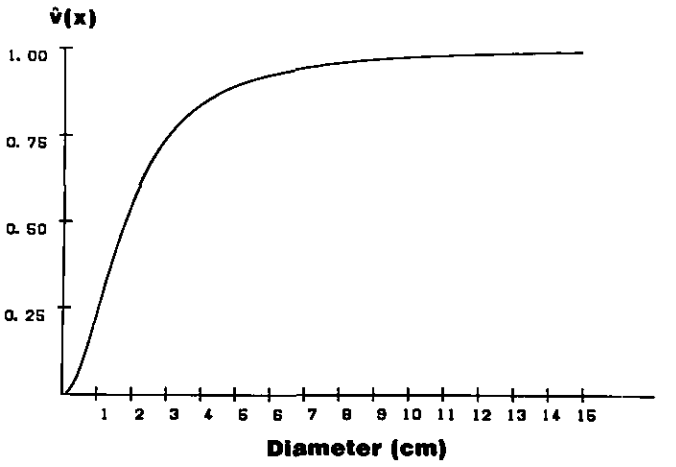

FIGURE 3. Probability of detection of gastric cancer by the indirect X-ray examination as a screening test.

calculated based on the estimated $r(x)$ and $v(x)$. Figure 3 illustrates the estimated $v(x)$.

Estimation of Growth Parameter. Finally, the MLE for β was obtained by substituting $\hat{r}(x)$ and $\hat{v}(x)$ for $r(x)$ and $v(x)$ in Eq. [5] and by maximizing the combined log likelihood of the data of diameter of gastric cancers detected by screening with one and two negative results in the past. The MLE and SE are 0.239 and 0.158, respectively. Based on the MLE, the doubling time of diameter in gastric cancer was estimated to be 2.90 months. By assuming that the diameter of a single cancer cell is 0.01 mm, the length of time from the start of growth to the point at which the diameter reaches 5 cm was estimated to be 3.0 years.

Discussion

Mathematical theories concerning the natural history of cancer and early detection by screening have been proposed by many researchers; one of the earliest studies was published by Zelen and Feinleib (*2*) as early as 1969. One of the finest mathematical theories is given in a series of papers by Albert et al. (*3*), Albert et al. (*4*), and Louis et al. (*5*). They formulated the natural history of chronic disease state in terms of the distribution of the age of a person at the time of entering the disease state, the sojourn time in that disease state, and the person's present age. The formulation was used by many authors for various studies. However, in the case of cancer, it is difficult to observe the point in time at which a cancer enters the preclinical state and even more difficult to define the state of preclinical state itself in a fashion that is useful for biostatistic research.

In the present study, we took into account the fact that the primary determinant of the distribution of sojourn time was the pathological variables of the growing tumor, and based on this fact, we developed a mathematical model by which the change in the pathological variables over time can be estimated. As noted above,

the pathological feature of a tumor determines not only whether it manifests any clinical symptoms but also whether it can be detected by a screening test and even determines the prognosis or curability of the patient with the tumor. Thus, pathological variables seem more useful in describing the natural history of cancer than the time measure itself.

Also, the present method enables us to estimate the probability of detection by screening test as a function of the pathological variables, in contrast to the models introduced so far by other researchers (*6,7*), in which they assume that the probability has the same value throughout the entire preclinical period. The ability of a screening test to detect a cancer of a certain size cannot be measured directly from screening data, since false negative cases are not accessible. The method used to calculate the false negative rate is usually to review the past X-ray films of individuals in whom cancers were later found. However, this method may give biased results, since the probability of detection by screening is related not only to the ability of the test used but also to the pathological features of the tumor, and this method cannot take into account the latter factor.

In the present study, we assumed that all cancers of a certain site have the same growth rate. This assumption may be replaced by a weaker assumption that the growth rate differs in individual cases following a certain distribution family such as log normal distribution. Although the results are not presented in this paper, according to the simulation we conducted to compare the fixed growth rate model with the random growth rate model with β, log normally distributed showed good agreement for various values of β and its standard deviation.

The data to which we applied our method were not obtained from a population to which a screening program was randomly assigned. If the method is applied to data obtained through a randomized controlled trial, the validity of results will be increased. However, it should be noted that an analysis of existing data based on the method presented in this study can provide important information on the natural history of cancer fairly efficiently if the assumptions made on the data and model are plausible from a biological viewpoint.

REFERENCES

1. Akaike, H. Information theory and an extension of maximum likelihood principle. In: 2nd International Symposium on Information Theory (H. Petrov and F. Csaki, Eds.), Akademia Kiado, Budapest, 1973.
2. Zelen, M. and Feinleib, M. On the theory of screening for chronic diseases. Biometrika 56: 601–614 (1969).
3. Albert, A., Gertman, P. M., and Louis, T. A. Screening for the early detection of cancer, I. The temporal natural history of a progressive disease state. Math. Biosci. 40: 1–59 (1978).
4. Albert, A., Gertman, P. M., Louis, T. A., and Liu, S-I. Screening for the early detection of cancer. II. The impact of screening on the natural history of the disease. Math. Biosci. 40: 61–109 (1978).
5. Louis, T. A., Albert, A., and Heghinian, S. Screening for the early detection of cancer. III. Estimation of disease natural history. Math. Biosci. 40: 111–144 (1978).
6. Walter, S. D. and Day, N. E. Estimation of the duration of a pre-clinical disease state using screening data. Am. J. Epidemiol. 118: 865–886 (1983).
7. Flehinger, B. J., and Kimmel, M. The natural history of lung cancer in periodically screened population. Biometrics 43: 127–144 (1987).