Jpn. J. Cancer Res. 81, 1109–1117, November 1990

Analysis of Stomach Cancer Incidence by Histologic Subtypes Based on a Mathematical Model of Multistage Cancer Induction and Exponential Growth

Naohito Yamaguchi,[1,4] Shaw Watanabe,[2] Keiichi Maruyama[3] and Toshiteru Okubo[1]

[1]Department of Environmental Epidemiology, Institute of Industrial Ecological Sciences, University of Occupational and Environmental Health, 1-1, Iseigaoka, Yahatanishi-ku, Kitakyushu 807, [2]Division of Epidemiology, National Cancer Center Research Institute and [3]Division of Surgery, National Cancer Center Hospital, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104

A mathematical model incorporating the processes of both cancer induction and subsequent tumor growth has been developed. The model was applied to incidence data of stomach cancer classified into histologic subtypes: papillary adenocarcinoma (PAP), well and moderately differentiated tubular adenocarcinomas (WEL and MOD), poorly differentiated adenocarcinoma (POR), mucinous adenocarcinoma (MUC) and signet-ring cell carcinoma (SIG). The multistage theory was assumed for cancer induction as in the Armitage-Doll model. For the period of growth, exponential growth was assumed and clinical surfacing was formulated as a stochastic process related to tumor diameter. The number of stages in cancer induction and the tumor growth rate were simultaneously estimated for each histologic subtype using the maximum likelihood procedure. The present model showed better fits than the Armitage-Doll model in most histologic subtypes except WEL. PAP, WEL and MOD, which are characterized as differentiated subtypes with less mucous production, showed different features from POR, MUC and SIG: 1) the number of stages was estimated to be larger, 2) the differences in incidence rates between males and females were more marked, and 3) males tended to have larger growth rates in PAP and MOD, while in POR, MUC and SIG, females had larger values. The present study showed that an analysis by histologic subtypes is of importance in stomach cancer and that the period of tumor growth should not be ignored when formulating a model of the natural history of stomach cancer.

Key words: Stomach cancer — Mathematical model — Multistage theory — Tumor growth — Tumor size

In the light of recent advances in cancer biology, integration of information obtained in the fields of basic laboratory sciences, clinical oncology and epidemiology is becoming one of the main subjects of interest in cancer research. The mathematical modeling approach seems promising with regard to this problem. It can play a substantial role in building a bridge between observations in the laboratory and those in epidemiology, by formulating quantitative relationships in complicated chains of phenomena in carcinogenesis.

The mathematical approach in cancer research has a long history. Among the landmarks are the pioneering work by Nordling[1)] and its refinement by Armitage and Doll.[2)] The model put forth by these researchers was a mathematical formulation of the multistage theory of carcinogenesis. They showed theoretically that the incidence or death rate of cancer increases in proportion to the $(k-1)$th power of age, if the cancer is induced through a series of k steps of carcinogenesis. Subsequently, it was found that this power relationship was observed in various cancers in different countries.[3)]

Although the so-called Armitage-Doll model mentioned above has continued to provide us with a good conceptual framework to interpret the observations in epidemiology,[4)] it should also be noted that the model is based on a number of assumptions that are too simplistic and unreal to apply to observations in the laboratory. These include the assumption that the time period between the start of tumor growth and the observation of incidence or death is short enough to ignore in the mathematical formulation.[5)]

It should also be mentioned that we have limited information on the fit of the model with regard to various histological subtypes of cancer. It seems biologically more plausible to assume that the process of carcinogenesis could differ by histologic subtype even if various subtypes occur within the same organ.

In this paper, a new mathematical model is proposed, in which the period of cancer growth is taken into account in addition to the period of cancer induction. The model is applied to various histologic subtypes of stomach cancer and the number of stages in cancer induction and the growth rate of tumor mass are simultaneously estimated for each histologic subtype.

[4] To whom correspondence should be addressed.

MATERIALS AND METHODS

Database The incidence rate of stomach cancer of a certain histologic subtype must be estimated from different databases, since the incidence rate is usually reported only for stomach cancer as a whole. In this study, the relative frequency of a histologic subtype was estimated for each sex and 5-year age class from 35 to 79 years of age, based on data from 2,505 male and 1,250 female cases of stomach cancer diagnosed at the National Cancer Center (NCC) Hospital in Tokyo during the period 1962 to 1983. The cases under 35 and over 79 years of age were excluded because the numbers were too small to obtain reliable estimates for relative frequencies of histologic subtypes. The numbers of cases in 5-year age classes ranged from 105 to 442 cases in males and from 85 to 186 cases in females. As the histological subtypes of stomach cancer, papillary adenocarcinoma (PAP), well differentiated tubular adenocarcinoma (WEL), moderately differentiated tubular adenocarcinoma (MOD), poorly differentiated adenocarcinoma (POR), mucinous adenocarcinoma (MUC) and signet-ring cell carcinoma (SIG) were considered.[6] The 5-year age-specific incidence rates of each subtype were then estimated by portioning the incidence rates of total stomach cancer based on the relative frequency of that age group. The incidence rates of stomach cancer as a whole come from the results of the Second Cooperative Study of 6 Population-based Cancer Registries for Estimating Cancer Incidence in Japan conducted from 1971 to 1972.[7] The survey covered a population of 14.7 million (7.3 million males and 7.4 million females). The distribution of tumor diameter in each histologic subtype was also estimated from the data of the NCC Hospital. The tumor diameter of each cancer was measured when the tumor was surgically removed. Inoperable cases were therefore excluded from the database. The results were used to estimate the parameters in the process of clinical surfacing.

Model The entire process of carcinogenesis was divided into two distinct processes in the model: the period of cancer induction and the period of tumor growth. In the period of cancer induction, a normal cell undergoes inheritable changes to become a cancer cell. The period of tumor growth starts immediately following the period of cancer induction and continues until the tumor surfaces clinically when it has become large enough to be detected. Hence, the distribution of the time period of tumor growth is determined by the process of clinical surfacing, which can be considered as a probabilistic phenomenon. Letting $h(a)$ denote the hazard rate for induction of a cancer cell at age a and $V(t)$ the probability of a tumor's not being detected during the period of t measured from the start of tumor growth, the joint probability of a cancer cell's being induced at age $a-t$ and the tumor's being detected as a clinical cancer at age a can be formulated as follows:

$$g(a, t)=h(a-t)\left[-\frac{dV(t)}{dt}\right].$$

The probability density function of the incidence at age a can then be obtained by integrating $g(a, t)$ over t:

$$z(a)=\int_0^a h(a-t)\left[-\frac{dV(t)}{dt}\right]dt. \qquad (1)$$

In formulating equation (1), the cancer induction is assumed to be a rare event, so that the probability density function for cancer induction can be approximated by the hazard function $h(a)$.

As in the Armitage-Doll multistage model, it is assumed that a normal cell must undergo a certain number of inheritable changes to become a cancer cell and that the transition rate in each change is constant over time.[5] Hence, the hazard rate $h(a)$ can be formulated as follows[5]:

$$h(a)=ra^w, \qquad (2)$$

where w denotes the number of stages in the process of cancer induction minus one and r a constant. w, which will be called the power parameter in this paper, is treated as a continuous quantity.

A functional form for clinical surfacing $V(t)$ is not known *a priori*. In the present study, the process of clinical surfacing was first modeled as a function of tumor diameter x and then the function $V(t)$ was obtained by substituting the time period of tumor growth t for the tumor diameter x. Exponential growth of the tumor mass was assumed in relating the tumor diameter x with the growth period t. The following conditional probability was chosen to model the process of clinical surfacing:

$$g(x)=\lim_{dx->+0}\frac{\Pr[X<x+dx;\ X\geq x]}{dx}, \qquad (3)$$

where X is a random variable for the tumor diameter at clinical detection (at surgical operation, strictly speaking). The function $g(x)$ can be interpreted as a rate of clinical surfacing while the tumor grows from x to $x+1$, given that the tumor was not detected until the diameter reached x. The exponential increase in tumor diameter with time is formulated as follows:

$$x=f(t)=c\exp(dt), \qquad (4)$$

where c is the diameter of a single cancer cell, d a growth parameter, and t the growth period until the tumor diameter reaches x. Since the $V(t)$ is equal to the probability of a tumor's not being detected until the diameter

reaches x, it can be formulated as a function of the time t by combining equations (3) and (4):

$$V(t)=\Pr[X>x]=\Pr[x>f(t)]$$
$$=\exp\left[-\int_c^{f(t)} g(u)du\right]. \quad (5)$$

The model expressed by equations (1) to(5) is called the MIEG (multistage induction and exponential growth) model in this paper.

Estimation procedure The maximum likelihood (ML) estimation procedure was used to estimate parameters in various models and the best-fit model for which the AIC (Akaike Information Criterion)[8] score shows the minimum value was selected. An AIC score for a model is calculated by the following formula:

$$\text{AIC}=2(\#\text{parameters in model})-2(\log \text{likelihood}).$$

As the first step, the function $g(x)$ was estimated from the distribution of tumor diameter. The distribution of diameter of clinical cancers detected at age a can be expressed by the following equation, based on equation (1):

$$\Pr[X>x_1]=\frac{\int_{t_1}^{a} h(a-t)\left[-\frac{dV(t)}{dt}\right]dt}{\int_0^a h(a-t)\left[-\frac{dV(t)}{dt}\right]dt}, \quad (6)$$

where t_1 is the growth period for a tumor to reach a diameter x_1 [$x_1=f(t_1)$]. The function $h(a-t)$ can be cancelled out of equation (6), based on the fact that the distribution of tumor diameter is not affected by age at detection.[9] Based on the facts that $V(a)\fallingdotseq 0$ and that $V(0)=1$, equation (6) can be further reduced to the following form:

$$\Pr[X>x_1]=\frac{V(t_1)-V(a)}{V(0)-V(a)}\fallingdotseq(Vt_1)$$
$$=\exp\left[-\int_c^{x_1} g(u)du\right].$$

Thus, $g(x)$ is directly estimable from the distribution of tumor diameter in clincal cancers. As a functional form of $g(x)$, the following function with unknown parameters b_1 and b_2 was assumed on an empirical basis:

$$g(x)=b_1x+b_2x^2.$$

The quadratic term, b_2x^2, was included in the model only when the AIC score became smaller if included. Similarly, separate models for males and females were selected only when the AIC score became larger for the model with a common parameter for males and females.

As the second step, the power and growth parameters in the MIEG model [w in equation (2) and d in equation (4)] were simultaneously estimated by fitting the MIEG model to the incidence data. The conditional likelihood was constructed to cancel out the nuisance parameter r. The parameter c (diameter of a single cancer cell) is set to 0.001 cm. The Armitage-Doll model, which can be considered as a special case of the MIEG model with a very large tumor growth rate, was also fitted to the incidence data.

RESULTS

The ML estimates of parameters in the models of clinical surfacing are shown in Table I. The observed distributions of tumor diameter are compared in Fig. 1 with the expected curves calculated from the ML estimates to check visually the goodness of fit. Separate models for males and females showed better fit in PAP, POR and SIG. Inclusion of the quadratic term improved the fit in histologic subtypes except female PAP, MOD and MUC. WEL showed the largest skew to the left, indicating higher detection probability while the tumor remained small.

Table I. Maximum Likelihood Estimates of Parameters in the Models for Clinical Surfacing of Stomach Cancer

Histologic subtypes[a]		Male[b]	Female[b]	Combined[c]
PAP	b_1[d]	0.030	0.050	–
	b_2[d]	0.002	–	–
WEL	b_1	–	–	0.107
	b_2	–	–	−0.005
MOD	b_1	–	–	0.050
	b_2	–	–	–
POR	b_1	0.035	0.031	–
	b_2	−0.001	−0.001	–
MUC	b_1	–	–	0.027
	b_2	–	–	–
SIG	b_1	0.050	0.035	–
	b_2	−0.002	−0.001	–

a) PAP=papillary adenocarcinoma, WEL=well differentiated tubular adenocarcinoma, MOD=moderately differentiated tubular adenocarcinoma, POR=poorly differentiated adenocarcinoma, MUC=mucinous adenocarcinoma, SIG=signet-ring cell carcinoma.

b) The separate models in which parameters were estimated separately for males and females showed better fit than the combined model.

c) The combined model which assumed that males and females had the same parameter values showed better fit than the separate models.

d) Parameters in the model of clinical surfacing:
$\lim_{dx\to+0}\Pr[X<x+dx;\ X\geq x]/dx=b_1x+b_2x^2$,
where X denotes the tumor diameter.

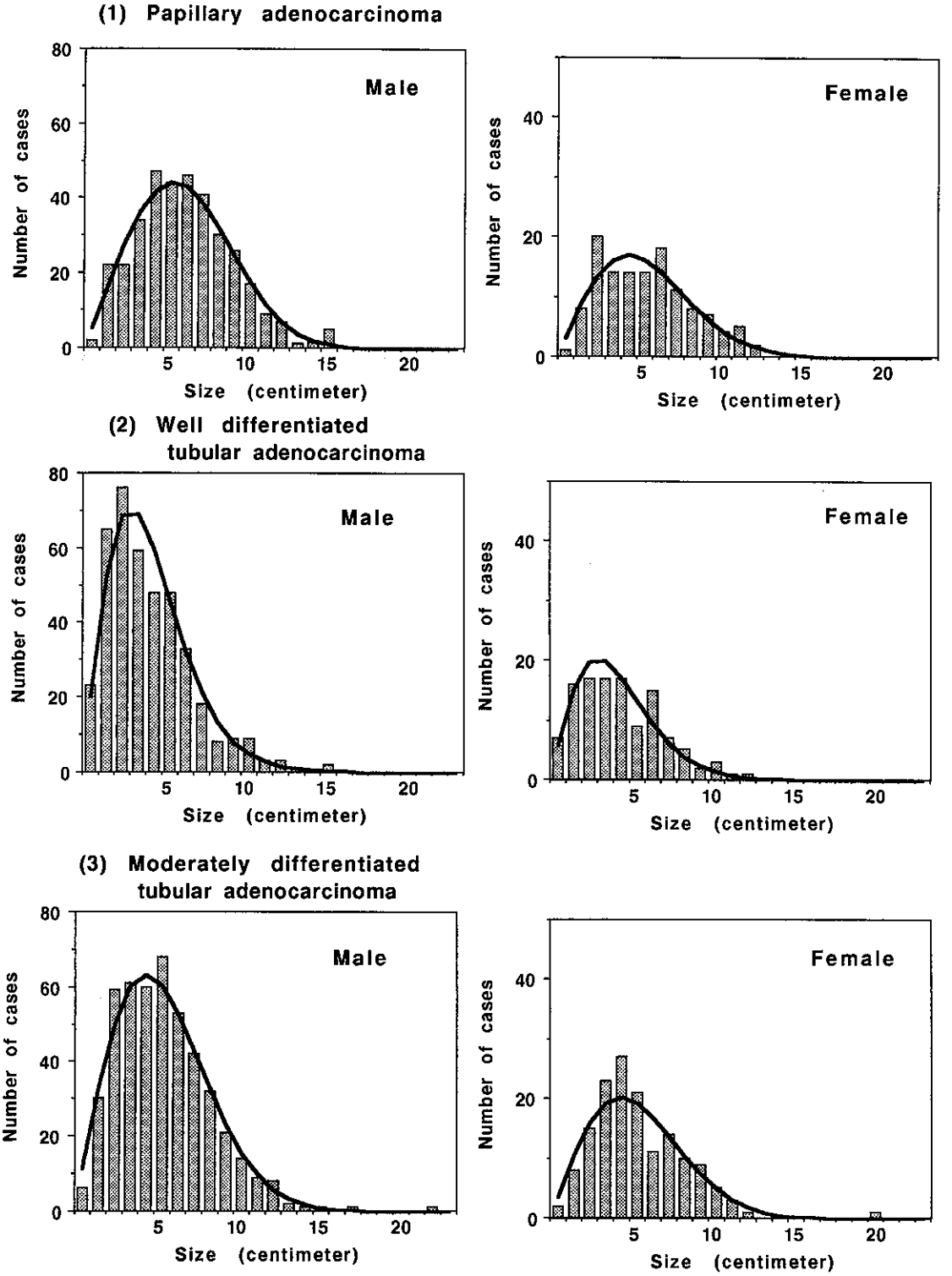
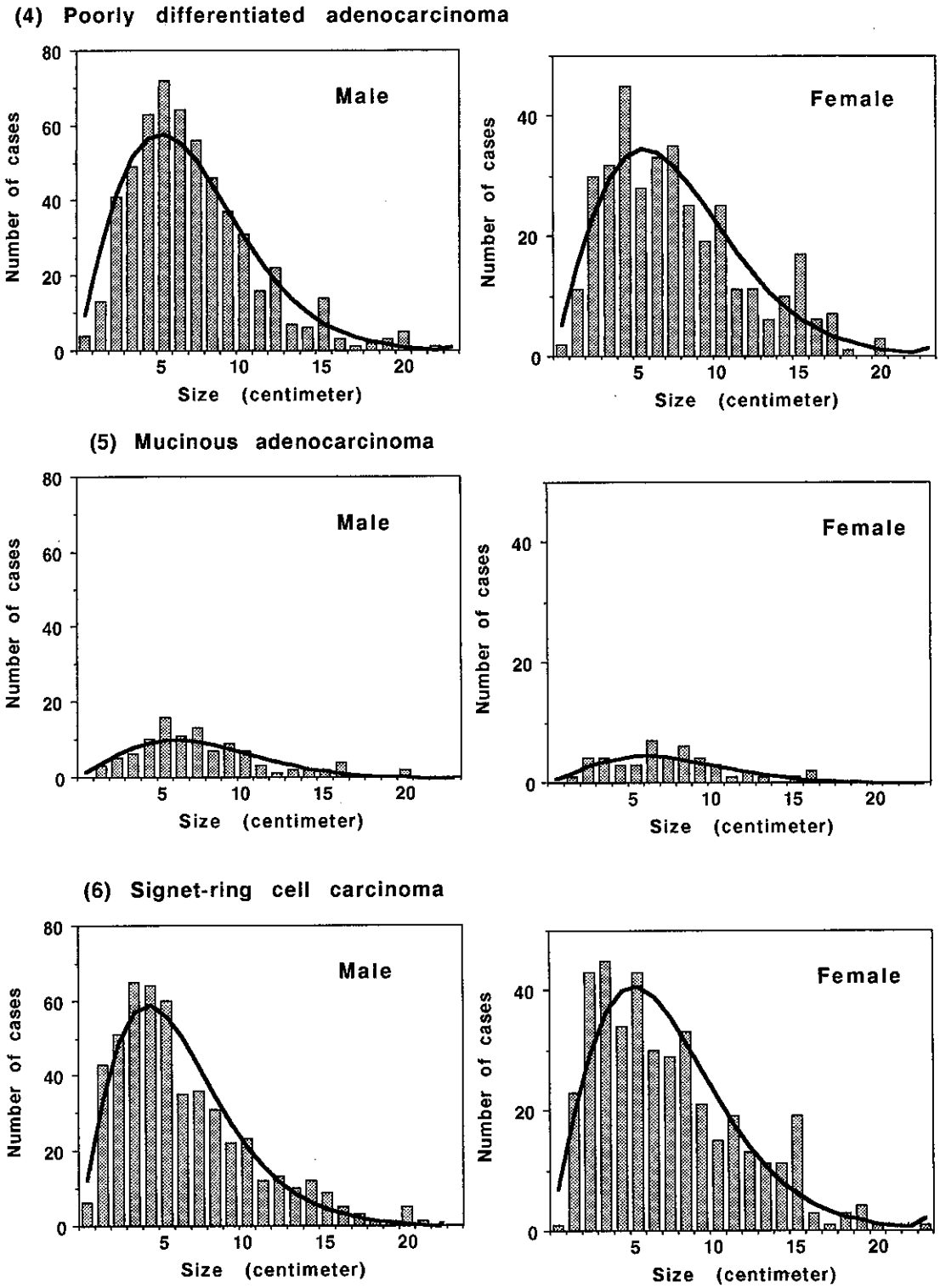


Fig. 1. Observed (bars) and fitted (solid lines) distributions



of tumor diameter in different histologic subtypes.

Table II. Maximum Likelihood Estimates of the Power and Growth Parameters in the MIEG Model

Model[a]		Sex	Power[b]	Growth[c]	AIC[d]
PAP	1	m	5.47	1.243	6973.4
		f	3.96	0.383	
	2	m	5.04	0.830	6973.9
		f		0.635	
	3	m	6.28		6977.9
		f	6.53		
	4	m+f	6.35		6976.7
WEL	1	m	4.86	0.718	7377.1
		f	5.50	2.075	
	2	m	5.24	0.992	7375.7
		f		1.318	
	3	m	6.17		7376.3
		f	5.96		
	4	m+f	6.12		7374.8
MOD	1	m	3.53	0.484	8647.9
		f	2.82	0.294	
	2	m	3.36	0.447	8647.6
		f		0.348	
	3	m	5.37		8678.3
		f	5.98		
	4	m+f	5.43		8682.4
POR	1	m	3.07	0.641	13897.0
		f	3.15	1.522	
	2	m	3.07	0.641	13895.0
		f		1.299	
	3	m	4.30		13907.6
		f	3.60		
	4	m+f	3.99		13929.8
MUC	1	m	1.71	0.301	2170.6
		f	1.76	0.488	
	2	m	1.72	0.302	2168.6
		f		0.477	
	3	m	4.43		2190.0
		f	3.06		
	4	m+f	3.91		2202.7
SIG	1	m	1.48	0.420	13498.2
		f	1.23	0.527	
	2	m	1.39	0.403	13498.8
		f		0.606	
	3	m	2.93		13564.2
		f	2.10		
	4	m+f	2.52		13611.5

a) Model 1: MIEG model with different power and growth parameters for males and females. Model 2: MIEG model with a common power parameter and different growth parameters for males and females. Model 3: Armitage-Doll model with different power parameters for males and females. Model 4: Armitage-Doll model with a common power parameter for males and females.
b) Estimates of power parameter in the model of cancer induction.
c) Estimates of growth parameter in the model of tumor growth.
d) AIC=2(#parameter in the model)−2(log likelihood).

PAP=papillary adenocarcinoma, WEL=well differentiated tubular adenocarcinoma, MOD=moderately differentiated tubular adenocarcinoma, POR=poorly differentiated adenocarcinoma, MUC=mucinous adenocarcinoma, SIG=signet-ring cell carcinoma.

The ML estimates of parameters in the MIEG model are shown in Table II. Model 1 (different power and growth parameters for males and females) was selected as the best-fit model for PAP and SIG, but the differences in AIC's between models 1 and 2 were marginal in both cases. Model 2 (common power parameter for males and females) was selected for MOD, POR and MUC. For WEL, model 4 (Armitage-Doll model with a common power parameter for males and females) was found to be the best-fit model.

When the ML estimates of the power parameter in model 1 were compared among the subtypes excluding WEL, PAP had the largest values among both males and females, followed by MOD, POR, MUC and SIG in males and by POR, MOD, MUC and SIG in females. The differences between males and females were not statistically significant at the 5% level for any subtype, when compared by likelihood ratio statistics.

The ML estimates of the growth parameter showed the largest value for PAP in males and for POR in females. It had the smallest value for MUC in males and for MOD in females. Females had larger growth rates in POR, MUC and SIG than males. When the differences between males and females were tested by likelihood ratio statistics in the subtypes excluding WEL, all differences were found to be statistically significant at the 5% level.

The age-incidence relationships in different subtypes are illustrated in Fig. 2. The expected curves calculated based on the best-fit models are also illustrated. The differences between males and females were larger at all ages for PAP, WEL and MOD, while they became less clear for POR, MUC and SIG, especially at the younger ages.

DISCUSSION

In the present study, the differences among histologic subtypes were highlighted by both the power and growth parameters in the model and the male-to-female ratio of incidence rates. As shown in Table II, PAP, WEL and MOD, which can be characterized as differentiated subtypes with less mucous production, tended to have a larger power, when compared with POR, MUC and SIG. This indicates that a normal cell must undergo a larger number of changes to become a cancer cell. PAP, WEL and MOD are also characterised by the larger male-to-female ratio of incidence rates as shown in Fig. 2. These facts suggest that, in SIG, MUC and POR, risk factors are operating in earlier life at the same magnitude for males and females, while in PAP, WEL and MOD, risk factors operate at larger magnitudes for males at all ages. It is possible to further postulate that some host factors such as genetic predisposition play a role in SIG, MUC and POR, while environmental factors such as lifestyle

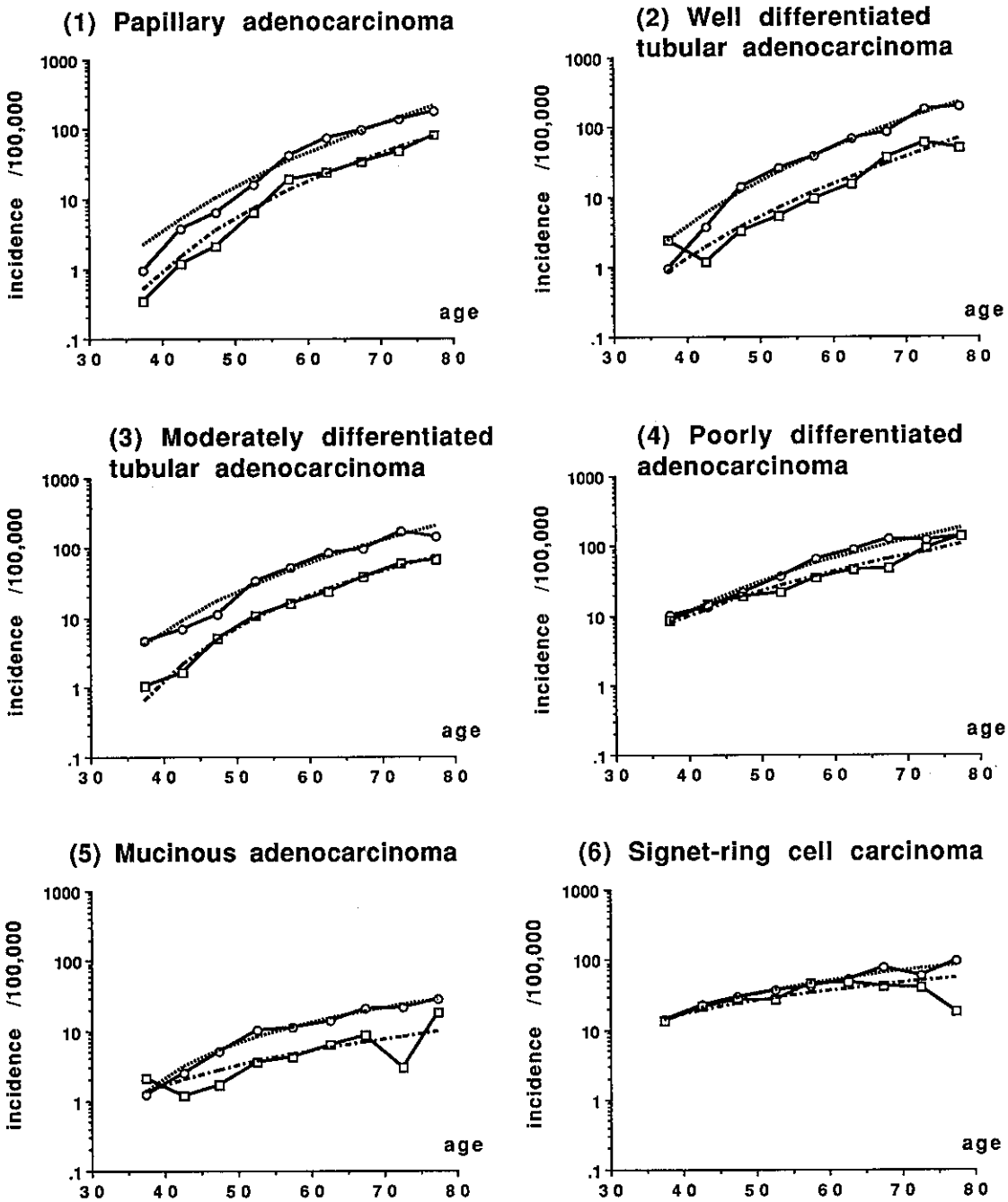


Fig. 2. Observed and fitted age-incidence curves for different histologic subtypes. Male observed (—○—), female observed (—□—), male fitted (·····) and female fitted (–•–•–•).

factors play a role in PAP, WEL and MOD. To examine this hypothesis, further epidemiological studies in which stomach cancer is classified by the histologic subtypes are needed.

Compared to the Armitage-Doll model, the MIEG model proposed in this paper showed better fit in all subtypes except WEL. This indicates that the process of growth should not be ignored when estimating the power

parameter in these subtypes. There are several differences to be pointed out between the MIEG model and the Armitage-Doll model. Firstly, it should be noted that the estimated power always has a smaller value in the MIEG model than in the Armitage-Doll model. Thus, the number of stages is overestimated if the Armitage-Doll model is applied to these subtypes. Secondly, the difference in the power parameter between males and females is not statistically significant in the MIEG model for all subtypes, while it is significant in the Armitage-Doll model for MOD, POR, MUC and SIG. It can therefore be inferred that the difference in the power parameter of males and females in the Armitage-Doll model is attributable to the difference in the growth parameter in these cases. The Armitage-Doll model showed a better fit in WEL; this can be explained by the larger growth rate as well as by the tendency for WEL to be clinically detected while the diameter is still small (Fig. 1).

The estimates of the growth parameter showed the largest value of 1.299 in female POR and the smallest value of 0.302 in male MUC. These values correspond to diameter doubling times of 6.4 and 27.5 months, respectively, and are in agreement with those reported in other studies.[10-12] In the subtypes excluding WEL, for which the Armitage-Doll model showed a better fit, there is no clear difference between the group of PAP and MOD and the others, which showed marked contrasts in the power parameter and the male-to-female ratio of incidence. However, when we focus on the difference between males and females, there is a contrast between the two subgroups. Males have larger growth rates in PAP and MOD, while females have larger growth rates in the other types. The hormonal environment may be a factor, as in breast cancer or uterine cancer[13] but further studies are needed to investigate the reasons for this difference.

Since the results in the present study were obtained from clinical data in a single hospital, the validity of the study relies on whether the cases in the hospital represent those in the whole country. The timing of diagnosis might be different from cases in other hospitals, since early cancers at a curable stage might accumulate in the NCC Hospital. Also, the distribution of histologic subtypes could differ for similar reasons. The distribution of tumor diameter showed a good agreement with 29,275 cases from 178 hospitals throughout Japan reported by the Japanese Research Society for Gastric Cancer[14]; the proportion of stomach cancer less than 5 cm was 44% in the NCC Hospital in contrast to 41% in the nationwide registry. As to the proportion of histologic subtypes, MOD and SIG showed larger differences between the NCC Hospital and nationwide registry; 17% and 25% for MOD and 21% and 11% for SIG, respectively. This indicates that access to a larger database covering wider regions is essential for further research.

The validity of the present study also depends on the assumptions introduced to formulate the MIEG model for the analysis. Of particular importance are the assumptions on the tumor growth. It was assumed in the present study that tumor growth is exponential and that the growth rate is the same for all tumors of a certain histologic subtype. These assumptions were not made on the basis of biological plausibility but on the basis of mathematical convenience. The number of unknown parameters must be as small as possible to make statistical inferences from limited epidemiological data. An analysis of stomach cancer cases with two or more consecutive observations is essential to develop better models for tumor growth.

The present study shows that analysis of cancer incidence by histologic subtypes provides us with further insights into the etiology and natural history of stomach cancer. Thus far, most epidemiologic studies on the risk factors of cancer have been conducted based on classification by organs, which may obscure underlying etiologic factors by mixing histologic subtypes which had different etiological backgrounds. Consideration of histological classification is thus of importance in the mathematical approach to cancer. However, to make further progress, several problems must be solved. First, it is essential to develop cancer databases which include information on histopathology as well as other potentially causative factors to allow more detailed analyses by mathematical modeling.[15] Second, scientific communication between laboratory scientists, epidemiologists and those specializing in mathematical modeling must be stimulated. It is necessary to test and reformulate mathematical models in order to enhance current progress in cancer research. Continued efforts to improve mathematical models will play an important role in achieving a better understanding of cancer etiology.

(Received June 20, 1990/Accepted August 30, 1990)

REFERENCES

1) Nordling, C. O. A new theory of the cancer-producing mechanism. *Br. J. Cancer,* **7**, 68–72 (1953).
2) Armitage, P. and Doll, R. The age distribution of cancer and a multi-stage model of carcinogenesis. *Br. J. Cancer,* **8**, 1–12 (1954).
3) Cook, P., Doll, R. and Fellingham, S. A. A mathematical model for the age distribution of cancer in man. *Int. J. Cancer,* **4**, 93–112 (1969).

4) Breslow, N. E. and Day, N. E. "Statistical Methods in Cancer Research, II. The Design and Analysis of Cohort Studies," IARC Scientific Publications No. 82, pp. 256–262 (1987). IARC, Lyon.
5) Whittemore, A. The age distribution of human cancer for carcinogenic exposures of varying intensity. *Am. J. Epidemiol.*, **106**, 418–432 (1977).
6) Japanese Research Society for Gastric Cancer. "The General Rules for the Gastric Cancer Study, 11th Edition," pp. 46–57 (1985). Kanehara Publishing Co., Tokyo (in Japanese).
7) Hanai, A., Sakagami, H., Fujimoto, I., Takano, A., Inoue, R., Nishida, M., Nose, T. and Senoo, I. The second cooperative study of 6 population-based cancer registries for estimating cancer incidence in Japan. *J. Health Welfare Stat.*, **22**(4), 3–15 (1975) (in Japanese).
8) Akaike, H. Information theory and an extension of maximum likelihood principle. *In* "Second International Symposium on Information Theory," ed. H. Petrov and F. Csaki, pp. 267–281 (1973). Akademia Kiado, Budapest.
9) Yamaguchi, N., Watanabe, S., Maruyama, K. and Okubo, T. Estimation of the prevalence of occult cancer and its application to the epidemiology of multiple primary cancer. *Jpn. J. Clin. Oncol.*, **15** (Suppl. 1), 313–323 (1985).
10) Yamashita, S., Kawai, K., Kohli, Y., Ashihara, T. and Fujita, S. Autoradiographic studies on human stomach epithelium by local labeling method. *Abstr. Book 8th Conf. on Radioisotopes,* 226 (1967).
11) Kawai, K., Miyaoka, T. and Kohli, Y. Evaluation of early gastric cancer from the clinical point of view. *In* "Early Gastric Cancer, Current Status of Diagnosis," ed. E. Grundmann, H. Grunze and S. Witte, p. 63 (1974). Springer, Berlin.
12) Fujita, S. Biology of early gastric cancer. *Pathol. Res. Pract.*, **163**, 297 (1978).
13) James, V. H. and Reed, M. J. Steroid hormones and human cancer. *Prog. Cancer Res. Ther.*, **14**, 471–487 (1980).
14) Japanese Research Society for Gastric Cancer. The report of treatment results of stomach carcinoma in Japan (1974, 75, 76, 77, 78). *Rep. Nation-wide Regist. Stomach Carcinoma Japan,* **25**, 63–64 (1986).
15) Watanabe, S. Why do we need cancer information. *Jpn. J. Clin. Oncol.*, **20**, 7–15 (1990).