J Chron Dis Vol. 37, No. 6, pp. 465–480, 1984
Printed in Great Britain.

0021-9681/84 $3.00 + 0.00

THE MULTIHIT MODEL OF CARCINOGENESIS: ETIOLOGIC IMPLICATIONS FOR COLON CANCER

JEFFREY V. SUTHERLAND and JOHN C. BAILAR III

Department of Radiology C278, University of Colorado School of Medicine, Denver, CO 80262 and Department of Biostatistics, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115, U.S.A.

(*Received in revised form* 22 *October* 1983)

Abstract—A new multihit model of carcinogenesis is developed for use in evaluating age-specific cancer incidence rates in human populations. The model allows for some heterogeneity in both risk (perhaps genetic) and pathway (number of hits). Fitting the model yields estimates of (1) levels of effect of background exposure to environmental agents, (2) tumor growth times after initiation of a malignant cell, and (3) relative sizes of high-risk groups in a human population. Maximum likelihood procedures are used to fit the model to the polyposis coli data of Veale and the colon cancer incidence data from the Third National Cancer Survey. Model estimates may be verified in some cases by review of independent data in the literature and results have both theoretical and practical implications. Findings are generally consistent with the adenoma-carcinoma etiologic sequence postulated by Hill, Morson and Bussey with one exception. A large proportion of the population may be at risk of four-hit colon tumors following a non-adenoma etiologic sequence.

1. INTRODUCTION

SEVERAL multihit models of carcinogenesis have been proposed to explain epidemiologic and laboratory data on cancer incidence [1–11]. Such mathematical formulations of tumor development may be used to study variations in tumor development time, estimate the size of high risk groups, develop new etiological hypotheses of tumor induction, and assess strategies for cancer prevention. We adopt Berenblum's view [12] that a tumor originates from a single cell and moves through several stages of change and growth before it becomes manifest as a clinical malignancy. A heuristic model of this process has been proposed [13, 14], in which a malignant cell is induced through alteration of cellular DNA by radiation, chemicals, viruses, or other factors. This process may be retarded by phenomena such as DNA repair, or accelerated by promoting agents. A single malignant cell replicates to form a tumor at a rate dependent on numerous stimulating and inhibiting factors.

Farber [15, 16] suggested that a malignant cell is the result of an evolutionary process in which a normal cell undergoes several mutation-like events, each followed by selection as the altered cell reproduces itself. These events were considered rate-limiting steps, which could be viewed as "hits" in the multihit model of carcinogenesis. Since either mutation-like events or actual mutations have the same mathematical implications in the multihit model, these terms will be used interchangeably.

Previous work in theoretical epidemiology has set a precedent for this view. Lilienfeld and Lilienfeld [17] noted the similarity of a multihit model developed by Burch [18] to the

Partially supported by NIH grant CA31682 (Sutherland) and a grant from the Alfred P. Sloan Foundation (Bailar). A preliminary version of a portion of this manuscript was presented at the 70th Annual Meeting of the American Association for Cancer Research, New Orleans, May 16, 1979.

Reed–Frost model of epidemics. The conversion of a susceptible to a case by contact is analagous to a mutation which converts a normal to a malignant cell.

Multihit models have been fitted to human cancer incidence data from population-based tumor registries, but the usefulness of such models has been hampered by two major problems [7]. First, the number of hits required to produce a malignancy seems to vary for different cancers in the same population and from one population to another for specific forms of cancer. Second, some cancers, such as the leukemias, do not appear to fit the model at all. Development of new procedures to account for genetic, environmental, and other heterogeneity in the population at risk [19], variations in tumor growth patterns, particularly for slowly growing tumors such as those of the colon and prostate [20, 21], and the effects of temporal trends in cancer incidence on data from cross-sectional studies [22] can improve the utility of the multihit model of carcinogenesis.

In this paper a new model is developed that incorporates both (a) parameters necessary for heterogeneity in a population at risk and (b) variations in tumor growth patterns. The model is applied to human colon cancer incidence data from the Third National Cancer Survey [23] and etiologic implications are discussed. Extension of the model to allow for estimation of temporal trends in cancer incidence is developed in a separate article [22].

2. REVIEW

A review of quantitative theories of carcinogenesis has been provided by Whittemore and Keller [24]. The original ideas essential to recent mathematical models of carcinogenesis are found in three nearly simultaneous primary sources. A one-hit model dependent on monoclonal tumor origins (that is, that a tumor originates in a single cell) was proposed in 1950 by Iversen and Arley [25]. Fisher and Holloman [26] suggested a model based on multicellular tumor origins in 1951. And in the same year, Muller [1] became the initial proponent of the multiple mutation theory of radiation carcinogenesis, the forerunner of modern multihit models of tumor development.

The one-hit model was extended to a two-hit process [27], but this approach has not been successfully applied to human cancer incidence data until recently [28, 29]. Data indicating that most tumors are monoclonal [30] have cast doubt on the multicellular model introduced by Fisher and Holloman [26]. Nordling [2] examined cancer mortality statistics from several countries and concluded that a multistage process might explain why cancer mortality rates in males increased according to the sixth power of age. Stocks [3], Armitage and Doll [4, 5], and others [7, 8, 28] have extended this work.

3. DEVELOPMENT OF A NEW MODEL

We assume that the process of carcinogenesis has two clearly defined stages. First, there is a stage of induction of a single malignant cell; this is followed by a stage of tumor growth or multiplication of that cell in a promoting environment to form a clinically observable tumor. The concept of Iversen and Arley [25] of cells that require a single hit for indication of malignancy is extended to the case where cells require multiple hits. A cell is defined to be malignant when it has received a sufficient number of hits to become the original cell of a malignant monoclonal tumor.

3.1. A mathematical model of the initiation process

The waiting time, t_1, until the appearance of a malignant cell (the kth hit) has a gamma distribution if the hits are independent, and identically distributed exponential variables with parameter m, or

$$g(t) = m^k t_1 k^{-1} \exp(-mt_1)/(k-1)!. \tag{1}$$

Then if N cells in a tissue are susceptible to alteration, the distribution of the first order statistic (that is, the time the first cell acquires its kth hit and becomes malignant) is

$$g(t_1) = N\{1 - G(t_1)\}^{N-1} g(t_1),$$

where $G(t_1)$ is the gamma cumulative distribution function. This relationship was first noted by Armitage *et al.* [31, 32].

Mayneord and Clarke [33] estimated the number of cells comprising a person of average size to be 10^{12}. Small organs such as the thyroid gland have about 2×10^8 cells, while larger organs such as the lungs or skin have an estimated 5×10^{10} cells. We assume that N is large enough that $g_1(t)$ is approximated well by its limiting distribution as N approaches infinity. This limiting distribution can be shown to be a Weibull distribution [34]:

$$g_1(t_1) \sim w(t) = km^k t_1^{k-1} \exp\{-(mt_1)^k\}. \tag{2}$$

with cumulative distribution function:

$$W(t_1) = 1 - \exp\{-(mt_1)^k\},$$

and hazard function (or incidence rate at time t_1):

$$h(t_1) = km^k t_1^{k-1}.$$

Watson [35, 36] has shown that this process may be approximated by the Weibull distribution under more general conditions. If the probability that any given cell becomes malignant remains small during the human life span, the hits need not be independent, the rate-limiting steps may be epigenetic rather than mutational, and hits may occur in any sequence, or even be reversible by DNA repair or other phenomena.

Considerable evidence supports the notion that carcinogenesis in any tissue or organ may involve several independent biological mechanisms; each mechanism may require one or more hits. More specifically, background exposures to environmental agents, or genetic predisposition to cancer, may move a cell part way through the process for some mechanisms, but not others. Hence, if $k - 1$ hits have been accumulated in one pathway from background exposures, an added dose increment capable of providing the last hit may produce a linear increment in risk, while agents that act through other multihit pathways may produce sublinear responses. Thus heterogeneity (environmental or genetic) in tumor induction systems may result in linear, quadratic, or higher order polynomial dose–response curves.

Also, individuals may be disposed to a tumor arising from different numbers of hits with varying probabilities due to both genetic and environmental factors. For example, multiple carcinogens may alter the probability of tumor induction through a specified pathway. Nevertheless, estimation of the distribution of probabilities of induction of one, two, . . . , or M-hit tumors in an aggregate population will yield some indication of the size of predisposed subgroups.

3.2. Tumor growth after cellular initiation

Many authors who have fitted the multihit model to human cancer incidence data have assumed that the time from the appearance of a single malignant cell to clinical appearance of a tumor is a constant [7, 10, 37, 38]. While this assumption may be a reasonable first approximation, Sutherland and Bailar [20] have shown that it also may lead to erroneous estimates of the number of hits required to induce a cancer, particularly for slowly growing tumors.

Suppose that the first malignant cell proceeds to a clinically observable tumor without regression and that cells independently producing other malignant tumors at a later time grow at the same rate as the first. Departures from these assumptions occur if (1) regression or immunologic suppression of an aberrant cell is common or (2) cells independently rendered malignant proceed to a clinically observable tumor at radically different rates, i.e. the first malignant cell may produce a clinically observable tumor after the appearance of a tumor caused by a second, independent malignant cell. Evidence in the literature suggests that the first situation is rare or not (yet) documented [16, 39, 40, 41, 42] for human cancers induced by low-level environmental exposure to carcinogens. No evidence is available on human tumor growth rates prior to clinical diagnosis of malignancies. Tumor growth rates are highly variable after diagnosis and such variability would effect

parameters estimated by the model. However, animal data on microscopic colon tumors indicate that variability during early stages of growth is not great enough to produce many clinically detectable tumors out of sequence with appearance of the original malignant cells that form the tumors [43].

There is evidence that many tumors grow exponentially, at least to a first approximation [43–48]. For the moment suppose that tumor growth is exponential during the entire tumor growth period. Let n denote the number of doublings in cell number required for a tumor to grow to detectable size, and D denote time for a tumor to double in cell number (or, approximately, size). Then t_G, the duration of tumor growth at the time of detection can be estimated as

$$t_G = n\,D.$$

We now assume that n and D are independent lognormal random variables; this implies that tumor growth time up to the point of detection is also lognormal. These assumptions of lognormality are supported by data on lung tumors described by Charbit *et al.* [49], Geddes [45], Chahinian [47], Meyer [50], Schwartz [51], and Weiss [52]. Available data [47, 50–52] suggest that the correlation between the required number of tumor doublings n and size at time of clinical detection D is less than 0.3. For present purposes this does not seem unacceptably out of line with our technical assumptions of independence.

Data on tumor sizes at clinical diagnosis from the Cancer Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute [53] also support lognormality. Figure 1 shows that the lognormal cumulative distribution function provides a remarkably good fit to the distribution generated by empirical data on 13,540 individual patients with information on the size of first observed colon tumor. A Kolmogorov–Smirnov test [54] for goodness of fit, however, indicated a slight (but statistically significant) deviation in a positive direction from lognormality at small tumor sizes (see Fig. 1). Excess numbers of small tumors may be attributable to differential tumor detection caused by screening of asymptomatic patients. The discussion below assumes lognormality of D as a useful approximation.

3.3. Relaxing the assumption of exponential growth

Some evidence [43, 45, 46, 55] supports the conjecture that tumor growth is not strictly exponential. However, whenever growth is a function of attained tumor size only (and not time), a change of the time scale can be used to create exponentiality. With this assumption, actual tumor growth time t'_G from the appearance of a single malignant cell to clinical observation of a tumor is a constant percentage of the time t_G expected under the assumption of exponential growth. This constant percentage or proportionality constant can be patient-specific, tumor-specific, or both, to accomodate heterogeneity.

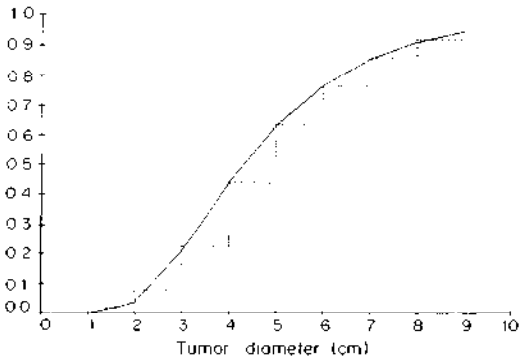

FIG. 1. Cumulative distribution function (CDF) of tumor sizes of 13,540 colon tumors diagnosed 1975–1977. Solid line is lognormal CDF. Dotted line is empirical CDF.

In order to allow for non-exponential tumor growth, suppose this percentage is denoted C_k where k is the number of hits required to induce a malignant cell. Then, $t'_G = C_k t_G$. If t_G is lognormally distributed, then so is t'_G. Thus we obtain the following density function:

$$L(t'_G) = (\sqrt{2n}\,\sigma\, t'_G)^{-1} \exp\{-(\ln t'_G - \mu')^2/2\sigma^2\}, \tag{3}$$

where $\mu' = \mu + \ln C_k$, $\mu = E(\ln t_G)$ and $\sigma^2 = \mathrm{Var}(\ln t_G)$.

The convolution of the distribution of time to initiation, equation (2), with the distribution of tumor growth time, equation (3), yields the distribution, $f(t)$, of the time to clinical appearance of a k-hit tumor:

$$f(t) = f(t_1 + t'_G) = w(t_1; m, k) * L(t'_G; C_k, \mu, \sigma), \tag{4}$$

where m is the gamma hazard rate of a k-hit tumor.

3.4. Heterogeneity of the population at risk

It is well known that some individuals are genetically predisposed to certain malignancies [56]. Crump [57] has commented on the desirability of a mathematical model which would allow for a variety of genetically susceptible subgroups in a population. In the context of a hit-theory model, Burch [58] suggested that genetically predisposed subgroups of a human population may require fewer hits to produce a malignant tumor. Knudson [59–62] and Ashley [63, 64] have fitted models to cancer incidence data to illustrate the possible effects of heterogeneity.

Not all heterogeneity in risk is genetic. Concomitant disease (e.g. immunodeficiency) can cause marked variations in risk. Heterogeneity in exposure, rather than response, may be extremely important. Here we deal directly with only genetic heterogeneity of a population but extensions to other kinds of mixtures should be straightforward.

Genetic heterogeneity is incorporated into the present model by assuming that the number of hits required to produce a malignant cell varies from individual to individual, $k = 1, 2, \ldots, M$. A variety of discrete and continuous distributions have been evaluated [65] for modeling the probability of induction of a malignant cell after k hits (binomial, geometric, negative binomial, Poisson, negative hypergeometric, lognormal, gamma, normal, beta distributions). The use of continuous distributions in this context (indeed, of any distribution not strictly on the positive integers) has not been closely argued, but seems to imply that the multihit model is being used solely for its descriptive properties and not for the interpretive/explanatory properties that led to its original development (and its name). Lack of flexibility in these unmixed distributions, both continuous and discrete, often results in a fit of the model to incidence data that is poor compared to the fit attained by allowing for M discrete subgroups in a population where the sizes of the M subgroups are considered the proportions of the population at risk of $1, 2, \ldots, M$-hit tumors.

3.5. Incorporation of heterogeneity in the model

Let p_k be the probability that an individual is at risk of initiation of a malignant cell after k hits, where $k = 1, 2, \ldots, M$. The distribution of time to clinical appearance of a tumor can be expressed by compounding the probability distribution function, $f(t)$, of the time to clinical appearance of a k-hit tumor equation (4) with respect to the probability of an individual being susceptible to a k-hit tumor. The density $v(t)$ can be written,

$$v(t) = \sum_{k=1}^{M} f(t \mid k; m_k, C_k, \mu, \sigma) p_k. \tag{5}$$

The heterogeneous multihit model, equation (5) has $3M + 2$ parameters: M, μ, σ, m_k and C_k, where $k = 1, 2, \ldots, M$, and p_k, where $k = 1, 2, \ldots, M - 1$. The mean and variance of tumor doubling times and the number of doublings prior to clinical detection may be estimated from independent data in the literature and used to calculate μ and σ as noted previously. We assume that μ and σ are not dependent on mechanisms of induction.

4. FITTING THE MODEL TO COLON CANCER DATA

Fitting the model to cancer incidence data requires a thorough understanding of the type of cancer under study because malignant neoplasms of various tissues have a variety of risk factors, growth patterns, and high risk groups. It is desirable to have a large number of cases of a specific cancer from a homogeneous population whose exposure to relevant etiologic agents is thought to have been constant over many decades. Available data from a carefully studied high-risk subgroup in the population can be helpful in refining and validating the model. Because epithelial tumors may have different underlying mechanisms of origin than other tumors such as leukemias or sarcomas [11] it is desirable to utilize cancer incidence data coded with sufficient detail to identify epithelial tumors. Analysis can then be restricted to this tumor type.

Colon cancer incidence data come close to satisfying the above criteria. Data are available on colon tumor doubling times in animals and humans. In particular, polyposis coli, a genetic disease that produces an extremely high risk of colon cancer, has been carefully studied. Colon cancer occurs with high incidence in western industrialized countries and during this century in the United States, colon cancer incidence and mortality rates have been more stable than rates for most other sites [66], suggesting that exposure to etiologic agents has also been rather stable.

In this section, a method for fitting the multihit model of carcinogenesis (equation 5) to cancer incidence data is illustrated. Parameters for colon tumor growth time are calculated. The model is fitted to data on polyposis coli patients, then to cancer incidence data from the population of Colorado during the period 1969–1971.

4.1. Maximum likelihood fit of incidence data

Cancer incidence data are determined from the number of cases, x_i, and numbers of persons at risk, n_i, for $i = 1, 2, \ldots, g$ age groups. Groups may be further subdivided by sex, race, known exposures, or other factors. Assuming individuals respond independently with low individual probabilities, the probability of x cases occurring within the ith age interval may be regarded as approximately Poisson distributed:

$$P(x_i; m_i) = m_i^{x_i} \exp(-m_i)/x_i!.$$

The likelihood L that exactly x_i cases of cancer occur in g age groups of a population, $i = 1, 2, \ldots, g$, with n_i persons at risk in the ith age group is:

$$L = \prod_{i=1}^{g} P(x_i; m_i), \tag{6}$$

where m_i, the expected number of cases in the ith age interval, is closely approximated by,

$$m_i = n_i \int_{t_{i-1}}^{t_i} h(t)\,dt,$$

where $h(t) = v(t)/(1 - V(t))$ and $V(t)$ is the cumulative distribution function of $v(t)$.

Since $v(t)$ (see equation 5) is a function of the maximum number of hits allowed in the model, m, the model can be fitted sequentially for $M = 1, 2, \ldots, c$. As M is increased the value of the likelihood function will increase. We find the smallest value of i such that the difference in the value of the likelihood for $M = i$ vs $M = i + r$ for each $r = 1, 2, \ldots, c$ and large c is not statistically significant. We then select $M = i$ as the appropriate parameter for the tumor type and population under study.

It should be noted that small sample sizes would result in low values of M being selected because of low statistical power. The likelihood is the product of g Poisson probabilities where g is fixed as the number of age groups specified. Small sample sizes within age groups would cause numbers of cases observed to fluctuate about expected numbers more than for large sample sizes. This effect is essentially random, and disappears as the sample size approaches infinity. In order to minimize this problem, we selected a tumor type for which large numbers of cases are available for this analysis.

4.2. Colon tumor development time

Failure to include variable tumor development time in the multihit model has inhibited its use as an effective tool in the analysis of human cancer incidence data. Colon cancer, a case in point, is a slowly growing tumor taking many years to develop from a single initiated cell to a clinically detectable neoplasm. Sutherland and Bailar [20] have shown that underestimates of colon tumor development time in previous work resulted in an overestimate of the number of hits required to initiate a malignancy.

Previously, it was shown that tumor growth time may be approximately lognormally distributed with parameters μ' and σ (equation 3). The parameter μ' is equal to the sum of three terms,

$$\mu' = E(\ln n) + E(\ln D) + \ln C_k,$$

The mean size of colon tumors at detection was 4.49 cm in a series of 13,540 measurable colon tumors in the SEER data [53]. Welin *et al.* [67] noted that colon tumors approximate the geometric shape of a cylinder. Using this assumption, the mean tumor size at detection is approximately 36.4 doublings of a cell 10 μm in diameter. The data of Steel [46] shows that the average doubling time for 19 human adenocarcinomas of the colon and rectum is 1.73 years. Then, approximately,

$$\mu' = \ln 36.4 + \ln 1.73 + \ln C_k,$$

where C_k is a constant to be estimated by fitting the model. Since C_k is anticipated to be on the order of 0.5, the average tumor growth time will be about 30 years. The second parameter in the lognormal distribution (equation 3) is estimated by the square root of the sum of the variance of log of the number of doublings required prior to clinical detection and the variance of the log doubling time.

4.3. Computer procedures

A computer program was developed to estimate the parameters required to fit the multihit model of carcinogenesis to colon cancer incidence data. Convolutions were calculated using a numerical analysis procedure developed by the authors [65]. Choice of interval widths of time in the Weibull and lognormal distributions were found to be critical. Selection of a very small interval width required excessive use of computer time. Large interval widths generated errors in evaluation of the convolution. A satisfactory width was found to be 0.25 years. This required calculation of 340 values of the Weibull and lognormal distributions (to cover ages 0–85) prior to evaluating the convolution at each step of the iterative process of maximizing the likelihood.

Parameter estimates must be greater than zero in most cases to be biologically or statistically meaningful. Also, inappropriately large values of parameter estimates often caused excessive use of computer time and produced meaningless results. Parameter values were therefore restricted to meaningful intervals using a logistic transform.

A subroutine which utilized a quasi-Newton method with different approximations to derivatives was used to minimize -2 log likelihood. Confidence intervals were readily calculated on parameters generated by the model since differences in -2 log likelihood asymptotically follow the chi-square distribution [68]. A FORTRAN listing of the program is available from the authors.

4.4. Fitting polyposis coli data

A well-known hereditary predisposition to carcinoma of the colon occurs in persons with polyposis coli. Virtually all persons with untreated polyposis of the colon get colon cancer unless they die prematurely. Since polyposis coli patients form a major genetically predisposed subgroup for the induction of colon cancer and since these individuals apparently require fewer hits [63] to initiate malignancy, we utilize data on such patients to develop preliminary estimates for the probability of single-hit tumor induction.

The polyposis coli data of Veale [69] when plotted as incidence curves show a bimodal pattern that is more pronounced in males than in females (Figs 2 and 3). Reed and Neel

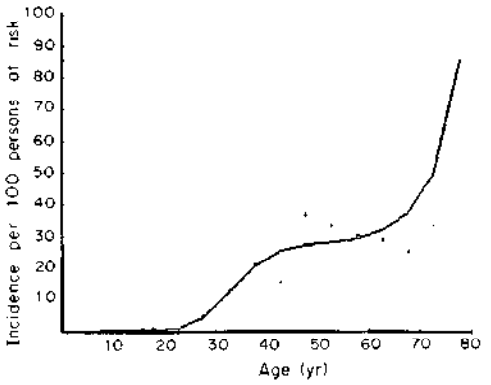

FIG. 2. Annual colon cancer incidence in 156 male polyposis coli patients. The multihit model is fit to the data of Veale [69]. Plus signs are empirical data and the curve is estimated by the model.

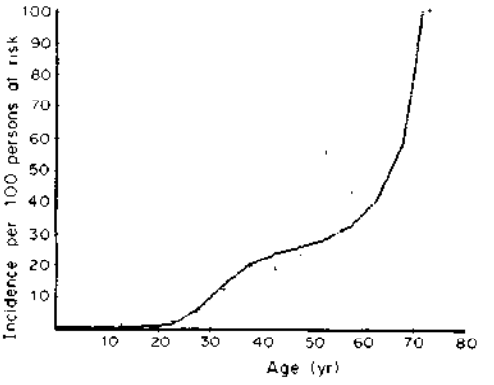

FIG. 3. Annual colon cancer incidence in 114 female polyposis coli patients. The multihit model is fit to the data of Veale [69]. Plus signs are empirical data and the curve is estimated by the model.

[70] noted the same phenomenon when their cases were combined with the data of Dukes [71]. Utsunomiya *et al.* [72] noticed bimodality in Japanese data but reported that it was not statistically significant. Ashley [63] reviewed the data of Veale [69] on colon cancer occurring in polyposis coli patients in the context of the multihit model of carcinogenesis. He estimated that persons with polyposis coli require one or two fewer hits than normal persons for the induction of colon cancer.

The data of Veale [69] (after Ashley [63]) were retabulated into a form suitable for fitting by the model under discussion (Table 1). The model was fitted for M equal to 1, 2, or 3. Differences in values of the likelihood function indicated that $M = 2$ provided a significantly better fit than $M = 1$ for males ($p < 0.005$) but not for females. No significant improvement in the fit was found for $M = 3$.

Model parameters for $M = 2$ are tabulated in Table 2. Sex differences in parameter values generated by the model are not statistically significant. Model results suggest that one-hit tumors occur with much higher probability and are more slowly growing than two-hit tumors. However, only about 22% of individual predisposed to polyposis coli are susceptible to one-hit tumors, whereas 78% are susceptible to two-hit tumors. Colon cancer arises in 100% of the one-hit group, but only 15% of the two-hit group. The majority of

TABLE 1. POLYPOSIS COLI DATA OF VEALE [69]*

Midpoint of age interval	Males		Females	
	Cases	Pop	Cases	Pop
2.5	0	156	0	114
7.5	0	156	0	114
12.5	0	153	1	113
17.5	1	147	0	106
22.5	0	137	2	97
27.5	6	127	5	84
32.5	14	109	8	66
37.5	17	88	10	51
42.5	8	53	7	37
47.5	14	38	6	26
52.5	6	18	10	18
57.5	3	10	3	7
62.5	2	7	1	3
67.5	1	4	0	1
72.5	1	3	1	1
77.5	2	2		

*Data retabulated from Ashley [63]. Note that the population at risk decreases with advancing age, due to individuals being treated for polyposis coli by removing the colon, sometimes before the appearance of any malignant tumor. Pop = population at risk.

TABLE 2. POLYPOSIS COLI DATA ON VEALE [69] MAXIMUM LIKELIHOOD PARAMETER ESTIMATES AND 95% CONFIDENCE INTERVALS

Parameter	Male
M	2
m_1	0.07559 (0.05289–0.10688)
m_2	0.00326 (0.00214–0.00359)
C_1	0.52882 (0.48820–0.56992)
C_2	0.16318 (0.06285–0.31917)
p_1	0.21845 (0.19327–0.22652)
	Female
M	2
m_1	0.05542 (0.03869–0.06562)
m_2	0.00366 (0.00234–0.00400)
C_1	0.49793 (0.44347–0.55771)
C_2	0.07248 (0.00000–0.21258)
p_1	0.22536 (0.19052–0.23564)

*The parameter M is the maximum number of hits allowed in the model. The average probability of a hit is m and varies for one and two-hit tumors. C_k is the expected proportion of tumor development time under the assumption of exponential growth and varies for one and two-hit tumors. The parameter p is the proportion of the population at risk of a one hit tumor. The proportion of the population at risk of a two-hit tumor is equal to one minus p_1. See equation (5).

observable tumors arise in the one-hit subgroup. Caution must be used in interpreting this result because the majority of polyposis coli patients are treated by removal of the colon and are no longer in the population at risk (see discussion in Section 4.6).

4.5. Fitting colon cancer data

Incidence data for colon cancer have been evaluated by Cook *et al.* [7] and Ashley [76] yielding hit parameters on the order of 6.5 for males and 5.5 for females. Ashley [64] comments extensively on six differences suggested by the data. However, sex differences are variable and sometimes non-existent in different countries [7]. Genetic or environmental heterogeneity of male and female populations in different parts of the world could generate such anomalies.

Data from the U.S. Third National Cancer Survey for Colorado are evaluated here [23]. Sarcomas and tumors of unspecified type were removed from the data since sarcomas are a different biological entity that happens to occur in the same anatomic location. They tend to appear at earlier ages and to be faster growing than epithelial tumors [11, 77]. Table 3 shows numbers of remaining cases occurring in white males and females aged 0–34, and for single-year age groups from 35–36 to 84–85 years of age and the 1970 census population data available for each age group. Single-year age groupings were used in order to provide the maximum likelihood estimation procedure with as much information as possible. Data for cases over age 85 were not used because census population figures for age groups by year were available only up to age 85 and incidence data from groupings at older ages may be unreliable [10, 22].

Model parameters were fitted to the data by maximizing the likelihood function (equation 6). The probability of a hit in the development of a single-hit tumor was assumed to be the m_1 obtained by fitting the model to the polyposis coli data, because most single-hit tumors appear to arise in polyposis coli patients. The parameters m_k, $k = 2, 3, \ldots, M$, C_k, $k = 1, 2, \ldots, M$ and p_k, $k = 1, 2, \ldots, M - 1$, were estimated from the data. M was fixed initially at two and increased in subsequent runs until the likelihood function reached an apparent maximum (evaluated by its showing no subsequent significant increase). A value of $M = 4$ was found sufficient to describe the data (see Fig. 4 and 5). Concurrently, the data were fitted by assuming that a proportion of the population was not susceptible but this model did not maximize the value of the likelihood function and was discarded. A

TABLE 3. COLON CANCER IN COLORADO, 1969–1971*

Midpoint of age interval	White males		White females	
	Cases	Pop	Cases	Pop
0 5 29.5	4	588,589	5	574,827
30.5	0	13,421	1	13,620
31.5	1	12,560	1	13,092
32.5	0	12,479	1	12,794
33.5	1	12,220	0	12,505
34 5	1	12,202	4	12,471
35.5	3	12,242	1	12,273
36.5	1	11,539	3	11,943
37.5	1	11,863	0	12,313
38.5	1	12,154	3	12,136
39.5	3	12,367	3	12,348
40.5	0	12,759	3	12,748
41.5	3	12,051	3	12,414
42.5	4	11,996	2	12,587
43.5	3	11,824	4	12,485
44.5	6	11,955	4	12,374
45.5	9	11,741	6	12,614
46.5	2	11,842	3	12,547
47.5	5	11,775	6	12,277
48.5	9	12,073	6	12,490
49.5	11	11,320	16	11,710
50.5	4	11,511	11	11,667
51.5	6	10,910	8	11,005
52.5	4	10,481	10	10,608
53.5	10	9220	5	10,116
54.5	8	9628	8	10,016
55.5	11	9257	16	9878
56.5	12	9007	11	9638
57.5	10	8743	7	9439
58.5	17	8580	17	8977
59.5	18	8182	16	8962
60.5	15	8247	16	9090
61.5	17	7646	14	8086
62.5	13	7368	17	8115
63.5	12	6880	20	7538
64.5	18	6588	13	7477
65.5	15	5889	18	7120
66.5	19	6039	18	6932
67.5	20	5540	27	6562
68.5	21	4918	15	6046
69.5	17	5074	27	6375
70.5	18	4837	20	6334
71.5	17	4343	24	5698
72.5	12	3876	21	5313
73.5	22	3865	22	5117
74.5	27	3716	21	5011
75 5	17	3475	29	4831
76.5	22	3200	29	4639
77.5	18	3023	29	4420
78.5	18	2584	23	3877
79.5	15	2382	24	3633
80.5	15	2386	35	3542
81.5	16	2056	24	3128
82.5	16	1629	23	2727
83.5	8	1460	20	2363
84.5	16	1333	22	2213
Total	639	1,040,193	805	1,070,554

*Data selected from Third National Cancer Survey tape supplied by the NCI Biometry Branch and include white cases with site codes 150.0–150.9 except for eight sarcomas and nineteen tumors with unspecified histology. Pop = population at risk.

summary of parameter estimates and 95% confidence intervals on these estimates is presented for male and female colon tumors in Table 4.

Average tumor development times are estimated from equation (6) to be in the range 21.4–40.3 years for one-, two-, three-, and four-hit tumors with one-hit tumors growing fastest in both males and females ($p < 0.05$). The suggestion of fast growing one-hit tumors may be an artifact generated by underreporting of single-hit tumors in older polyposis coli patients or due to a multiplicity effect. The fastest growing tumor may be the first tumor detected in a polyposis coli patient with multiple tumors. Alternatively, the proportion of susceptible persons in the population surviving to older ages may be differentially depleted

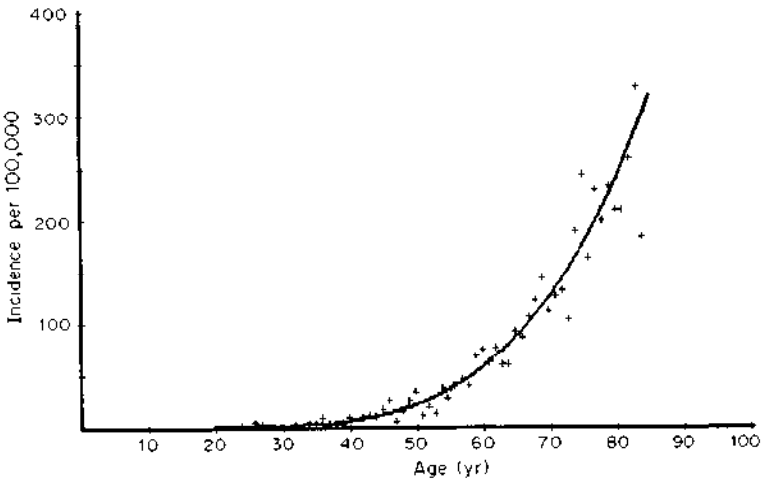

FIG. 4. Annual colon cancer incidence based on 639 epithelial tumors occurring in 1,040,193 white Colorado males ages 0–85 during the Third National Cancer Survey (1969–1971). Plus signs are empirical data and the curve is estimated by the model.

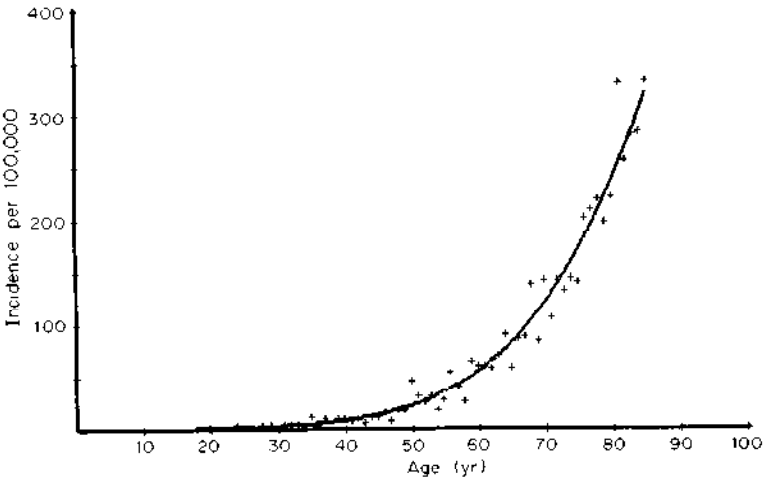

FIG. 5. Annual colon cancer incidence based on 805 epithelial tumors occurring in 1,070,554 white Colorado females ages 0–85 during the Third National Cancer Survey (1969–1971). Plus signs are empirical data and the curve is estimated by the model.

due to cancer at younger ages, or prophylactic colectomies may artifactually lower estimates of single-hit tumor incidence rates.

Table 5 displays in summary form the implications of the model concerning proportions of the population at risk for one-, two, three, and four-hit tumors and the percent of cancers observed each year that correspond to the various mechanisms of carcinogenesis. The parameters p_i, $i = 1, \ldots, 3$, were used to estimate the size of population subgroups at risk of one, two, three, and four-hit tumors. Cumulative incidence is the sum of the age-specific incidence rates for persons aged 0–74 and indicates the probability of disease in a subgroup over a lifespan. The percent of tumors observed is proportional to the size of subgroups at risk and the cumulative incidence of colon cancer within subgroups. Less than 1% of observed tumors arise in patients with polyposis coli. Over 30% are estimated to be two-hit tumors arising in persons genetically predisposed to isolated polyps, 17.2% arise as three-hit tumors, and 51.9% are four-hit tumors.

4.6. Discussion

The polyposis coli data suggest that one-hit tumors occur with higher probability and grow more slowly in a subgroup of 22% of patients with polyposis coli than two-hit tumors

TABLE 4. COLORADO COLON CANCER INCIDENCE DATA THIRD NATIONAL CANCER SURVEY (1969–71) MAXIMUM LIKELIHOOD PARAMETER ESTIMATES AND 95% CONFIDENCE INTERVALS FOR MULTIHIT MODEL WITH HETEROGENEITY

Parameter	Male
M	4
m_1	fixed at 0.07559 (see Table 2)
m_2	0.00446 (0.00371–0.00515)
m_3	0.00366 (0.00297–0.00418)
m_4	0.00681 (0.00653–0.00707)
C_1	0.39032 (0.29755–0.47978)
C_2	0.56421 (0.50465–0.63167)
C_3	0.51915 (0.43377–0.64303)
C_4	0.60970 (0.58158–0.63991)
p_1	0.00006 (0.00002–0.00015)
p_2	0.04740 (0.03150–0.06479)
p_3	0.28041 (0.02906–0.52498)
	Female
M	4
m_1	fixed at 0.05542 (see Table 2)
m_2	0.00334 (0.00293–0.00374)
m_3	0.00342 (0.00257–0.00401)
m_4	0.00714 (0.00689–0.00739)
C_1	0.33920 (0.23403–0.42826)
C_2	0.50193 (0.45154–0.55454)
C_3	0.54220 (0.43301–0.71743)
C_4	0.63963 (0.61497–0.66594)
p_1	0.00006 (0.00001–0.00015)
p_2	0.08388 (0.06182–0.10819)
p_3	0.27459 (0.10923–0.42955)

*The parameter M is the maximum number of hits allowed in the model. The average probability of a hit for an k-hit tumor is m_k. C_k is the expected proportion of tumor development time expected under the assumption of exponential growth. The parameter p_k is the proportion of the population at risk of a one-, two-, or three-hit tumor. The proportion of the population at risk of a four-hit tumor is equal to $1 - p_1 - p_2 - p_3$. See equation (5).

occurring in a subgroup of 78% of polyposis coli patients since m_1 is larger than m_2, and C_1 is larger than C_2 for both males and females (see Table 2). This is consistent with the hypothesis of Veale [69], who originally suggested that three allelic genes form the genetic basis of colon tumor etiology. He postulated a dominant gene "P" linked to the development of polyposis coli, a recessive gene "p" related to the induction of isolated polyps, and a normal gene "+". The Pp genotype results in early onset of polyposis coli, the P+ genotype produces late onset disease, and the pp genotype results in the appearance of isolated polyps. The very rare PP genotype was never observed with certainty and the p+ genotype was considered indistinguishable from the normal ++ genotype.

Veale's hypothesis may be true for some two-hit tumors, but is unlikely to be true for others. Fifty-two percent of the polyposis coli patients were selectively removed from the population at risk by prophylactic colectomy [69]. If only low risk patients remained at risk, the model would tend to select these individuals into a two-hit subgroup with short tumor growth time. The P+ group postulated by Veale could still be a small percentage of the 78% of polyposis coli patients in the two-hit subgroup. In any event, individuals

TABLE 5. SUMMARY TABLE IMPLIED BY PARAMETER ESTIMATES*

Genome	Mechanism of carcinogenesis	Frequency (%) of genome in population	Cumulative incid. (0–74) of cancer in genome	Percent of cancers observed in population
Polyposis + polyp	1-hit	~0	100	0.3
Polyposis + normal	2-hit	~0	100	0.1
Polyp + polyp	2-hit	6.6	9.8	30.5
Polyp + normal	3-hit	27.7	1.2	17.2
Normal + normal	4-hit	65.7	1.6	51.9

*Male and female estimates are averaged

in the two-hit subgroup of polyposis coli patients produce a negligible number of tumors in a general population series.

The analysis of the Third National Cancer Survey series is of interest because whether or not colon cancer can arise *de novo* is an area of controversy. Muto *et al.* [78], after reviewing the literature and evaluating their own data, stated the following:

(1) Considerable evidence supports the hypothesis that adenomatous polyps and villous adenomas predispose to colon cancer. It is generally agreed that hyperplastic and juvenile polyps have no malignant potential.

(2) Some authorities argue that adenomatous polyps have no malignant potential. However, all are agreed that villous adenomas predispose to colon cancer [79].

(3) A recent prevailing view is that adenomatous polyps and villous adenomas are histologic variants of the same neoplastic process.

(4) Definitive evidence is not available to show that all colon cancer arises from adenomatous polyps and villous adenomas. However, sufficient evidence exists to suggest that at least half of the colon cancers develop from previously benign adenomatous polyps and villous adenomas.

The parameter estimates generated by fitting the multihit model to colon cancer incidence data suggest that half of the colon cancers develop *de novo* when viewed in relation to current theories of the pathogenesis of colon cancer. The rationale for this interpretation is discussed below. Development of cancer in patients with ulcerative colitis is not considered here since they appear to account for less than 1% of colon tumor incidence [80].

Hill *et al.* [74] extended the hypothesis of Veale for the pp genotype. They postulated three environmental agents A, B, and C, which may be viewed as hits in the multihit model. Then,

$$\text{pp} + \text{A} \rightarrow \text{small adenomas},$$
$$\text{pp} + \text{A} + \text{B} \rightarrow \text{large adenomas},$$
$$\text{pp} + \text{A} + \text{B} + \text{C} \rightarrow \text{carcinoma},$$
$$\text{and occasionally, pp} + \text{A} + \text{C} \rightarrow \text{carcinoma}.$$

This hypothesis was supported by pathological and epidemiological data on the development of adenomas and carcinomas. Note that the pathway pp + A + C can result in a two-hit carcinoma in a person without polyposis coli.

Parameter estimates indicate that the majority of individuals (67.2% of males, 64.1% of females) are at risk of four-hit tumors. In addition, the risk of a four hit tumor is estimated to be higher ($p < 0.05$) than the risk of a three-hit tumor. If the four-hit tumors were simply an extension of the three-hit adenoma sequence, the four-hit tumors should be less probable. Since they are not, this implies that at least one pathway to a four-hit tumor is independent of the three-hit adenoma sequence tumors.

Morson [82] observed that 10% of colon carcinomas at St Marks Hospital showed evidence of benign adenomatous or villous tumors in continuity with invasive growth. The proportion was 50% for early tumors. It was suggested that growing cancers destroy previously benign tumors. One of the possible alternative explanations is that over 50% of colon cancers follow a four-hit pathway different from the adenoma-carcinoma sequence. Early cancers might tend to show excess adjacent benign adenomatous or villous tumor tissue because these cancers would be more likely to be two or three-hit tumors. This hypothesis could explain the observation of Burkitt [83] that colon cancer exists in rural communities in developing countries where polyps are virtually unknown. These colon cancers may be four-hit tumors that do not require environmental exposures essential to development of polyps which are precursors to one-, two, or three-hit tumors. (Of course, it is by no means clear that relevant environmental carcinogens are absent, or even uncommon in less developed areas.)

Results imply (see Table 5) that the percentage of four-hit tumors is almost the same as the percentage of observed colon tumors for which Morson *et al.* cannot document the

occurrence of the adenoma-carcinoma etiologic sequence. Slightly less than half the tumors may arise from a polyp sequence. In addition, 6.6% of the population is at risk of non-polyposis two-hit tumors, a group which probably represents individuals with familial predisposition to colon cancer. We note that Lynch *et al.* [84] estimated that 6.58% of persons in the U.S. have two or more first degree relatives with cancer and are at elevated risk for colon cancer. This group, although small, is at high risk and produces a substantial portion of the numbers of tumors observed in the population at large.

4.7. Advantages and disadvantages of the present model

Available data on the epidemiology of adenomatous polyps do not permit comparison of hit probabilities estimated by the multihit model with the incidence of polyps in different countries or in regions of the United States. If such data were available, examination of parameter estimates generated by fitting the model to cancer incidence data (which are now available for many regions and countries) and the relation of these estimates to the incidence of polyps, dietary factors, and lifestyle variations in different population groups could suggest specific etiologic hypotheses.

Variability in the genetic heterogeneity between male and female populations should be examined through use of additional colon cancer incidence data available in the United States and other countries. Evaluation of data from other tumor sites will also be of interest, particularly if the size of a genetically predisposed subgroup can be estimated (breast cancer might be a good example).

The decreasing incidence of colon and other cancers in some populations at older ages in the 1969–1971 data could indicate that only a subset of the total population is at risk of developing colon cancer. However, reduced risk, such as of lung cancer, in older cohorts or underreporting in old age groups [10] could produce the same pattern, making interpretation difficult. The collection of more complete cancer incidence data, particularly in older age groups, would help to reduce the number of alternative hypotheses.

In its present form our biological interpretation of the model is strictly valid only for continuous constant exposure to carcinogens. Devesa and Silverman [66] state that U.S. colon cancer incidence rates have increased from 23.8 per 100,000 in 1947–48 to 29.0 in 1969–71 in white males and decreased from 26.0 to 24.8 per 100,000 in white females during the same periods. These changes are small and should result in minor variations in parameter values. Further research in this area is needed, however, particularly if this model is now to be applied to tumor sites where incidence is highly variable over time. The model is extended to the more complex case of variable exposure in a separate manuscript [22].

The primary value of this multihit model is its capability of generating parameter estimates which can be directly related to independent data in the literature. As new data become available on tumor cell kinetics, the size of tumors at diagnosis, the epidemiology of polyps and carcinomas, and the effects of various environmental carcinogens, the model can be refined and updated. Although alternative models may provide an adequate fit to cancer incidence data, few at present provide parameter estimates that can be empirically verified.

Acknowledgements—The authors acknowledge the helpful suggestions of Drs Emmanuel Farber, Alfred G. Knudson Jr, Henry T. Lynch, Suresh H. Moolgavkar, Strother H. Walker, Alice S. Whittemore, and an anonymous reviewer who suggested the format for results presented in Table 5.

REFERENCES

1. Muller HJ: Radiation damage to the genetic material. **Sci Prog** 7: 93–493, 1951
2. Nordling CO: A new theory on the cancer inducing mechanism. **Br J Cancer** 7: 68–82, 1953
3. Stocks P: A study of the age curve for cancer of the stomach in connection with a theory of the cancer producing mechanism. **Br J Cancer** 7: 407–428, 1953
4. Armitage P, Doll R: The age distribution of cancer and a multi-stage theory of carcinogenesis. **Br J Cancer** 8: 1–12, 1954
5. Armitage P, Doll R: Stochastic models for carcinogenesis. In **Proc 4th Berkeley Symp Math Stat Prob IV**, Neyman J (Ed). Berkeley: University of California Press, 1961. pp. 19–38

6. Neyman J, Scott EL: Statistical aspect of the problem of carcinogenesis. In **Proc 5th Berkeley Symp Math Stat Prob IV,** Lecom LM, Neyman J (Eds). Berkeley: University of California Press, 1967. pp. 745–776
7. Cook PJ, Doll R, Fellingham SA: A mathematical model for the age distribution of cancer in man. **Int J Cancer** 4: 93–112, 1969
8. Townsend JL: Smoking and lung cancer: A cohort data study of men and women in England and Wales 1935–70. **J R Stat Soc** 141: 95–107, 1978
9. Doll R: Cancer and aging: the epidemilogic evidence. In **Oncology 1970. Proc. of the 10th Int Cancer Cong V,** Clark RL *et al.* (Eds). Chicago: Year Book Medical Publishers, 1970. pp. 1–28
10. Doll R: The age distribution of cancer: implications for models of carcinogenesis. **J R Stat Soc Ser A,** 134: 133–166, 1971
11. Peto R: Epidemiology, multistage models and short-term mutagenicity tests. In **Origins of Human Cancer.** Hiatt HH, Watson JD, Winsten JA (Eds). Cold Spring Harbor Laboratory, 1977. pp. 1403–1428
12. Berenblum I: **Carcinogenesis as a Biological problem.** New York, Elsevier, 1974.
13. Trosko JE, Chu EHY: The role of DNA repair and somatic mutation in carcinogenesis. **Adv Cancer Res** 21: 391–425, 1975
14. Boutwell RK: The function and mechanism of promoters of carcinogenesis. **CRC Crit Rev Toxicol** 17: 419–443, 1974
15. Farber E: Carcinogenesis—cellular evolution as a unifying thread: presidential address. **Cancer Res** 33: 2527–2550, 1973
16. Farber E: Sequential analysis of chemical carcinogenesis. **Laboratory Workshop, Cancer Biology I—Etiology, Diagnosis and Treatment.** Given Institute of Pathobiology, 1979
17. Lilienfeld AM, Lilienfeld DE: **Foundations of Epidemiology** (2nd edn). New York: Oxford Univ Press, 1980
18. Burch PRJ: Age and sex distribution for some idiopathic non-malignant conditions in man. Some possible implications for growth-control and natural and radiation-induced ageing. In **Radiation and Ageing.** Lindop PJ, Sacher GA (Eds). London: Taylor & Francis, 1966. pp. 117–155
19. Knudson AG Jr: Genetic predisposition to cancer. In **Origins of Human Cancer.** Hiatt HH, Watson JD, Winsten JA (Eds). Cold Spring Harbor Laboratory, 1977. pp. 45–52
20. Sutherland JV, Bailar JC III: Multihit models: Assumptions and generalizations. In **Advances in Medical Oncology, Research and Education I:** Carcinogenesis. Margison GP (Ed) Oxford: Pergamon Press, 1979. pp. 309–313
21. Sutherland JV, Bailar JC III: The multihit model of carcinogenesis: Application to colon cancer data from the Third National Cancer Survey (abstract). **Proc Am Assoc Cancer Res** 20: 77, 1979
22. Sutherland JV, Bailar JC III: The multihit model of carcinogenesis: Evaluation of temporal trends in cancer incidence data from Third National Cancer Survey (abstract). **Proc Am Assoc Cancer Res** 21: 81 1980
23. Cutler SJ, Young JL: Third National Cancer Survey: Incidence Data. **Nat Cancer Inst Monogr** 41, 1975
24. Whittemore AS, Keller JB: Quantitative theories of carcinogenesis. **SIAM Rev** 20: 1–30, 1978
25. Iversen S, Arley N: On the mechanism of experimental carcinogenesis. **Acta Pathol Microbiol Scand** 27: 773–803, 1950
26. Fisher JC, Holloman JH: A hypothesis for the origin of cancer foci. **Cancer** 4: 916–918, 1951
27. Armitage P, Doll R: A two-stage theory of carcinogenesis in relation to the age distribution of human cancer. **Br J Cancer** 11: 161–169, 1957
28. Ashley DJ: The two "hit" and multiple "hit" theories of carcinogenesis. **Br J Cancer** 23: 313–328, 1969
29. Moolgavkar SH, Venzon DJ: 2-event models for carcinogenesis—Incidence curves for childhood and adult tumors. **Math Biosci** 47: 55–77, 1979
30. Fialkow PJ: Clonal origin of human tumors. **Biochim Biophys Acta** 458: 283–321, 1976
31. Armitage P, Doll R, Pike MC: Somatic mutation (letter). **Br Med J** 1: 723–724, 1965
32. Pike MC: A method of analysis of a certain class of experiments in carcinogenesis. **Biometrics** 22: 142–161, 1966
33. Mayneord WV, Clarke RH: Carcinogenesis and radiation risk: A biomathematical reconnaissance. **Br J Radiol** (Suppl) 12: 1–112, 1975
34. Gumbel EJ: **Statistics of Extremes.** New York: Columbia University Press, 1958
35. Watson GS: Extreme values in samples from m-dependent stationary stochastic processes. **Ann Math Stat** 25: 798–800, 1954
36. Watson GS: Age incidence curves for cancer. **Proc Natl Acad Sci USA** 74: 1341–1342, 1977
37. Manton KG, Stallard E, Burdick D, Tolley HD: A stochastic compartment model of stomach cancer with correlated waiting time distributions. **Int J Epid** 8: 283–291, 1979
38. Manton KG, Stallard E: A two-disease model of female breast cancer: Mortality in 1969 among white females in the United States. **J Natl Cancer Inst** 64: 9–16, 1980
39. Cole WH: Opening address: Spontaneous regression of cancer and the importance of finding its cause. **Nat Cancer Inst Monogr** 44: 5–9, 1976
40. Stutman O: Immunodepression and malignancy. **Adv Cancer Res** 22: 261–422, 1975
41. Stutman O: Immunological surveillance. In **Origins of Human Cancer.** Hiatt HH, Watson JD, Winsten JA (Eds). Cold Spring Harbor Laboratory, 1977. pp. 729–750
42. Prehn RT: Tumor progression and homeostasis. **Adv Cancer Res** 23: 203–236, 1976
43. Maskens AP: Histogenesis and growth pattern of 1,2-dimethylhydrazine-induced rat colon adenocarcinoma. **Cancer Res** 1536: 1585–1592, 1976
44. Collins VP, Loeffler RK, Tivey H: Observations on growth rates of human tumors. **Am J Roentgenol** 76: 988–1000, 1956
45. Geddes DM: The natural history of lung cancer: A review based on rates of tumour growth. **Br J Dis Chest** 73: 1–17, 1979
46. Steel GG: **Growth Kinetics of Tumours.** Oxford: Clarendon Press, 1977

47. Chahinian P: Relationship between tumour doubling time and anatomoclinical features in 50 measurable pulmonary cancers. **Chest** 61: 340, 1972
48. Spratt JS, Ackerman LV: The growth of a colonic adenocarcinoma. **Am Surg** 27: 23–28, 1961
49. Charbit A, Malaise P, Tubiana M: Relation between the pathological nature and the growth rate of human tumors. **Eur J Cancer** 7: 307–315, 1971
50. Meyer JA: Growth rate versus prognosis in resected primary bronchogenic cancer. **Cancer** 31: 1468, 1973
51. Schwartz M: A biomathematical approach to clinical tumor growth. **Cancer** 14: 1272–1294, 1961
52. Weiss W: Peripheral measurable bronchogenic carcinoma: Growth rate and period of risk after therapy. **Am Rev Resp Dis** 103: 198, 1971
53. Wolfson R: **SEER data** (letter and data tape). Biometry Branch, National Cancer Institute, April 8, 1980
54. Hoel PG: **Introduction to Mathematical Statistics.** New York: John Wiley, 1971. pp. 324–401
55. Mendelsohn M: **Tumor Growth and the Cell Cycle.** Lecture, University of Colorado School of Medicine, February 18, 1977
56. Mulvihill JJ: Congenital and genetic diseases. In **Persons at High Risk of Cancer.** Fraumeni JF (Ed). New York: Academic Press, 1975. pp. 3–37
57. Crump KS: Dose response problems in carcinogenesis. **Biometrics** 35: 157–167, 1979
58. Burch PRJ: A biological principle and its converse: Some implications for carcinogenesis. **Nature** 195: 241–243, 1962
59. Knudson AG: Mutation and cancer: Statistical study of retinoblastoma. **Proc Natl Acad Sci** 68: 820–823, 1971
60. Knudson AG: Retinoblastoma: A prototype hereditary neoplasm. **Semin Oncol** 5: 57–60, 1978
61. Knudson AG, Strong LC: Mutation and cancer: A model for Wilm's tumor. **J Natl Cancer Inst** 48: 313–324, 1972
62. Knudson AG, Strong LC: Mutation and cancer: Neuroblastoma and pheochromocytoma. **Am J Hum Genet** 24: 514–532, 1972
63. Ashley DJ: Colonic cancer arising in polyposis coli. **J Med Genet** 6: 376–378, 1969
64. Ashley DJ: A systematic sex difference in intestinal carcinoma. **Cancer** 25: 966–971, 1970
65. Sutherland JV: **The Multihit Model of Carcinogenesis: Applications to Human Colon Cancer Incidence Data.** Ph.D. Thesis. University of Colorado School of Medicine: Department of Biometrics, 1980
66. Devesa SS, Silverman DT: Cancer incidence and mortality trends in the United States: 1935–74. **J Natl Cancer Inst** 60: 545–571, 1978
67. Welin S, Youker J, Spratt JS: The rates and patterns of growth of 375 tumors of the large intestine and rectum observed serially by double contrast enema study (Malmo technique). **Am J Roentgenol** 90: 673–687, 1963
68. Mood AM, Graybill FA, Boes DC: **Introduction to the Theory of Statistics.** New York: McGraw-Hill, 1974
69. Veale AMO: **Intestinal Polyposis. Eugenics Laboratory Memoirs XL.** London: Cambridge University Press, 1965
70. Reed TE, Neel JV: A genetic study of multiple polyposis of the colon. **Am J Hum Genet** 7: 236–263, 1955
71. Dukes CE: Familial intestinal polyposis. **Ann Eugen** 17: 1–29, 1952
72. Utsunomiya J, Murata M, Tanimura M: An analysis of the age distribution of colon cancer in adenomatosis coli. **Cancer** 45: 198–205, 1980
73. Asman HB, Pierce ER: Familial multiple polyposis, a statistical study of a large Kentucky kindred. **Cancer** 25: 972–981, 1970
74. Hill MJ, Morson BC, Bussey HJR: Aetiology of adenoma–carcinoma sequence in large bowel. **Lancet** 1: 245–247, 1978
75. Bussey HJR: **Familial Polyposis Coli.** Baltimore: Johns Hopkins University Press, 1975
76. Ashley DJ: Incidence and mortality of intestinal cancer. **Cancer** 25: 959–965, 1970
77. Lamerton LF: Cell population kinetics in normal and malignant tissues. In **Scientific Foundations of Oncology.** Symington T, Carter RL (Eds). William Heinemann Medical Books, 1976. pp. 119–125
78. Muto T, Bussey HJR, Morson BC: The evolution of cancer of the colon and rectum. **Cancer** 36: 2251–2270, 1975
79. Koppel M, Bailar JC III, Weakley FL, Shimkin MB: Incidence of cancer in the colon and rectum among polyp-free patients. **Dis Colon Rectum** 5: 349–355, 1962
80. Riddell RH: The precarcinomatous lesion of ulcerative colitis. In **The Gastrointestinal Tract.** Morson BC (Ed). Baltimore: Williams & Wilkins, 1977. pp. 109–123.
81. Correa P: Epidemiology of polyps and cancer. In **The Pathogenesis of Colorectal Cancer.** Morson BC (Ed). Philadelphia: WB Saunders, 1978. pp. 126–152
82. Morson BC: Evolution of cancer of the colon and rectum. **Cancer** 34: 845–849, 1974
83. Burkitt DP: Some neglected leads to cancer causation. **Natl Cancer Inst Monogr** 52: 5–12, 1979
84. Lynch HT, Brodkey FD, Lynch P, Lynch J, Maloney K, Rankin L, Kraft C, Swartz M, Westercamp T, Guirgis HA: Familial risk and cancer control. **JAMA** 236: 582–584, 1976