Taylor & Francis
Taylor & Francis Group

Evidence that natural immunity to breast cancer and prostate cancer exists in the majority of their risk populations is predicted by a novel, inherently saturated, ordered mutation model†

IVAN KRAMER*

Physics Department, University of Maryland Baltimore County, 1000 Hilltop Circle, Catonsville, MD 21250, USA

The series of *ordered* mutations that cause a specific cell to become cancerous is modeled so that the fraction of a risk population (*e.g.* White men) that has developed a specific cancer (*e.g.* melanoma) at any age can be calculated. The *saturated* model constructed and solved here is isomorphic to the physical model describing an ordered chain of radioactive nuclei decays with the exception that it allows for the possibility that a fraction of a risk population may be *immune* to developing a specific cancer.

The simplest model developed here depends on only three independent parameters: the number of ordered mutations necessary for a cell to become cancerous, the fraction of the risk population that is immune to developing a specific cancer and the average time between mutations (a time defined as the *mutation lifetime*). The values of these independent parameters are determined by fitting the model's cancer incidence function to the cancer incidence data.

This model was applied to five widely different cancers: melanoma, pancreatic cancer, female breast cancer, non-Hodgkin lymphoma and prostate cancer. The modeling predicts that *all* White males in the USA are vulnerable to developing melanoma, *five*-ordered mutations are required to develop it and the mutation lifetime is 48.3 years. By contrast, the modeling predicts that 80.7% of White females in the USA are *immune* to developing melanoma, *three*-ordered mutations are required to develop it and the mutation lifetime is 78.9 years. Remarkably, it was also found that about 70% of females are immune to developing breast cancer and about 70% of males are immune to developing prostate cancer, predictions that fit in with the experimental evidence of cancer immunosurveillance and immunoediting.

Clearly, different risk populations can develop the same cancer through different pathways. Delineating the mechanism underlying the prevalence of immunity to specific cancers in specific risk populations should become a research priority. Finding ways of blocking or repairing cellular mutations and/or destroying mutated, *potentially cancerous* cells would prevent cancers from developing altogether and eliminate a major cause of mortality.

Keywords: Cancer; Natural immunity prevalence; Immunosurveillance; Incidence rate; Saturated ordered mutation model; Immmunoediting

1. The saturated, ordered, mutation model of cancer development

Research data suggests that a cancerous cell results from an *ordered* series of *random* mutations. A potentially cancerous cell spontaneously mutates from one *mutation state* to another in a definite order until the final mutation causes the cell to become cancerous. Thus, for example, a normal, unmutated cell can only mutate into the first mutation state *and no other* and cells that are in the first mutation state can only mutate into the second mutation state *and no other*, etc. It is assumed that the characteristics of the cell in each mutation state are unique and different from the characteristics of the cell in every other mutation state. It is also assumed that a specific cancer results from an ordered series of mutations of a single cell of a specific organ.

Clearly, the number of mutations necessary to cause a specific cancer in a particular risk population is important to determine. Equally important to ascertain is the average time a cell spends in any particular mutation state, a quantity that will be called the mutation *lifetime* of the state. Perhaps, the most important thing to calculate from modeling the incidence rate of a specific cancer in a particular risk population is the prevalence of natural immunity to the cancer if, indeed, any exists. The model to be presented below will enable all three of these quantities to be calculated by fitting the model's cancer incidence function to cancer incidence data.

An important, novel feature of the cancer model is that it is inherently saturated, *i.e.* the maximum percentage of a risk population that can develop a particular cancer can never exceed 100%. Thus, the model's cancer incident rate function always increases, peaks and subsequently declines towards zero as the age of the risk population increases to arbitrarily large values (sometimes the peak lies *above* the maximum human lifespan).

The ordered, saturated, mutation model to be developed here will be fitted to the incidence data for five very different cancers: melanoma, female breast cancer, pancreatic cancer, non-Hodgkin lymphoma (NHL) and prostate cancer. Perhaps, the biggest surprise that results from the modeling is the prediction that significant natural immunity does indeed exist to breast and prostate cancers.

To construct a cancer model, the link between the mutation states of the cells of an organ must be connected to the risk states for individuals. To begin to be precise, it will be assumed that an ordered series of mutations, numbering m, of a particular cell of a particular organ is required for the cell to become cancerous (*e.g.* a skin cell and melanoma). At any age t, each cell of an organ of a person is in one of the mutation states. The cancer risk to the person will be gauged by the cell of the organ of the person *in the highest mutation state*, no matter how many cells are in this state. Thus, a person who has *at least one* cell in the pth mutation state ($p \leq m$) will be said to be in the pth mutation state. Thus, once a cell of an organ of a person mutates into the mth mutation state, the person develops cancer since it is assumed that only one cell is need to cause it.

Now consider a number of N_s *newly born* infants that represent a random sample of a risk population (e.g. White males) all of whom are susceptible to developing a particular cancer (e.g. melanoma). If the age of the infant cohort is denoted by t (birth is coincident with time $t = 0$), then the number of members of the cohort that are in the rth mutation state by age t will be denoted by $N(r/m, t)$ or $N_r(t)$ for short, where $r = 0, 1, 2,\ldots, m$. Assuming that all the infants are in the zeroth mutation state at birth, then at age $t = 0$, $N_r(0) = 0$ for $r = 1, 2, 3,\ldots, m$ and $N_0(0) = N_s$. A schematic representation of the sub-cohorts $N_p(t)$ at age t is shown in diagram 1 where the ordered mutations are represented by the arrows connecting two sequential mutation states.

Consider the first mutation necessary for a particular cell to eventually become cancerous. The fraction of the cohort with zero mutations that experiences the first mutation per unit time will be denoted by k_1 and called the first *mutation rate*. The average time required for

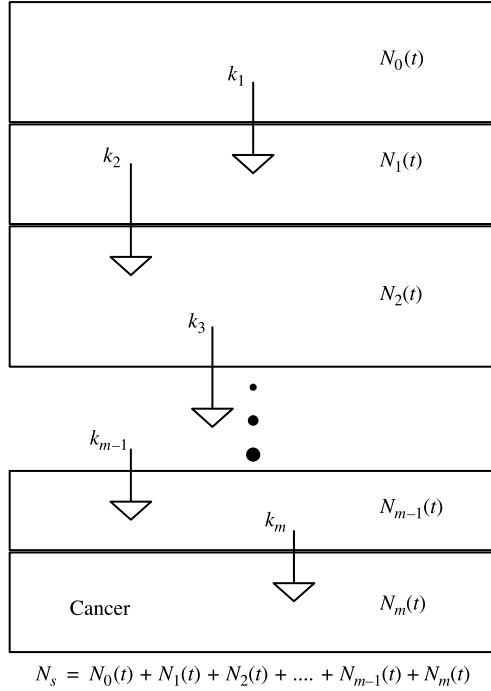$$N_s = N_0(t) + N_1(t) + N_2(t) + \ldots + N_{m-1}(t) + N_m(t)$$

Diagram (1). Mutation model leading to cancer

the cohort to experience the first mutation will be defined as the first *mutation lifetime* and will be denoted by T_1. The mutation lifetime and the mutation rate will be shown to be reciprocals of each other so that $k_1 \equiv T_1^{-1}$. Continuing in this way, the number of members of the cohort $N_r(t)$ who are in the rth mutation state at age t depends on the values of the r mutation rates $(k_r) \equiv (k_1, k_2, \ldots, k_r)$ as depicted in diagram 1. It will be assumed that all the mutation rates are constants so that the mutations experienced by the cohort *occur randomly*.

The appendix contains a complete description on how the functions $N_r(t)$ are computed from the model just described. One of the most important features of the model is that it allows the possibility that a fraction f_i of a risk population is *naturally immune* to a particular cancer. The fraction of the risk population that is susceptible to cancer f_s is related to the fraction that is immune to it since $f_s + f_i = 1$.

The most accurate way to directly measure the number of people in the populations $N_r(t)$ is to follow the cohort from birth and determine the number of cohort members who are in the rth mutation state at age t, where $r = 1, 2, 3, \ldots, m$. Aside from the fact that this experiment would last a lifetime to complete, this experiment is impractical for the simple reason that scientists do not know how to identify all the ordered mutations of a cell that cause it to become cancerous.

To acquire the data needed to determine the number of mutations m needed to generate a particular cancer, the mutation rates $(k_r) \equiv (k_1, k_2, \ldots, k_r)$ and the fraction f_i of the population that is naturally immune to developing cancer of a certain type, the age-related annual cancer incidence rate data can be used as a substitute for the ideal experiment discussed above. This substitution assumes that the biomedical profile of the general population at any fixed age extrapolated back in time is no different from the population of newborn infants. Of course, the population of any country is fluid, with immigration and

emigration occurring all the time. In using recorded cancer incidence statistics to determine the populations $N_r(t)$, it is assumed that the genetic profile of the country at any age is unchanged by immigration or emigration.

The model constructed in the appendix leads to a particular simple cancer incidence function if it is assumed that all of the mutation rates are identical (see equation (5c)). This simplified model was applied to five diverse cancers in turn. The USA cancer incidence statistics used in this paper comes from the Surveillance Epidemiology and End Results (SEER) data which can be found on the National Cancer Institute Website. The New York State (NYS) cancer incidence statistics comes directly from the NYS Department of Health Website.

2. Melanoma

Skin cancers are unique in the study of cancers in that *all* of the cells of this organ are observable at any time; thus, melanoma is a good place to start.

Figure 1a is a graph of SEER White male melanoma incidence rate data. To fit this data, the simplest model incidence function described in the appendix will be tried first, namely, that all the mutation rates are identical, equal to the common value k. In this case the trial melanoma incidence function involves only three independent parameters, m, k and f_s and is exceedingly simple. However, a least-squares fit to the data requires that the value of the susceptibility parameter f_s exceed unity, which is physically impossible. Thus, f_s was set equal to one so that all males are susceptible to developing melanoma and the subsequent least squares fit involved determining the values of only two parameters, m and k. Thus, carrying out a least-squares fit to the data in figure 1a using the incidence function in (5c) with $f_s = 1$ yields the excellent results also shown in figure 1a.

Given the simplicity of this three-parameter model function, the fit shown in figure 1a is remarkably good. It is found that five mutation are required to cause melanoma in the White male population ($m = 5$) and the average time between consecutive mutations (the mutation lifetime) is 48.3 years, the reciprocal of the mutation rate which was found to be $k = 0.02068\,\text{year}^{-1}$. The fact that a *single* mutation rate k could characterize all five mutations is significant since it implies that all the mutations may have a common cause.

Extrapolating the model incidence function obtained by the fit shows that the function peaks and monotonically declines towards zero as the age of the risk population increases. In this case, however, the peak and subsequent decline occurs at ages *above* that of the male human life-span. The incidence function in equation (5c) must peak and subsequently decline towards zero because the model assumed here is inherently *saturated*, *i.e.* the cumulative number of cancers of any type in a risk population at any age can never exceed the total number in the risk population. Using equation (5c) in the appendix, the peak in the incidence function occurs at $(5 - 1)/(0.02068\,\text{year}^{-1}) = 193$ years with the value of $404\,\text{year}^{-1}$ and monotonically declines thereafter.

The number of White males in a cohort of 100,000 who have accumulated four mutations (one less than the five mutations necessary to acquire melanoma) at age t is denoted by $N(4/5, t)$ and calculated from equation (5a) with $m = 5$ and $k = 0.02068\,\text{year}^{-1}$. The result is shown in figure 1b so that the pool of White males who are one mutation away from acquiring melanoma steadily increases with age. The number of White males who have come down with melanoma at age t, $N_m(t) \equiv N(5/5; t)$, is also plotted in figure 1b; from this curve it is clear that by the age of 82.5 years about 3% of White men will have developed melanoma, a result that agrees with published data.
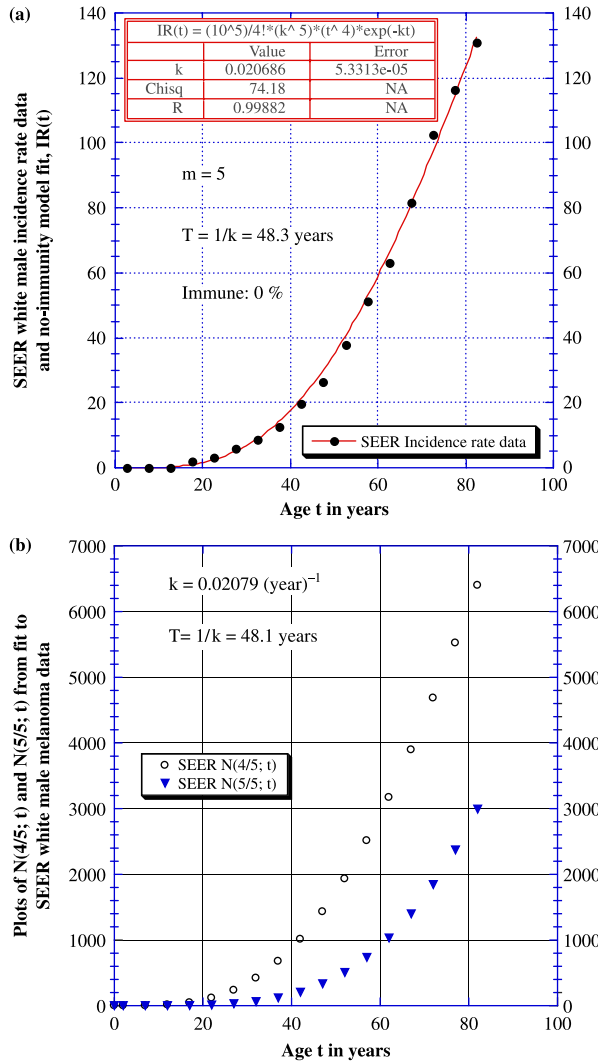
(a)

(b)

Figure 1. (a) Surveillance Epidemiology and End Results (SEER) White male melanoma incidence rate data and no-immunity model fit, IR(t). (b) Number of White males who have accumulated four mutations $N(4/5; t)$ and acquired melanoma (five mutations) $N(5/5; t)$ at age t per 100,000.

Figure 2a is a graph of SEER White *female* melanoma incidence rate data and this curve is clearly very different from the White male one shown in figure 1a. Thus, we expect very different model results for White women than what was obtained for White men. Clearly, the sex of the cohort has an important impact on the melanoma incidence rate.

To fit the White women melanoma incidence rate data, the simplest model incidence function described in the appendix will again be tried first, namely, one that assumes that all the mutation rates are the same. A credible fit to this data was impossible if the susceptibility parameter f_s is chosen to be one, so a fit was tried with three independent parameters: f_s, m and k. A least-squares fit to the data in figure 2a yields the results also shown in figure 2a. Remarkably, it is found that 80.7% of White females are *immune* to acquiring melanoma

(a)

(b)

(c)

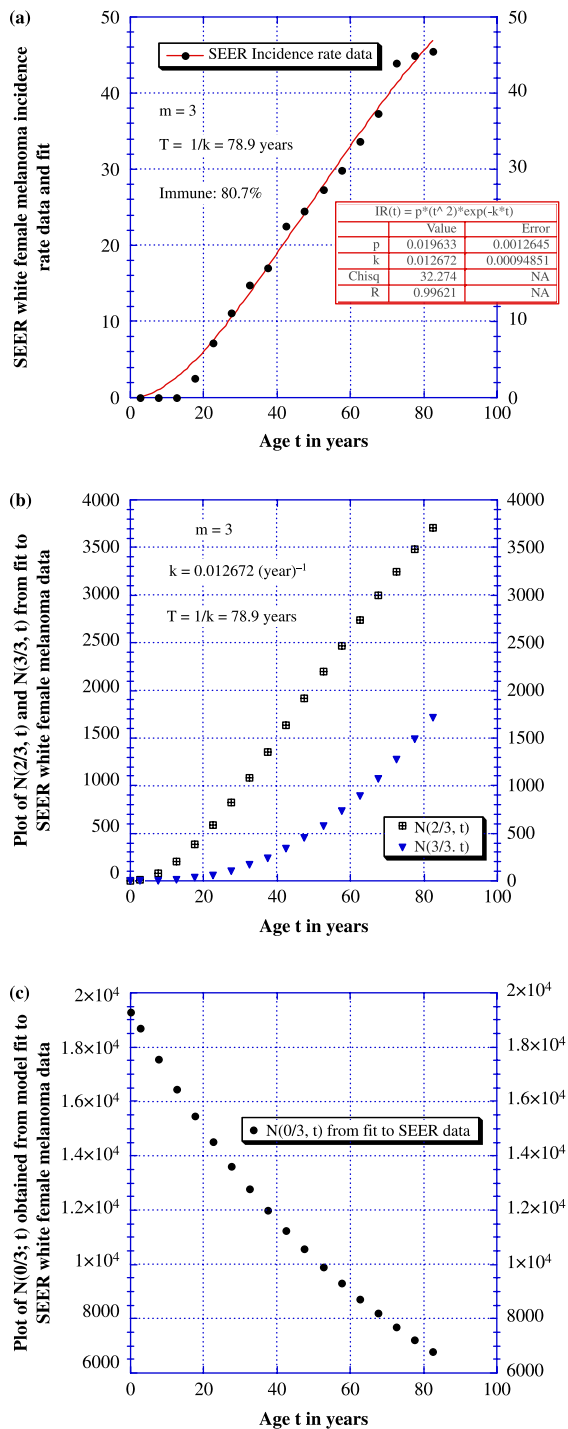Figure 2. (a) SEER White female melanoma incidence rate data and immunity model fit, IR(*t*). (b) Number of White females who have accumulated two mutations *N*(2/3; *t*) and acquired melanoma (three mutations) *N*(3/3; *t*) at age *t* per 100,000. (c) Model prediction of the number of USA White females susceptible to developing melanoma that have not experienced a single mutation *N*(0/3; *t*) at age *t* per 100,000.

($f_s = 0.1929$), three mutations are required to cause it ($m = 3$) and the average time between consecutive mutations (the mutation lifetime) is 78.9 years.

The number of White females in a cohort of 100,000 who have accumulated two mutations (one less than the number necessary to acquire melanoma) at age t is denoted by $N(2/3, t)$ for clarity and calculated from equation (5a) in the appendix with $m = 3$ and $k = 0.01267\,\text{year}^{-1}$. The result is shown in figure 2b so that the pool of White females who are one mutation away from acquiring melanoma steadily increases with age. The number of White females who have come down with melanoma at age t, $N_3(t) \equiv N(3/3, t)$, is also plotted in figure 2b and from this curve it is clear that by the age of 82.5 years about 1.7% of White females will develop melanoma, a result that also agrees with published data.

A plot of the number of *melanoma-susceptible* White females who have *not acquired a single mutation* at age t in a random sample of 100,000 White females, a function defined as

Table 1. Summary of model parameters stemming from fits.

Cohort	Figure/model[†]	m	k year^{-1}	$T = k^{-1}$ years	Percent immune (%)	Fit error $(\chi^2)^{‡}$ years^{-2}
(1) Melanoma						
SEER White male	Figure 1a/NI	5	0.02068	48.3	0	74.1
	I		Physically impossible ($f_s > 1$)			
SEER White female	Figure 2a/I	3	0.01267	78.9	80.7	32.2
	NI	3	0.006347	157.5	0	120.9
(2) Pancreatic cancer						
NYC male	NI	8	0.03884	25.7	0	75.2
	Figure 3a/I	8	0.04100	24.3	23.5	68.6
Upper NYS male	NI	8	0.03898	25.6	0	127.0
	Figure 3b/I	8	0.04353	22.9	41.5	94.7
NYC female	NI	9	0.04483	22.3	0	59.9
	Figure 4a/I	9	0.04536	22.0	6.29	59.7
Upper NYS female	Figure 4b/NI	8	0.03768	26.5	0	90.2
	I		Physically impossible ($f_s > 1$)			
SEER female	NI	8	0.03829	26.1	0	114.1
	Figure 4c/I	8	0.04294	23.2	43.0	86.1
(3) Breast cancer						
SEER White female	Figure 5a/I	11	0.1307	7.65	70.8	610.1
	NI	5	0.03407	29.3	0	16,929
SEER Black female	Figure 5b/I	9	0.1021	9.79	72.5	696.4
	NI	5	0.03189	31.3	0	11,615
NYS female	Figure 5c/I	9	0.1008	9.92	69.0	707.8
Upper NYS female	Figure 5d/I	10	0.1165	8.58	69.5	868.7
NYC female	Figure 5e/I	8	0.08539	11.7	69.3	1,143
(4) Non-Hodgkin lymphoma						
NYS male	Figure 6a/CNI	2	0.001359	735	0	51.5
		7	0.03265	30.6	0	
NYS female	Figure 6b/CNI	3	0.004145	241	0	93.2
		7	0.03019	33.1	0	
(5) Prostate cancer						
NYS male	Figure 7a/I	40	0.5314	1.88	69.9	1161
	NI	11	0.1017	9.82	0	231,080
NYC male	Figure 7b/I	40	0.5345	1.87	71.8	878.0
Upper NYS male	Figure 7c/I	40	0.5299	1.88	68.8	1,392

[†] Immunity (I), no-immunity (NI), or composite no-immunity (CNI) model.
[‡] See equation (5f) in the appendix.

$N(0/3, t)$ here, is plotted in figure 2c. According to the model fit, in every random sample of 100,000 White female infants there are 19,296 who are susceptible to acquiring melanoma; by age 82.5 years about 6784 of these women *have still not acquired a single mutation* that would lead to melanoma. These women are certainly not immune to acquiring melanoma, they have just been fortunate in not acquiring a single mutation. Moreover, in 100,000 White female infants, 80,703 are predicted by the modeling to be actually *immune* to acquiring all the mutations necessary to cause melanoma. It may be impossible today to tell the difference between females that are naturally immune to melanoma and those that have not experienced a single mutation by sheer luck, but if melanoma biology progresses to a point where such a distinction can be ascertained, this model prediction could then be tested.

To test the credibility of the immunity model fit in figure 2a, a fit to the same data assuming that all females are susceptible to developing melanoma was executed. Setting $f_s = 1$, the best fit found that $m = 3$ (again), $k = 0.00634$ year^{-1} (about half the value returned by the immunity model fit), but the fit error (χ^2) was four times that of the immunity model fit (see (5f) in the appendix). Although the no-immunity model does not return as credible a fit to the data as does the immunity model, the disparity in fit errors is not large enough to prove that immunity to melanoma exits in White females.

All of the model fit results discussed above and to follow are summarized in table 1 for convenience. Clearly, the saturated, ordered mutation model constructed here yields results about melanoma that are novel and important. Although there are differences in the values of the model parameters m, k and f_s between the White male and female cohorts, the mutation model successfully describes their melanoma incidence functions.

Since different mutation numbers m are required to fit male and female melanoma incident data, different risk populations can develop melanoma through *different pathways*.

The only conclusive way to determine the precise mutations necessary to produce any particular cancer is to complete population-based genetic association studies involving cancer patients and non-cancer controls for specific risk populations (males, females, etc.). Such studies have already been undertaken for the major cancers.

Is this model credible?

In 2002 it was discovered that the mutation of the BRAF gene (the gene that makes the protein called B-RAF) is the key mutation causing most melanoma [1,2]; the discovered mutation renders BRAF cancer gene (oncogene) deaf to control signals. Normally, BRAF is part of a series of genetic switches that must be "on" (via mutations) for a cell to uncontrollably divide.

There are three different types of genes within a cell that are important in determining whether a cell becomes cancerous or not: (1) genes that signal a cell to multiply, (2) genes that stop a cell from multiplying (*e.g.* the p53 tumour suppressor gene) and (3) genes that repair other damaged (mutated) genes.

Clearly, the *order* of these mutations is important. For a cell to become cancerous, the gene that repairs the DNA damage to the cell's other genes must first itself suffer a damaging mutation.

Before or after this mutation, a damaging mutation must take place in the gene that stops cell division so that when a cell becomes cancerous it will remain so. Thus, it would appear that a *minimum* of three ($m = 3$) *ordered* mutations are required to cause a normal cell to become cancerous, the same number found here to cause melanoma in White females. An infant may be born with one or more of these mutations thereby developing the related cancer sooner, on the average, than an infant born with none of these mutations. Thus, the ordered mutation model for melanoma constructed here fits this minimum mutation model for cancer development.

The immunity to melanoma by about 80% of White females in the USA predicted by the modeling here can be ascribed to an ability in these females to rapidly (1) repair a precancerous mutation before the next ordered mutation can occur or (2) recognize and destroy a precancerous mutated skin cells before the next one can occur.

What triggers the ordered mutations in the oncogenes of a cell that cause the cell to become cancerous?

In the model fits to the melanoma incidence curves discussed above each mutation takes place with the *same* mutation rate. This implies that the *same* cause or series of causes *could be* responsible for each and every mutation, such as factor(s) in the general environment, including antigens. As an example, it is possible that each mutation is triggered by a particular virus whose principle effect is to change or mutate the DNA of a specific gene in the target cell it infects. If these viruses exist in the general environment and if infection with these viruses occur randomly with the same probability of infection in susceptible individuals over time, it could account for the results of the melanoma modeling obtained here.

3. Pancreatic cancer

The model fit to the incidence of pancreatic cancer will now be addressed. Although there is a long list of plausible potential causes for melanoma, the potential causes of pancreatic cancer are more speculative.

The analysis here will consider the incidence of pancreatic cancer in risk populations living in different regions of NYS to test the dependence of this disease on the sex, genetic profile and life style of the cohort.

The immunity model fit to the pancreatic cancer incident rate data for New York City (NYC) males is shown in figure 3a. This excellent fit leads to the conclusion that eight mutations are required to cause pancreatic cancer in males on the average, the mutation rate is $k = 0.04100 \, \text{year}^{-1}$ and 23.5% of males are immune to developing pancreatic cancer, results that are included in figure 3a. Using this value for the mutation rate, the corresponding mutation lifetime is $T = 24.3$ years.

Assuming that all males are susceptible to developing pancreatic cancer by setting $f_s = 1$ and then re-executing the fit with only two parameters (m and k) yields similarly excellent results but with a slightly larger fit error; again $m = 8$, but k has slightly decreased in value. Since both of these fits are excellent, neither can be ruled out as being realistic and we can only conclude that the percentage of NYC males that are immune to pancreatic cancer lies *anywhere* in the 0–23.5% range, *including the 0% possibility*.

The immunity model fit to the pancreatic cancer incident rate data for *upper* NYS males, *excluding those that live in NYC*, is shown in figure 3b, while the *no*-immunity model fit ($f_s = 1$) to the same data gives a similar result but with a 34% increase in fit error. As before, both fits give $m = 8$ but slightly differ in their values of k. Since both fits are very good, we can only conclude that the percentage of upper NYS males that are immune to pancreatic cancer lies *anywhere* in the 0–41.5% range, *including the 0% possibility*.

The immunity model fit to the pancreatic cancer incident rate data for NYC *females* is shown in figure 4a. This excellent fit leads to the conclusion that nine mutations are required to cause pancreatic cancer in NYC females, the mutation rate is $k = 0.04536 \, \text{year}^{-1}$ and only 6.29% of this risk population are immune to this cancer, results that are included in figure 4a. Thus, the values of m and k for NYC females are slightly larger than their corresponding values for NYC males. Fitting the no-immunity model to this same data leads to virtually
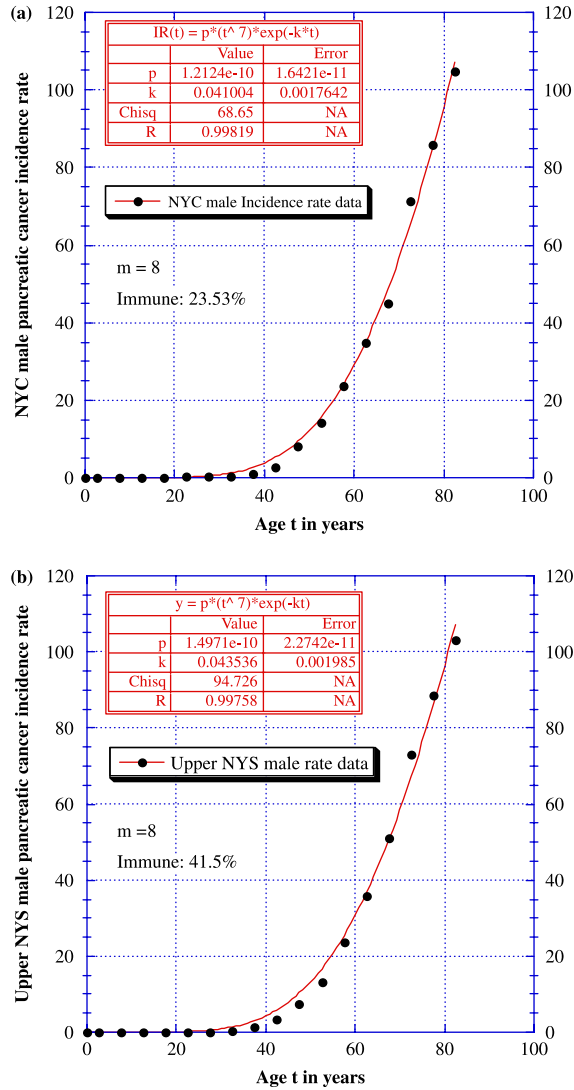
Figure 3. (a) New York City (NYC) male pancreatic cancer incidence rate data and immunity model fit. (b) Upper New York State (NYS) male pancreatic cancer incidence rate data and immunity model fit.

identical results with virtually the same fit error as in figure 4a (see table 1). Thus, no claim of immunity to pancreatic cancer can be made for NYC females.

The immunity model fit to the pancreatic cancer incident rate data for *upper* NYS *females*, excluding those that live in NYC, leads to a value for the susceptibility constant f_s that exceeds 1, a physical impossibility. Thus, setting $f_s = 1$, the no-immunity model fit to the data is shown in figure 4b. As for NYS males, the excellent fit here leads to the conclusion that all members of this cohort are susceptible to this cancer and that eight mutations are required to cause it. The value of the mutation rate for upper NYS females turned out to be $k = 0.03768 \, \text{year}^{-1}$, extremely close to its value for NYC males. The modeling results for this cohort are also included in figure 4b.

(a)

IR(t) = p*(t^ 8)*exp(-kt)		
	Value	Error
p	1.892e-12	3.1423e-13
k	0.045367	0.0021408
Chisq	59.714	NA
R	0.99793	NA

NYC female incidence rate data

m = 9

Immune: 6.29%

(b)

IR(t) = (10^5)*(k^ 8)*(t^ 7)*exp(-kt)/7!		
	Value	Error
k	0.037685	0.00011935
Chisq	90.255	NA
R	0.99684	NA

Upper NYS female rate data

m = 8

Immune: 0%

(c)

IR(t) = p*(t^ 7)*exp(-kt)		
	Value	Error
p	1.308e-10	2.0869e-11
k	0.042946	0.0020832
Chisq	86.119	NA
R	0.99738	NA

SEER Incidence rate data
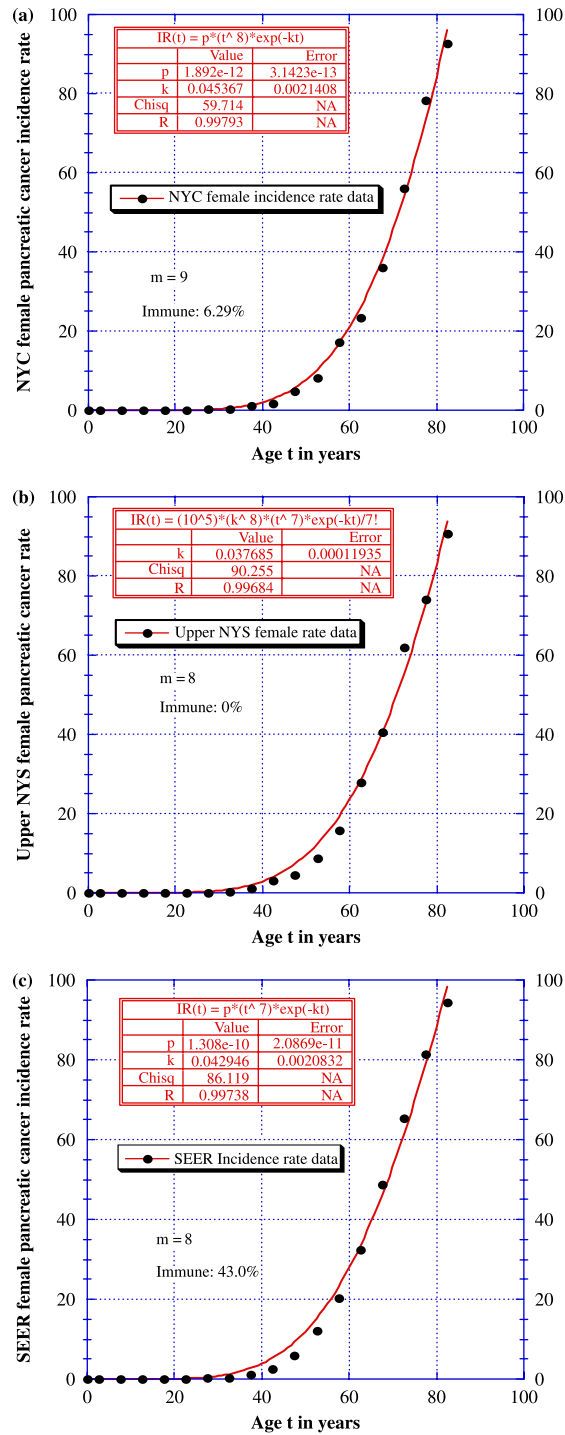
m = 8

Immune: 43.0%

Figure 4.   (a) NYC female pancreatic cancer incidence rate data and immunity model fit. (b) Upper NYS female pancreatic cancer incidence rate data and no-immunity model fit. (c) SEER female pancreatic cancer incidence rate data and immunity model fit.

Fitting the immunity model to the SEER (USA) female pancreatic cancer incidence rate data yields the results shown in figure 4c; this fit predicts that 43.0% of USA females are immune to developing pancreatic cancer, an immunity not predicted from the NYS data fit (see above). The no-immunity model gives similar results (both predict that $m = 8$), but with a 32% increase in fit error (see table 1).

Viewing the results of modeling the pancreatic cancer incidence rate data for the five different risk populations as a whole leads to the conclusion that the incidence of pancreatic cancer is generally independent of the sex, genetic profile and geographical location of the cohort. This cancer is most likely caused by $m = 8$ mutations and a sizable fraction of both males and females *may* have immunity to developing it.

The exact cause(s) of pancreatic cancer is not known although there is a list of known risk factors, such as cigarette smoking and chronic pancreatitis, which can increase the risk of acquiring it. To date, the results of a genome-wide association study of prostate cancer patients and controls has not appeared in the literature.
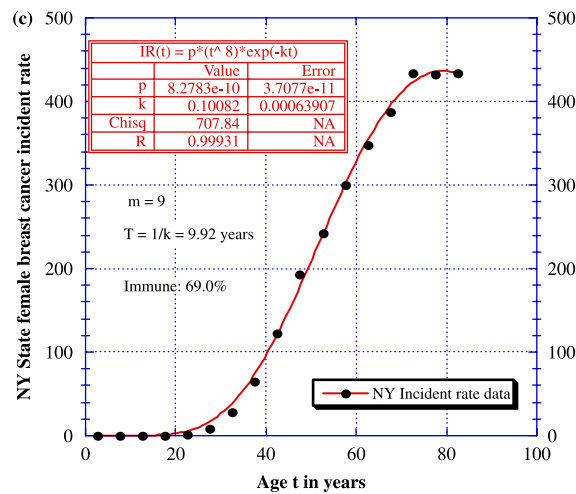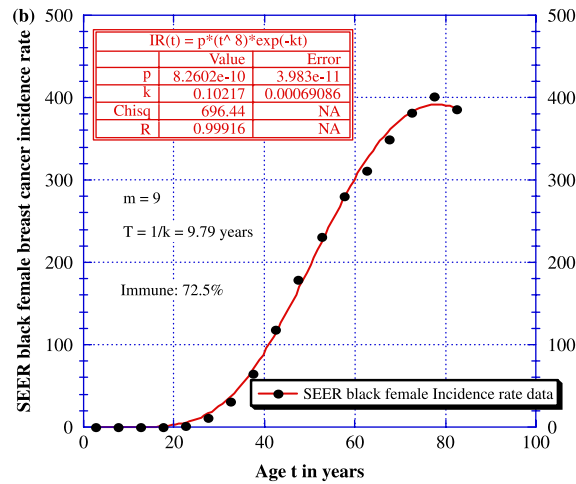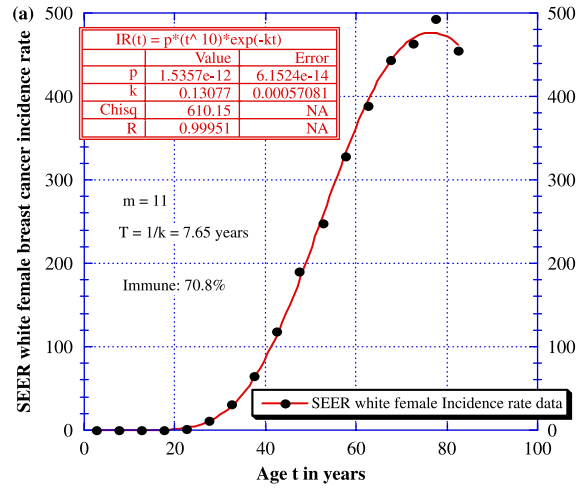
4. Female breast cancer

In this section the models developed here will be applied to the breast cancer incidence rate data for five different female risk populations: USA Whites (SEER data), USA Blacks (SEER data), NYS, upper NYS and NYC.

The model fit to the breast cancer incident rate data for USA White females (SEER data) is shown in figure 5a. The fit leads to the conclusion that 70.8% of the cohort is *immune to developing* this cancer, 11 mutations are required to cause it and the mutation rate is $k = 0.1307\,\text{year}^{-1}$, results that are included in figure 5a. Using this value for the mutation rate, the corresponding mutation lifetime is 7.65 years. As seen in figure 5a the incidence rate curve peaks at age 78.4 years and approaches zero thereafter, an inherent feature of the *saturated* model developed here. Since the modeling yields a susceptible fraction of $f_s = 0.29139$, out of 100,000 White USA females, only 29,139 are susceptible to developing breast cancer; however, only about half of these vulnerable females will develop this cancer because the average human lifetime is less than 100 years. In other words, the majority of females susceptible to developing breast cancer will not develop it because they will die of something else first.

To test the credibility of the immunity model results above for USA White women, the no-immunity model ($f_s = 1$) was used to fit the same data and the results are shown in table 1. Clearly, the no-immunity model must be rejected since it incurs a fit error (χ^2) of more than 27 times that of the immunity model.

The corresponding model fit to the breast cancer incident rate data for USA *Black* females is shown in figure 5b. The fit leads to the conclusion that 72.5% of Black females are *immune to developing* this cancer, nine mutations are required to cause it and the mutation lifetime is 9.79 years, results that are included in figure 5b. Thus, although about the same percentage of White and Black women are immune to developing breast cancer according to the model fits, the values of m and k for these two cohorts are somewhat different.

To test the credibility of the immunity model results for USA Black women, the no-immunity model was used to fit the same data and the results are shown in table 1. Again, the no-immunity model must be rejected since it incurs a fit error (χ^2) of more than 16 times that of the immunity model.
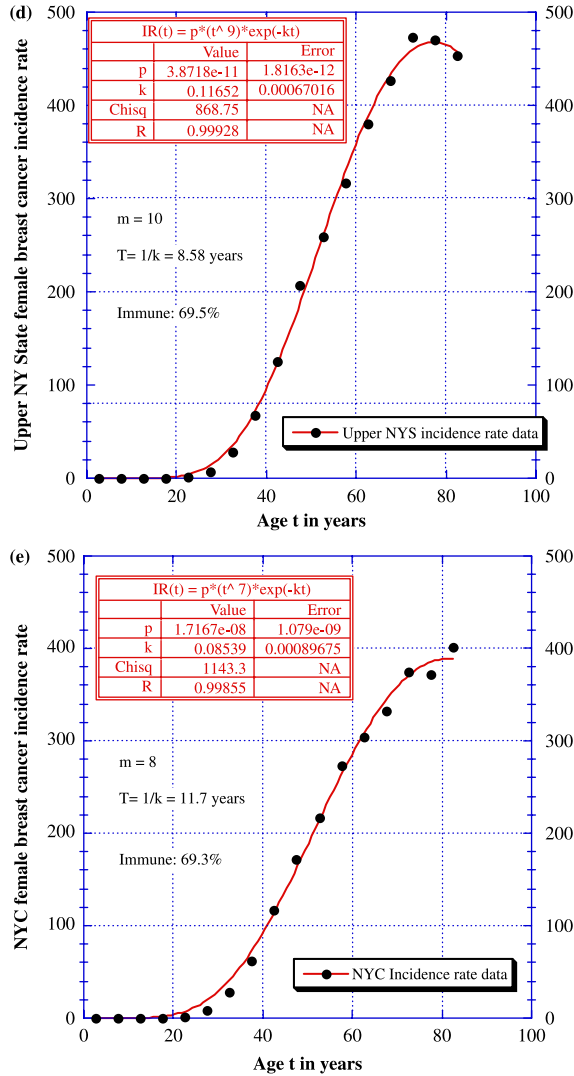
(a)

IR(t) = p*(t^ 10)*exp(-kt)		
	Value	Error
p	1.5357e-12	6.1524e-14
k	0.13077	0.00057081
Chisq	610.15	NA
R	0.99951	NA

m = 11

T = 1/k = 7.65 years

Immune: 70.8%

SEER white female Incidence rate data

SEER white female breast cancer incidence rate

Age t in years

(b)

IR(t) = p*(t^ 8)*exp(-kt)		
	Value	Error
p	8.2602e-10	3.983e-11
k	0.10217	0.00069086
Chisq	696.44	NA
R	0.99916	NA

m = 9

T = 1/k = 9.79 years

Immune: 72.5%

SEER black female Incidence rate data

SEER black female breast cancer incidence rate

Age t in years

(c)

IR(t) = p*(t^ 8)*exp(-kt)		
	Value	Error
p	8.2783e-10	3.7077e-11
k	0.10082	0.00063907
Chisq	707.84	NA
R	0.99931	NA

m = 9

T = 1/k = 9.92 years

Immune: 69.0%

NY Incident rate data

NY State female breast cancer incident rate

Age t in years

(d)

IR(t) = p*(t^ 9)*exp(-kt)

	Value	Error
p	3.8718e-11	1.8163e-12
k	0.11652	0.00067016
Chisq	868.75	NA
R	0.99928	NA

m = 10

T= 1/k = 8.58 years

Immune: 69.5%

Upper NYS incidence rate data

(e)

IR(t) = p*(t^ 7)*exp(-kt)

	Value	Error
p	1.7167e-08	1.079e-09
k	0.08539	0.00089675
Chisq	1143.3	NA
R	0.99855	NA

m = 8

T= 1/k = 11.7 years

Immune: 69.3%

NYC Incidence rate data

Figure 5.   (a) SEER White female breast cancer incidence rate data and immunity model fit. (b) SEER Black female breast cancer incidence rate data and immunity model fit. (c) NYS female breast cancer incident rate data and immunity model fit. (d) Upper NYS female breast cancer incidence rate data and immunity model fit. (e) NYC female breast cancer incidence rate data and immunity model fit.

The model fit to the breast cancer incident rate data for NYS females is shown in figure 5c. The fit leads to the conclusion that 69.0% of the cohort are *immune to developing* this cancer, nine mutations are required to cause it and the mutation lifetime is 9.92 years, results that are included in figure 5c. These results are in general agreement with the SEER results for USA White and Black females above.

The model fit to the breast cancer incident rate data for *upper* NYS females, excluding NYC females, is shown in figure 5d. The fit leads to the conclusion that 69.5% of the cohort are *immune to developing* this cancer, 10 mutations are required to cause it and the mutation

lifetime is 8.58 years, results that are included in figure 5d and which are close to that of NYS as a whole.

The model fit to the breast cancer incident rate data for NYC females is shown in figure 5e. The fit leads to the conclusion that 69.3% of this risk population are *immune to developing* this cancer, eight mutations are required to cause it and the mutation lifetime is 11.7 years, results that are included in figure 5e. The differences in the values of the mutation number m and the mutation rate k between NYC females and upper NYS females can be attributed to the expected differences in the genetic profiles of these two risk populations and differences in their environments.

Viewing the results of the breast cancer modeling for the five different risk populations considered as a whole (see table 1) leads to the predictions that around 70% of females are *immune* to developing breast cancer, somewhere between 8 and 11 mutations are needed to cause it *on the average* and the *average* value of the mutation lifetime lies somewhere between 7.65 years and 11.7 years. Different female risk populations may develop breast cancer through different pathways; these different pathways may account for the differences in the values of the model parameters that describe their cancer incidence rates.

The first gene associated with breast cancer, BRCA1 (for breast cancer), was identified on chromosome 17 in 1994. A year later a second gene associated with breast cancer (BRCA2) was discovered on chromosome 13. Since not all hereditary breast cancers are associated with BRCA1 and BRCA2, it is currently thought that this cancer can be caused by mutations in many other (unidentified) genes. To date, the results of a comprehensive genome-wide association study of breast cancer patients and controls has not appeared in the literature.

The exact causes of the ordered mutations that induce a breast cell to become cancerous are not known.

5. Non-Hodgkin lymphoma

NHL is a collection of about 35 different, distinct cancers of the cells of the immune system (B-cells, T-cells and NK-cells). The incidence rates of these distinct cancers for a given risk population could be *different*, so that the composite incidence rate curve for NHL may look different from that of the individual lymphomas that compose it. The SEER incidence data for NHL is a fairly sensitive function of the gender and race of the cohort.

Figure 6a contains the NYS NHL incidence rate data for *males*. A least-squares fit to the early data points (below 22.5 years of age) using the *no-immunity* model gives a mutation number of $m_1 = 2$, whereas an identical fit to the remaining data points gave a mutation number of $m_2 = 7$. Thus, a fit to all the data in figure 6a was tried by the composite incidence function $I(t) \equiv IR(m_1, k_1, t) + IR(m_2, k_2, t)$ where k_1 and k_2 are the only independent parameters of the fit. As can be seen from figure 6a, an excellent fit to the data was obtained using $k_1 = 0.001359 \, \text{year}^{-1}$ and $k_2 = 0.03265 \, \text{year}^{-1} = 24.0 \, k_1$ with all males assumed to be susceptible to acquiring NHL.

A similar fit to NYS *female* NHL data using $m_1 = 3$ and $m_2 = 7$ gave the excellent results shown in figure 6b with, $k_1 = 0.004145 \, \text{year}^{-1}$ and $k_2 = 0.03019 \, \text{year}^{-1} = 7.28 \, k_1$, again assuming that all females are susceptible to acquiring NHL.

Thus, for both males and females there are two different pathways to acquiring NHL, one that involves either two or three mutations and one that involves seven mutations. It is important to note that the $m_1 = 2$ or 3 pathway occurs at relatively early ages. These two distinct pathways could be describing two genetically different risk populations, one born with a predisposition to

developing NHL and one that is not. However, since our results show that $k_1 \ll k_2$, it is more likely that these pathways are describing different sets of NHLs in the collection.

Neither of the fits in figure 6a and b would be materially improved by using the immunity model for the fits. Thus, there is no evidence in this modeling that there is any immunity to NHL in the population.

NHLs result from somatic translocations in the genes of lymphoid cells. The H2AFX gene encodes a key histone involved in the detection of DNA double-stranded breaks that can lead to these translocations. A population-based genetic association study of H2AFX variants in 487 NHL cases and 531 controls supports the hypothesis that genetic variation in the H2AFX gene can lead to the acquisition of NHL by destroying genetic stability [3].
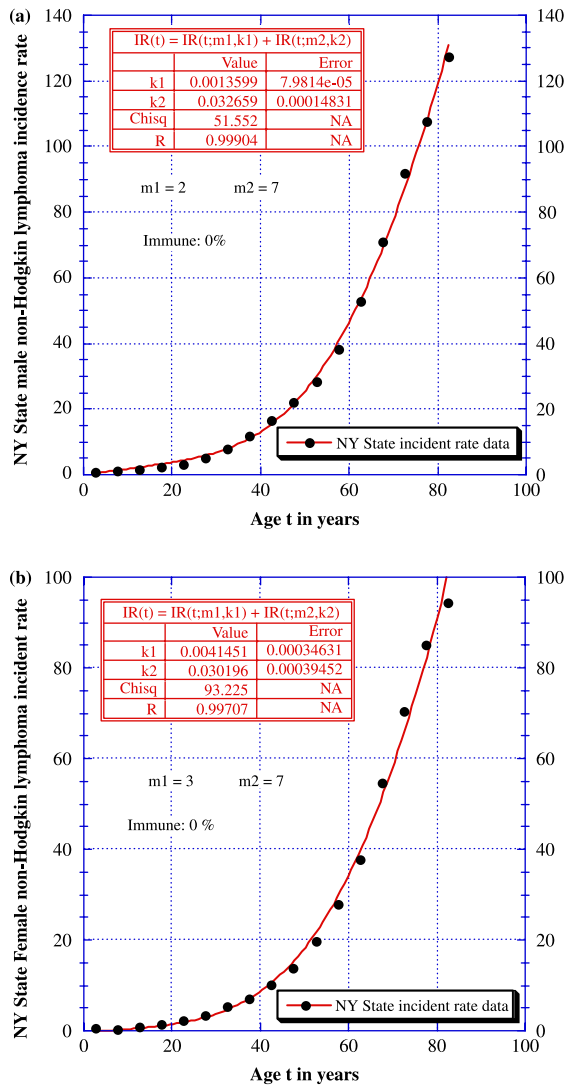

Figure 6.   (a) NYS male non-Hodgkin lymphoma (NHL) incidence rate data and no-immunity model fit. (b) NYS female NHL incidence rate data and no-immunity model fit.

6. Prostate cancer

The immunity model fit to NYS prostate cancer incidence data is shown in figure 7a. This excellent fit predicts that 69.9% of NYS men are immune to developing prostate cancer, 40 mutations are required to develop it and the mutation lifetime is only 1.88 years.

To test whether these results significantly change as the risk population changes, the identical fit was executed for both exclusively NYC prostate data (see figure 7b) and Upper NYS data, excluding NYC data (see (7c)). Both of these fits are also excellent and return values for the model parameters that are virtually identical to those of NYS as a whole shown in figure 7a. The values of the model parameters obtained for each of these three fits are shown in the respective figures and in table 1. From these results it is clear that the development of prostate cancer is largely independent of race and environment.

The results of the no-immunity model fit to the NYS prostate cancer data is shown in table 1. With a fit error 199 times that of the immunity model fit, the no-immunity model must be rejected. The no-immunity model similarly fails to credibly fit the data in figures 7b and c for NYC and Upper NYS, respectively. These negative results strengthen the credibility of the results of the immunity model.

Are these results for prostate cancer credible?

Direct proof that mutations in DNA can cause prostate cancer comes from an interesting controlled experiment by Li *et al*. [4] who exposed a colony of benign human epithelial cells to a single exposure of a 0.6 Gy dose of alpha particles (1 Gy ≡ 100 rads). Malignant cells emerged from the radiated sample while none did in the unirradiated controls. Compared to the controls, prominent changes in chromosome 6, 11 and 16, as well as deletions of the p53 gene were observed in the tumour outgrowth and tumour cells.

To determine what fraction of the risk of various cancers is attributable to genetic factors, a study of 44,788 Scandinavian twins was conducted [5]. The study concluded that 42% of the risk to developing prostate cancer was due to heritable factors, with the majority of the risk dependent on environmental factors. In fact this study concluded that environmental factors played the dominant role in causing *all* sporadic cancers.

A genome-wide association study of 3090 sporadic prostate patients and controls was undertaken by Nam *et al*. [6]. The study involved 40 prostate cancer patients and 40 controls and found that there were a total of 237 single nucleotide polymorphisms (SNPs) associated with prostate cancer. Significant associations of prostate cancer with two allele risk haplotypes on both chromosome 1q25 and chromosome 7p21 were found.

The roles of chromosome translocations and other aberrations as well as diet in causing prostate cancer are undergoing intense scrutiny [7–10].

To date, the exact number of mutations necessary to cause prostate cancer has yet to be determined, but with 237 SNPs to work with, the prediction of 40 mutations by the ordered mutation model here is certainly possible.

7. Discussion

A summary of all of the above model fit results appears in table 1.

Unexpectedly, the modeling predicts that about 70% of females are immune to acquiring breast cancer and up to 80% of White females *may be* immune to developing melanoma. The modeling also predicts that prostate cancer incidence is largely independent of race and environment, about 70% of men are immune to developing it, 40-ordered mutations are

(a)

IR(t) = p*(t^ 39)*exp(-kt)		
	Value	Error
p	1.5358e-53	7.4258e-55
k	0.53142	0.00066982
Chisq	1161.5	NA
R	0.99977	NA

NY State incidence data

m = 40

T = 1/k = 1.88 years

Immune: 69.9%

NY State prostate cancer incidence rate

Age t in years

(b)

IR(t) = p*(t^ 39)*exp(-kt)		
	Value	Error
p	1.8154e-53	8.0011e-55
k	0.53454	0.00061294
Chisq	878.05	NA
R	0.9998	NA

NY City incidence data

m = 40

T = 1/k = 1.87 years

Immune: 71.8%

NY City prostate cancer incidence rate

Age t in years

(c)

IR(t)= p*(t^ 39)*exp(-kt)		
	Value	Error
p	1.4287e-53	7.3595e-55
k	0.52998	0.00071235
Chisq	1392	NA
R	0.99974	NA

Upper NY State incidence data

m = 40

T = 1/k = 1.88 years

Immune: 68.8%

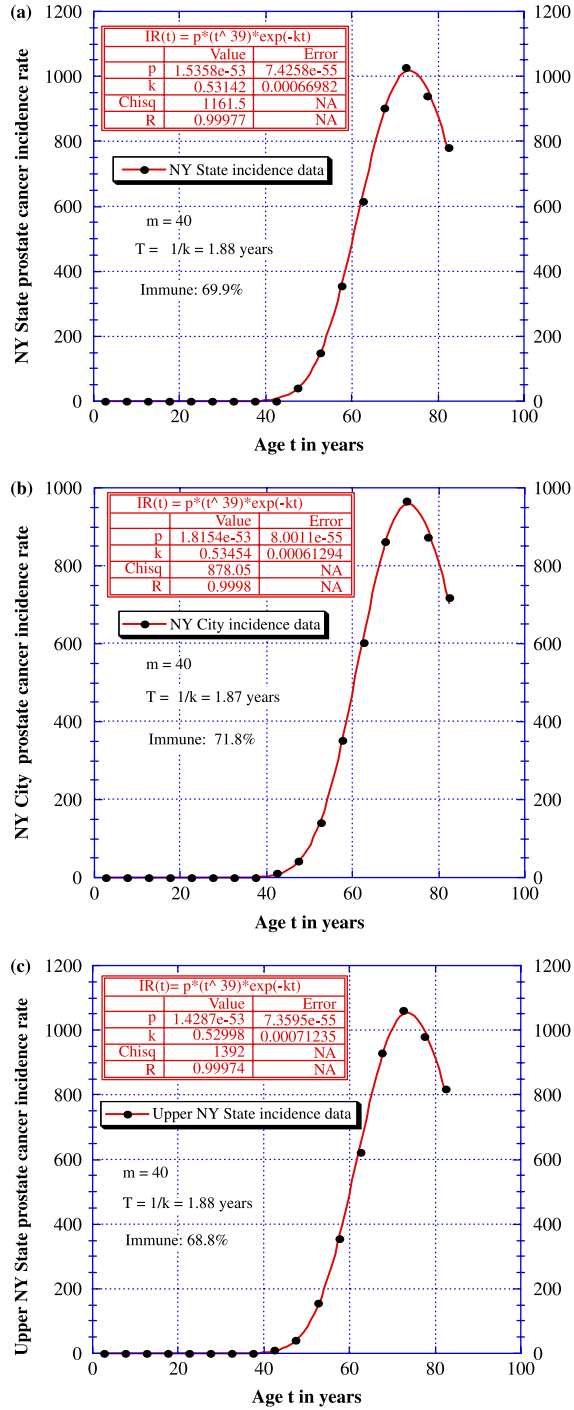**Upper NY State prostate cancer incidence rate**

Age t in years

Figure 7.   (a) NYS prostate cancer incidence rate data and immunity model fit. (b) NYC prostate cancer incidence rate data and immunity model fit. (c) Upper NYS prostate cancer incidence rate data and immunity model fit.

required to cause it and that the mutation lifetime is only 1.88 years. The modeling also suggests that somewhere between 0–43% of the population *may* be immune to developing pancreatic cancer although it should be emphasized that the modeling is consistent with there being no immunity to this cancer. The NHL mutation rate associated with the $m = 7$ pathway is very close to the pancreatic cancer mutation rate suggesting that the cause(s) of the mutations leading up to these cancers may be the same.

The immunity to melanoma and breast cancer in the majority of White females and the immunity to prostate cancer in the majority of men in the USA predicted by the modeling can be generated by an ability to either (1) rapidly repair a mutation to a potentially cancerous cell or to (2) destroy a mutated cell before the next ordered mutation can occur.

Proof that the immune system helps to control the incidence rate of a variety cancers stems from the fact that those infected with human immunodeficiency virus (HIV) experience increased incidences of Karposi's sarcoma, non-Hodgkin's lymphoma and invasive cervical cancer [11–16]. Similarly, patients who start taking immunosuppressive drugs after receiving an organ transplant begin experiencing a higher incidence rate of cancers; conversely, withdrawing immunosuppressive drugs from a patient who had received an organ transplant was found to contribute to the remission of malignant melanoma in the patient [17].

More recently, a cohort of HIV-infected men between 60 and 70 years old was found to have a 21-fold increased risk for developing prostate cancer [18], suggesting that the immunity to prostate cancer in about 70% of men predicted by the modeling here could be the result of effective immune system reaction against mutated prostate cells in these men.

Prostatitis and subclinical inflammation of the prostate indicated by elevated prostate-specific antigen (PSA) levels were found to be common among HIV-infected men [18]. Thus, chronic inflammation of the prostate may be the cause of the ordered mutations that eventually lead to cancer [19–21].

The possibility that a normal, unsuppressed, functioning immune system can help eliminate malignant melanoma was confirmed by a case report by Hodi, Granter and Antin [17]. Hodi *et al.* report on a patient with aplastic anaemia who was treated with immunosuppressive drugs for 6 years until the patient came down with malignant melanoma that had metastasized. Immuno-suppression was discontinued and a subsequent biopsy of a tumour of the patient revealed melanoma cells infiltrated with CD4+ and CD8+ T cells. The patient then received one cycle of chemotherapy and a complete remission followed. The authors conclude that the unsuppressed immune system contributed to the remission of the melanoma.

Experiments with mice demonstrate that immune systems exist that block injected virulent cancer cells from imparting cancer [22].

Cells of some cancers survive the immune system's counterattack in part by secreting Nagalase (α-N-acetylgalactosaminidase), a substance that is highly immunosuppressive of macrophage activation. In a study of 16 prostate cancer and 16 breast cancer patients, Yamamoto and his collaborators found that the Nagalase activity in these cancer patients varied between 3 and 10 times greater than in the study's healthy controls [23].

Moreover, Yamamoto and Urade found that administering vitamin D-binding protein (Gc protein)-derived macrophage activating factor (GcMAF) at the rate of 100 ng once a week over a period of no more than 6 months can overcome the effect of Nagalase and reduce Nagalase activity in breast and prostate cancer cohorts to within normal levels [23]. Since the number of cancerous cells in a patient is proportional to the patient's Nagalase activity, reducing the Nagalase activity to normal levels implies that all cancerous cells have been eliminated from the patient by an activated immune system, a conclusion that was confirmed by

computed axial tomography (CAT-scans) [23]. Indeed, it is possible that GcMAF therapy may be effective against *all* cancers.

Yamamoto and his collaborators have also shown that both the HIV [24] and influenza [25] virions contain Nagalase and that patients infected with these virions exhibit elevated Nagalase activity. Yamamoto also found that GcMAF therapy (100 ng once a week) was effective in reducing the elevated Nagalase activity in three HIV-infected patients with acquired immunodeficiency syndrome (AIDS) to the level of healthy controls [25]; the eradication of HIV and HIV-infected cells by GcMAF therapy was confirmed by the complete clearance of viral antigens (p24 and gp120) in the patients' blood.

The high prevalence of natural immunity in breast cancer and prostate cancer risk populations predicted by the saturated, ordered, mutation model in this paper fits in with the current concepts of cancer immunosurveillance and immunoediting [27–33].

From direct experiments with mice and humans, natural immunity or effective cancer immunosurveillance has been demonstrated [27]. CD4+ and CD8+ $\alpha\beta$ T cells have been shown to recognize tumour antigens and are two key cells that the immune system uses to *eliminate* cancer cells. It is now thought that immunosurveillance is only first phase of a broader picture of the possible results of the interaction between cancer cells and the immune system. This broader picture has been dubbed immunoediting and consists of an initial immunosurveillance phase (during which cancer cells are eliminated) followed by a possible equilibrium phase (during which the number of cancer cells *in vivo* are not zero and remain constant) followed by a possible escape phase (during which renewed tumour growth occurs) [27–33].

The availability of the human genome sequence has led to the discovery of cancer genes. Most recently, Christopher *et al*. have identified 158 cancer-promoting mutations in 119 genes [34] to be added to the approximately 350 cancer genes previously discovered. Thus, the total number of cancer genes discovered to date fits in with the number of mutations required to cause the cancers appearing in table 1 (see column two) calculated from the modeling in this paper.

Regardless of the number of mutations m necessary to fit the incidence data of a particular cancer in a specific risk population, it is possible that the cancer can develop through *different pathways* in that risk population.

The only conclusive way to determine the precise mutations necessary to produce a particular cancer is to complete population-based genetic association studies involving cancer patients and non-cancer controls for the specific risk populations (males, females, etc.).

It is clear from this analysis that at any given age the precancerous cells of a person may have one or more mutated cells that are at least one mutation short of becoming cancerous. If a method could be found to destroy these cells *before* they become cancerous, then cancer could be *prevented*. It should be easier to destroy a handful of mutated precancerous cells then it would be to destroy a huge number of metastasized cancer cells that are constantly dividing. Viewed through this prism, preventing a cancer from occurring ought to be easier to accomplish than trying to cure it.

Clearly, if mutated pre-cancerous cells can be identified and destroyed, the associated cancer would be prevented from occurring in the first place. A drug that stimulates macrophage to recognize and destroy prostate and breast cancer cells has been discussed in the literature [23] and studies to see if this or other drugs can identify and destroy mutated, pre-cancerous cells should become a research imperative.

The idea that cancer requires a number of stages to develop was first proposed in a model by Armitage and Doll in 1954 who used the function pt^m to fit cancer incidence data [35]. This model has been mathematically modified over the years in an effort to take into account

the fact that the cumulative number of cancers in any risk population must eventually saturate. For example, Pompei and Wilson arbitrarily modified the Armitage and Doll model by using the function $pt^m(1 - kt)$ to model the incidence data [36]; this change was designed to help fit the apparent plateauing in the data with age. It should be noted that for small values of the product kt, $\exp(-kt) \approx 1 - kt$ so that the model developed here reduces to that of Pompei and Wilson. The physical model developed in this appendix is inherently saturated, with no *ad hoc* mathematical functions used to describe the cancer incidence function.

However, Armitage and Doll showed that a two-stage model for cancer development with clonal expansion of intermediate cell populations [37] could generate similar age-specific incidence curves to those in reference [35]. Using this approach, a multi-stage clonal expansion model of cancer development was constructed by Moolgavkar, Venzon and Knudson (also called as the MVK model in the literature) [38–40]. One problem with this model is that not all of the parameters it must introduce can be determined from incidence data alone [41]. A good example of how this model is applied in practice is contained in a recent paper by Luebeck and Moolgavkar [42]. The appendix for reference [42] contains all of the mathematical details necessary to make this model work; clearly, to fit this model's incidence function to cancer data requires knowledge of the values of many more parameters than does the ordered mutation model introduced here. In fact, the simplest form of the ordered mutation model involves only three parameters and all of these parameters are determined from a least-squares fit of the very simple model incidence function in (5c) to cancer incidence data.

Acknowledgement

References

[1] Davies, H. *et al.*, 2002, Mutations of the BRAF gene in human cancer, *Nature*, **417**(6892), 949–954.

[2] Brose Marcia, S., Volpe Patricia, Feldman Michael, Kumar Madhu, Rishi Irum, *et al.*, 2002, BRAF and RAS mutations in human lung cancer and melanoma, *Cancer Research*, **62**(23), 6997–7000.

[3] Novik, K.L., Spinelli, J.J., Macarthur, A.C., Shumansky, K., Sipahimalani, P., Lai, A., Conners, J.M., Gascovne, R.D., Gallagher, R.P. and Brooks-Wilson, A.B., 2007, Genetic variation in H2AFX contributes to risk of non-Hodgkin lymphoma, *Cancer Epidemiology, Biomarkers & Prevention*, **16**(6), 1098–1106.

[4] Li, H., Gu, Y., Hukku, B., McLeod, D.G., Hei, T.K. and Rhim, J.S., 2007, Malignant transformation of human benign prostate epithelial cells by high linear energy transfer alpha-particles, *International Journal of Oncology*, **31**(3), 537–544.

[5] Lichtenstein, P., Holm, N.V., Verkasalo, P.K., Iliadou, A., Kaprio, J., Koskenvuo, M., Pukkala, E., Skytthe, A. and Hemminiki, K., 2000, Environmental and heritable factors in the causation of cancer—analyses of cohorts of twins from Sweden, Denmark and Finland, *New England Journal of Medicine*, Jul 13, **343**(2), 78–85.

[6] Nam, R.K., Zhang, W.W., Loblaw, D.A., Klotz, L.H., Trachtenberg, J., Jewett, M.A., Stanimirovic, A., Davies, T.O., Toi, A., Venkateswaran, V., Sugar, L., Siminovitch, K.A. and Naroid, S.A., 2007, A genome-wide association screen identifies regions on chromosomes 1q25 and 7p21 as risk loci for sporatic prostate cancer, *Prostate Cancer and Prostatic Disease*, Sep 18, PMID:17876339 to be published.

[7] Vecchione, A., Gottardo, F., Gomella, L.G., Wildemore, B., Fassan, M., Bragantini, E., Pagano, F. and Baffa, R., 2007, Molecular genetics of prostate cancer: clinical translational opportunities, *Journal of Experimental & Clinical Cancer Research*, **26**(1), 25–37.

[8] Lane, T.M., Strefford, J.C., Yanez-Munoz, R.J., Purkis, P., Forsythe, E., Nia, T., Hines, J., Lu, Y.J. and Oliver, R.T., 2007, Identification of a recurrent t(4;6) chromosome translocation in prostate cancer, *Journal of Urology*, **177**(5), 1907–1912.

[9] Saramaki, O. and Visakorpi, T., 2007, Chromal aberrations in prostate cancer, *Frontiers in Bioscience*, **12**, 3287–3301.

[10] Berguin, I.M., Min, Y., Wu, B., Wu, J., Perry, D., Cline, J.M., Thomas, M.J., Thornberg, T., Kulik, G., Smith, A., Edwards, I.J., D'Agnostino, R., Zhang, H., Wu, H., Kang, J.X. and Chewn, Y.Q., 2007, Modulation of prostate cancer genetic risk by omega-3 and omega-6 fatty acids, *The Journal of Clinical Investigation*, **117**(7), 1866–1875.

[11] Monfardini, S., Vaccher, E., Pizzocaro, G. *et al.*, 1989, Unusual malignant tumors in 49 patients with HIV infection, *AIDS*, **3**, 449–452.

[12] Remick, S.C., 1996, Non-AIDS-defining cancers, *Hematology/Oncology Clinics of North America*, **10**, 1203–1213.

[13] Smith, C., Lilly, S., Mann, K.P. *et al.*, 1998, AIDS-related malignancies, *Annals of Medicine*, **30**, 323–344.

[14] Cooley, T.P., 2003, Non-AIDS-defining cancer in HIV-infected people, *Hematology/Oncology Clinics of North America*, **17**, 889–899.

[15] Mbulaiteye, S.M., Biggar, R.J., Goedert, J.J. *et al.*, 2003, Immune deficiency and risk for malignancy among persons with AIDS, *Journal of Acquired Immune Deficiency Syndromes*, **32**, 527–533.

[16] Frish, M., Biggar, R.J., Engels, E.A. *et al.*, 2001, AIDS-Cancer Match Registry Study Group Association of cancer with AIDS-related immunosuppression in adults, *JAMA*, **285**, 1736–1745.

[17] F. S. Hodi, Scott Granter and Joseph Antin, Withdrawal of immunosuppression contributing to the remission of malignant melanoma: a case report, *Cancer Immunity*, **5**, 7 (19 May 2005).

[18] Nancy F. Crum, 2005, Increased risk of prostate cancer among HIV-infected men, *Contagion*, **2**(2), 66–70.

[19] Nelson, W.G., De Maizo, A.M. and Issacs, W.B., 2003, Mechanisms of disease: prostate cancer, *New England Journal of Medicine*, **349**, 366–381.

[20] Platz, E.A. and De Maizo, A.M., 2004, Epidemiology of inflammation and prostate cancer, *Journal of Urology*, **171**, S36–S40.

[21] De Maizo, A.M., Marchi, V.L., Epstein, J.L. *et al.*, 1999, Proliferative inflammatory atrophy of the prostate: implications for prostatic carcinogenesis, *American Journal of Pathology*, **155**, 1985–1992.

[22] Jessica Ruvinsky, 2006, Are You Immune to Cancer?, *Discover*, **27**, 8.

[23] Yamamoto, N. and Ueda, M., 2004, Therapeutic Efficacy of Vitamin D-binding Protein (Gc Protein)-derived Macrophage Activating Factor (GcMAF) for Prostate and Breast Cancers, *12th International Congress of immunology and 4th Annual Conference of FOCIS*, Montreal, Canada, July 18–23.

[24] Yamamoto, N., 2006, Pathogenic significance of α-N-acetylgalactosaminidase activity found in the envelope glycoprotein gp160 of human immunodeficiency virus type 1, *AIDS Research and Human Retroviruses*, **22**(3), 262–271.

[25] Yamamoto, N. and Ueda, M., 2005, Pathogenic significance of α-N-acetylgalactosaminidase activity found in the hemagglutinin of influenza virus, *Microbes and Infection*, **7**, 674–681.

[26] Yamamoto, N. and Ueda, M., 2004, Eradication of HIV by Treatment of HIV-infected/AIDS Patients with Vitamin D-binding Protein Derivative, *12th International Congress of immunology and 4th Annual Conference of FOCIS*, Montreal, Canada, July 18–23, 2004.

[27] Gavin P. Dunn, Lloyd J. Old and Robert D. Schreiber, 2004, The immunobiology of cancer immunosurveillance and immunoediting, *Immunity*, **21**, 137–148, August.

[28] Gavin P. Dunn, Allen T. Bruce, Hiroaki Ikeda, Lloyd J. Old and Robert D. Schreiber, 2002, Cancer immunoediting: from immunosurveillance to tumor escape, *Nature Immunology*, **3**, 991–998.

[29] Dunn Gavin P., Koebel Catherine M. and Schreiber, Robert D., 2006, Interferons, immunity and cancer immunoediting, *Nature Reviews. Immunology*, **6**(11), 836–848.

[30] Zhihai Qin and Thomas Blankenstein, 2004, A cancer immunosurveillance controversy, *Nature Immunology*, **5**, 4–5 (1 January 2004).

[31] Lawrence Zitvogel, Antoine Tesniere and Guido Kroemer, 2006, Cancer despite immunosurveillance: immunoselection and immunosubversion, *Nature Reviews. Immunology*, **6**(10), 715–727. Epub 2006 Sep 15.

[32] Germenis, A.E. and Karanikas, V., 2007, Immunoepigenetics: the unseen side of cancer immunoediting, *Immunology and Cell Biology*, **85**(1), 55–59. Epub 2006 Nov 28.

[33] Inman Brant A, Frigola Xavier, Dong Haidong and Kwon Eugene D., 2007, Costimulation, coinhibition and cancer, *Current Cancer Drug Targets*, **7**(1), 15–30.

[34] Christopher Greenman *et al.*, Patterns of somatic mutation in human cancer genomes, *Nature*, **446**, 153–158 (8 March 2007).

[35] Armitage, P. and Doll, R., 1954, The age distribution of cancer and a multi-stage theory of carcinogenesis, *British Journal of Cancer*, **8**(1), 1–12.

[36] Pompei, F. and Wilson, R., 2001, The age distribution of cancer; the turnover at old age, *Health and Environmental Risk Assessment*, **7**(6), 1619–1650.

[37] Armitage, P. and Doll, R., 1957, *British Journal of Cancer*, **11**, 161–169.

[38] Moolgavkar, S. and Knudson, A., 1981, Mutation and cancer: a model for human carcinogenesis, *Journal of the National Cancer Institute*, **66**(6), 1037–1052.

[39] Moolgavkar, S.H., Dewanji, A. and Venzon, D.J., 1988, A stochastic two-stage model for cancer risk assessment. I. The hazard function and the probability of tumor, *Risk Analysis*, **8**(3), 383–392.

[40] Moolgavkar, S.H. and Luebeck, E.G., 1990, Two-event model for carcinogenesis: biological, mathematical and statistical considerations, *Risk Analysis*, **10**(2), 323–341.

[41] Heidenreich, W.F., Luebeck, E.G. and Moolgavkar, S.H., 1997, Some properties of the hazard function of the two-mutation clonal expansion model, *Risk Analysis*, **17**(3), 391–399.
[42] George Lueback, E. and Moolgavkar, Suresh H., 2002, Multistage carcinogenesis and the incidence of colorectal cancer, *PNAS*, **99**(23), 15095–15100.

Appendix

Cancer of a given type requires a chain of *m ordered* mutations to occur within a cell of a given organ, where $m = 1, 2, 3, \ldots$ The model to be developed here will assume that these mutations occur randomly, are independent of time, but are ordered so that, say, mutation two cannot occur unless mutation one has occurred first, etc.

Consider a random cohort of infants (age zero) whose members are all susceptible to developing cancer of a particular type as they age. The number of susceptible infants in the cohort will be denoted by N_s. Assuming that birth is coincident with time $t = 0$, the number of members of the cohort that have experienced the first mutation but none higher are denoted by $N_1(t)$, the number of people who have experienced the first two-ordered mutations but none higher are denoted by $N_2(t)$, etc.

If $p = 0, 1, 2, 3, \ldots, m - 1$, the fraction of the number of people in the cohort with p mutations who experience the next mutation per unit time will be denoted by k_{p+1}, so that m random, ordered mutations are required to cause cancer of a certain type. In this paper the quantities $k_q, q = 1, 2, 3, \ldots, m$, will be called *mutation rate*s. Thus, at any given time t after birth, the number of people in the cohort with any given number of mutations is depicted in diagram 1. Since the last mutation in the *ordered* chain of mutations causes cancer, the number of people in the cohort that have developed cancer at time t is then $N_m(t)$. The numbers $N_0(t), N_1(t), \ldots, N_m(t)$ will be called the sub-populations of the cohort at time t.

It is possible that a fraction of a random sample of the population is actually *immune* to developing a particular cancer. If this immune fraction is denoted by f_i, then the fraction of the population that is susceptible to developing a particular cancer is $f_s = 1 - f_i$, where $0 \le f_s \le 1$. Thus, in a random cohort of N_0 people, the number of people susceptible to a particular cancer is given by $N_s = f_s N_0$. For example, it is a common practice in reporting cancer incidence data to take $N_0 = 100,000$.

The set of coupled, time-dependent equations involving the sub-populations in this model are

$$\frac{\mathrm{d}N_0(t)}{\mathrm{d}t} = -k_1 N_0(t) \tag{1a}$$

$$\frac{\mathrm{d}N_1(t)}{\mathrm{d}t} = k_1 N_0(t) - k_2 N_1(t) \tag{1b}$$

$$\frac{\mathrm{d}N_p(t)}{\mathrm{d}t} = k_q N_{q-1}(t) - k_{q+1} N_q(t), q = 1, 2, 3, \ldots, m-1, \tag{1c}$$

$$\frac{\mathrm{d}N_m(t)}{\mathrm{d}t} = k_m N_{m-1}(t). \tag{1d}$$

Adding these equations together and integrating leads to

$$N_s = N_0(t) + N_1(t) + N_2(t) + N_3(t) + \ldots + N_m(t) \equiv \sum_{p=0}^{m} N_p(t), \tag{2}$$

since $N_r(0) = 0$ for $r = 1, 2, \ldots, m$.

The solutions to each of the equations in (1) *in order* are

$$N_0(t) = N_s e^{-k_1 t} = f_s N_0 e^{-k_1 t}, \tag{3a}$$

$$N_q(t) = e^{-k_{q+1} t} \int_0^t k_q N_{q-1}(t') e^{k_{q+1} t'} \, dt', \quad q = 1, 2, \ldots, m-1, \tag{3b}$$

$$N_m(t) = k_m \int_0^t N_{m-1}(t') dt'. \tag{3c}$$

If $(k_r t) \ll 1$ for $r = 1, 2, \ldots, m$, then the equations in (3) yield the approximate solution

$$N_m(t) \approx f_s N_0 \cdot k_1 k_2 k_3 \ldots k_m \frac{t^m}{m!}, (kt) \ll 1. \tag{4}$$

Thus the value of m, the number of mutations necessary to cause a particular cancer, can be determined by fitting the function in (4) the *early part* of the $N_m(t)$ data curve.

If all the mutation rates k_r, $r = 1, 2, \ldots, m$, are equal to the same constant k, then the solutions in (3b) yield the particularly simple solutions

$$N_q(t) = f_s N_0 \frac{k^q}{q!} t^q e^{-kt}, \quad q = 1, 2, \ldots, m-1, \tag{5a}$$

and the number of people in the cohort that have developed cancer by time t is given by

$$N_m(t) = f_s N_0 \left[1 - e^{-kt} \cdot \sum_{p=0}^{m-1} \frac{(kt)^p}{p!} \right]. \tag{5b}$$

Notice that $N_m(0) = 0$, as it should and $N_m(\infty) = N_s = f_s N_0$ as it must. Thus, if people lived forever, the entire susceptible population would develop cancer.

Using either (3c) and (5a) or (5b), the number of people per unit time in the cohort who come down with cancer at age t (the *cancer incidence rate*) is given by

$$\text{IR}(m; k; t) \equiv \frac{dN_m(t)}{dt} = \frac{f_s N_0 k^m}{(m-1)!} t^{(m-1)} e^{-kt}, \tag{5c}$$

where m is the number of mutations necessary to cause a particular cancer, k is the mutation rate and f_s is the fraction of the population that is susceptible to developing this particular cancer. The values of m, k and f_s are determined by fitting the incident rate function in (5b) to cancer incidence data. Thus, the fit involves determining three independent parameters in

general. If every member of the population is susceptible to developing a particular cancer, then $f_s = 1$ and the fit involves determining the values of only two parameters, m and k.

An inherent feature of the model developed here is that the cumulative number of people in a risk population that develop cancer over time can never exceed the total number of people in the risk population, a characteristic known as *saturation*. The saturation of the model is responsible for the fact that the general cancer incidence function in (5c) monotonically increases, peaks and monotonically declines towards zero as the age of the risk population increases. The peak in the incidence function in (5c) occurs at the age of

$$t_{\max} = \frac{(m-1)}{k} \quad \text{with the value of } \mathrm{IR}(m; k; t_{\max}) = \frac{f_s N_0 k}{(m-2)!}(m-1)^{(m-2)}\mathrm{e}^{-(m-1)}. \quad (5d)$$

The following procedure was followed in obtaining the least-squares fits shown in this paper. Referring to (5c), the parameter

$$p \equiv \frac{f_s N_0 k^m}{(m-1)!} \quad (5e)$$

was regarded as an independent parameter to be determined, along with m and k, by the fit. Then, solving (5e) for f_s, the value of f_s can be calculated from the values of p, m and k returned by the fit. If the set of n consecutive data points used in the fit are denoted by $\{d_i\}$ and if the corresponding model fit for this points are denoted by the set $\{x_i\}$, then the square of the error of the fit, to be called χ^2, is defined as

$$\chi^2 \equiv \sum_{i=1}^{n}(x_i - d_i)^2. \quad (5f)$$

The lower the value of χ^2, the better the model fit is to the data.

If $kt \ll 1$, then the leading term in the solution in (5b) is

$$N_m(t) \approx f_s N_0 \frac{(kt)^m}{m!}, \quad (kt) \ll 1. \quad (5g)$$

Equations (5g) and (4) leads to the following relationship between the average mutation rate k and the set m mutation rates $k_1, k_2, \ldots, k_m$:

$$k^m = k_1 k_2 k_3 \ldots k_m. \quad (6)$$

The *physical* model that produces the result in (5c) for the cancer incidence rate has distinct advantages over the *phenomenological* models whose independent parameters generally lack a simple, clear physical interpretation.

If all the mutation rates have different values, then the time-dependent cancer incidence rate can be calculated in a straightforward way using the results in equations (3).

For example, if $m = 1$, then using (3a) and remembering that for $r = 1, 2, \ldots, m$

$$\mathrm{IR}(1; k_r; t) \equiv \frac{dN_1(t)}{dt} = k_1 N_0(t) = f_s N_0 k_1 \mathrm{e}^{-k_1 t} \quad (7a)$$

and

$$N_1(t) = f_s N_0 [1 - e^{-k_1 t}] \tag{7b}$$

Similarly, if $m = 2$, then

$$N_1(t) = f_s N_0 \frac{k_1}{(k_2 - k_1)} \left[e^{-k_1 t} - e^{-k_2 t} \right] \tag{8a}$$

and

$$IR_2(2; k_r; t) \equiv \frac{dN_2(t)}{dt} = k_2 N_1(t) = f_s N_0 \frac{k_1 k_2}{(k_2 - k_1)} \left[e^{-k_1 t} - e^{-k_2 t} \right] \tag{8b}$$

As a final example, if $m = 3$, then

$$\text{IR}(3; k_r; t) \equiv \frac{dN_3(t)}{dt} = k_3 N_2(t) = f_s N_0 \frac{k_1 k_2 k_3}{(k_2 - k_1)} \left[\frac{\left(e^{-k_1 t} - e^{-k_3 t} \right)}{(k_3 - k_1)} - \frac{\left(e^{-k_2 t} - e^{-k_3 t} \right)}{(k_3 - k_2)} \right]. \tag{9}$$

Continuing in this way, the cancer incidence rate $\text{IR}_m(t)$ for any value of m and arbitrary (positive) values of the mutation rates $k_1, k_2, k_3, \ldots, k_m$, can be computed.

Notice that the cancer incidence functions in (7b), (8b) and (9) all approach zero if the age of the cohort becomes large enough, a characteristic feature of *all* cancer incidence functions in this model. For $m > 1$, the incidence function must also vanish at age $t = 0$. Thus, for $m > 1$, it is always the case that the incidence function starts out at zero, monotonically grows until it reaches a peak and then monotonically declines towards zero as the age of the risk population continues increasing. However, the peak in the incidence rate function may occur at ages above the natural human life-span; in these cases not everyone in the risk population that is susceptible to developing cancer will get it before they die of something else. This situation occurs for melanoma, pancreatic cancer and NHL as discussed in this paper. For female breast cancer the peak in the cancer incidence curve occurs around 77 years of age.

For greater clarity, at points in this exposition the function $N_s(t)$ will be denoted by $N(s/m, t)$, $s = 0.1, 2, \ldots, m$, where m is the number of mutations necessary to cause a particular cancer in a characteristic cohort.