Widespread Immunity to Breast and Prostate Cancers is Predicted by a Novel Model that also Determines Sporadic and Hereditary Susceptible Population Sizes

I. Kramer¹ *

¹ Physics Department, University of Maryland Baltimore County

Abstract. Natural immunity to breast and prostate cancers is predicted by a novel, saturated ordered mutation model fitted to USA (SEER) incidence data, a prediction consistent with the latest ideas in immunosurveillance. For example, the prevalence of natural immunity to breast cancer in the white female risk population is predicted to be 76.5%; this immunity may be genetic and, therefore, inherited. The modeling also predicts that 6.9% of White Females are born with a mutation necessary to cause breast cancer (the *hereditary* form) and, therefore, are at the highest risk of developing it. By contrast, 16.6% of White Females are born without any such mutation but are nonetheless susceptible to developing breast cancer (the *sporadic* form). The modeling determines the required number of ordered mutations for a cell to become cancerous as well as the mean time between consecutive mutations for both the sporadic and hereditary forms of the disease. The mean time between consecutive breast cancer mutations was found to vary between 2.59 - 2.97 years, suggesting that such mutations are rare events and establishing an upper bound on the lifetime of a breast cell. The prevalence of immunity to breast cancer is predicted to be 79.7% in Blacks, 86.5% in Asians, and 85.8% in Indians. Similarly, the prevalence of immunity to prostate cancer is predicted to be 67.4% for Whites, 50.5% for Blacks, 77.7% for Asians, and 78.6% for Indians. It is of paramount importance to delineate the mechanism underlying immunity to these cancers.

Key words: cancer incidence, mutation model, immunity, sporadic, hereditary, pathways, proto-oncogene, immunosurveillance, cell lifetime.

AMS subject classification: 92C60

*E-mail: kramer@umbc.edu

1. The Ordered Mutation Model of Cancer Development

Research data suggests that a cancerous cell results from an *ordered* series of *random* mutations of a special subset of its genes (*cancer genes*). 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. A mutated gene that is necessary to cause a specific cancer is called an *oncogene*, while the un-mutated form of this gene is called a *proto-oncogene*. Thus, a cell becomes cancerous after sequential mutations occur in an ordered set of its proto-oncogenes. If this set is not unique, then the cancer can develop through different *pathways*.

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 which 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 here will enable all three of these quantities to be calculated by fitting the models cancer incidence function to cancer incidence data.

An important, novel feature of the cancer model to be presented here is that it is inherently saturated, i.e., the maximum percentage of a risk population (*e.g.* White Males) that can develop a particular cancer can never exceed 100%. Thus, the models simplest 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).

Discovering the causes of the mutations that lead to the development of cancers is one of the outstanding challenges to date. Clearly, if the etiological cause of any one of the mutations necessary to cause a particular cancer could be identified and its action blocked, then the cancer could never develop in the first place. The ordered, saturated, mutation model to be developed here will be fitted to the incidence data for breast and prostate cancers. One of the biggest surprises that results from the modeling is the prediction that significant natural immunity does indeed exist for *both* of these cancers.

The model used here is a more sophisticated version of the one developed for describing cancers in reference [1] and is completely described in the appendix here for the convenience of the readers. This mutation model constructed here was also used to describe the transition from HIV infection to AIDS [2] and the development of Alzheimers disease [3].

One of the most important features of this 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 $f_s + f_i = 1$. Thus, in a random cohort of N_0 people in a risk population, the number N_s that are susceptible to a particular cancer is given by $N_s = f_s N_0$, where the value of f_s is determined by a least-squares

fit to cancer incidence data. The number of people in this cohort that have acquired r cancerous mutations at age t will be denoted by $N_r(t)$, and a complete description of how the functions $N_r(t)$ are computed from the model appears in the appendix.

If a number m ordered mutations are required to cause a particular cancer, 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 r^{th} mutation state at age t , where $r = 1, 2, 3, \dots, m$. 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.

The age-related annual cancer incidence rate data will be used as a substitute for the ideal experiment discussed above in order to acquire the data needed to determine the number of mutations m needed to generate a particular cancer, the mutation rates, and the fraction f_i of the population that is naturally immune to developing cancer of a certain type. This alternative approach 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 solution of the model equations 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 (9.12)]. This simplified model will be applied to breast and prostate cancer data in this paper 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 (NCI) website.

2. Female Breast Cancer in the U.S.A.

The USA breast cancer incidence rate data for white females is shown in Figure 1A/WF. All of the data in this paper is the SEER 1992-2006 *annual* incidence data per 100,000 members of a risk population. In the labeling of figures, the notation WF stands for the White Female risk population. Figures were named in this way to make them easier to identify in this paper. To fit the data in Figure 1A/WF, the simplest model incidence function described in the appendix was tried first, namely, that all the mutation rates are identical, equal to the common value k . Thus, using the trial incidence function in (9.12) in the appendix, the fit to the WF data shown in Figure 1A/WF was obtained with only three independent parameters: f_s , k , and m . As seen in this figure, the (χ^2) error of this fit is $739(\text{year})^{-2}$ [see (9.14) in the appendix], but the fit is nonetheless credible. Fixing the value of susceptibility parameter to $f_s = 1$ and refitting this data with only two parameters leads to a clearly not credible fit with an error of $14,274(\text{year})^{-2}$, over 19 times what it was in the 3-parameter fit. Thus, the modeling suggests that the value of f_s is substantially less than unity and that immunity to breast cancer in the WF population exists. To improve the 3-parameter fit in the range below 50 years of age, the incidence rate function was generalized

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

Diagram (1). Mutation model leading to cancer

to include a subpopulation of susceptible females who were born with some of the mutations necessary to cause breast cancer. Thus, it was assumed that there were two subpopulations of susceptible white females. A fraction f_{s1} of white females is assumed to be born with the *sporadic* form of susceptibility to development of breast cancer, i.e., this fraction was born with *none* of the mutations required to develop the disease; the incidence function for this subpopulation is described by (9.12) with the parameters f_{s1} , k , and m_1 . On the other hand, a fraction f_{s2} of the white female population is not only susceptible to developing breast cancer but was born with some of these mutations necessary to cause it; this subpopulation is said to have the *hereditary* form of the susceptibility to breast cancer development and has an incidence function described by (9.12) with parameters f_{s2} , k , and $m_2 < m_1$. Females born with the hereditary susceptibility have a head-start in the development of breast cancer and, therefore, are more likely to develop it. Notice that the mutation rate parameter k is assumed to be *identical for both susceptible subpopulations*. Thus, the compound incidence rate function used for the WF breast cancer fit is given by

$$IR(t) = IR(f_{s1}, m_1, k, f_{s2}, m_2; t) = IR_1(f_{s1}, m_1, k; t) + IR_2(f_{s2}, m_2, k; t) \quad (2.1)$$

and the resulting fit is shown in Figure 1B/WF/CMPD along with the values of the 5 fit parameters. [The factor CMPD in this figure description stands for the *compound model* represented by equation (2.1).] Notice that the fit error has been reduced by a factor of 4.9 to only $150.7(\text{year})^{-2}$,

Fig. (1A/WF). SEER 1992-2006 White Female breast cancer incidence rate $IR(t)$ data as a function of age t together with ordered mutation model fit

only 20% of what it was in the 3-parameter fit. Clearly, this 5-parameter fit is excellent, $m_2 < m_1$ as it must, and $f_{s2} < f_{s1}$ as expected. According to this fit, about 16.6% of WF ($f_{s1} = 0.166$) require 30 ordered mutations ($m_1 = 30$) to develop breast cancer, while another 6.8% ($f_{s2} = 0.0686$) require only 21 ordered mutations ($m_2 = 21$) to develop it *on the average*. The value of the mutation rate k for both subpopulations is identical at $k = 0.360(\text{year})^{-1}$ so that the mean time between successive mutations is $T = k^{-1} = 2.77 \text{ years}$. Assuming that white females born with sporadic susceptibility (the m_1 subpopulation) are all born with none of the ordered mutations required to cause breast cancer, then all females born with hereditary susceptibility (the m_2 subpopulation) inherit the 9th ordered mutation at birth, on the average, and, therefore, have a head-start in developing breast cancer and are more likely to develop it. Since the value of the mean time between successive mutations for both subpopulations is identical, members of the m_2 subpopulation are more likely to develop breast cancer at earlier ages than members of the m_1 subpopulation.

The total fraction of the white female population that is susceptible to developing breast cancer is predicted to be $f_{s1} + f_{s2} = 0.234$ so that about 76.5% are immune to developing this cancer. The values of the fit parameters shown in Figure 1B/WF/CMPD are also shown in Table 1 along with the value of the fit error (see (9.14) in the appendix). Is immunity of breast cancer also predicted for the other USA female risk populations?

Fitting the basic model incidence function in equation (9.12) in the appendix to the SEER

Black Female breast cancer incidence data gives the results shown in Figure 2A/BF with a χ^2 error of $880(\text{year})^{-2}$. Applying the compound model to this data reduces this error by a factor of 2.6 and gives the much better fit shown in Figure 2B/BF/CMPD; the values of the 5 parameters that determine this fit are also shown in Table 1.

Again using the incidence function in equation (9.12) yields a fit to the SEER Asian Female breast cancer data shown in Figure 3A/AF with a χ^2 error of $944(\text{year})^{-2}$. Applying the compound model to this data reduces this error to $226(\text{year})^{-2}$, a reduction by a factor of 4.1, as shown in Figure 3B/AF/CMPD; again, the values of the 5 parameters that determine this fit are also shown in Table 1.

Finally, using the incidence function in equation (9.12) yields the fit to the SEER American Indian Female breast cancer data shown in Figure 4A/IF with a χ^2 error of $1319(\text{year})^{-2}$. Applying the compound model to this data reduces this error to $333(\text{year})^{-2}$, a reduction by a factor of 3.9, as shown in Figure 4B/IF/CMPD; the values of the 5 parameters that determine this fit are also shown in Table 1.

As seen in Table 1, the modeling predicts that 79.6% of Black Females, 86.5% of Asian Females, and 85.8% of American Indian Females in the USA are immune to developing breast cancer, percentages that are comparable to that obtained for the White Female population (76.5%). The mean time between consecutive cancerous mutations was found to vary between 2.59-2.97 years; these values constitute an upper limit on the lifetime of a breast cell since mutations occur during a cellular turn-over.

The modeling predicts that the number of ordered mutations required to develop *sporadic* breast cancer is 30 for Whites, 28 for Blacks, 30 for Asians, and 32 for Indians, results that are very close to each other but indicate that sporadic breast cancer can develop through slightly different pathways.

As already noted, the modeling also leads to the prediction that 6.86% of White Females are born with the 9th ordered mutation and require only 21 additional ordered mutations to develop *hereditary* breast cancer. Similarly, the modeling predicts that 6.09% of Black Females are born with the 9th ordered mutation and require only 19 additional ordered mutations to develop hereditary breast cancer, that 5.49% of Asian Females are born with the 9th ordered mutation and require only 21 additional ordered mutations to develop hereditary breast cancer (identical to the results for White Females), and that 5.24% of Indian Females are born with the 10th ordered mutation and require only 22 additional ordered mutations to develop hereditary breast cancer.

All of the above corresponding results are really very similar to each other indicating that the development of breast cancer is remarkable independent of race.

3. Prostate Cancer in the U.S.A.

The modeling that was used to fit the breast cancer incidence data above will now be applied to USA prostate cancer data.

The USA prostate cancer incidence rate data for white males is shown in Figure 5A/WM where

Fig. (1B/WF/CMPD). SEER 1992-2006 White Female breast cancer incidence rate $IR(t)$ data as a function of age t together with COMPOUND ordered mutation model fit

the notation “WM” stands for the White Male risk population. Using the trial incidence function in equation (A5c) in the appendix with three independent parameters, the fit to the WM data is also shown in Figure 5A/WM. As seen in this figure, the (χ^2) error of this fit is $947(\text{year})^{-2}$ [see (9.14) in the appendix], but the fit is nonetheless credible. Fixing the value of susceptibility parameter to $f_s = 1$ and refitting this data with only two parameters leads to a clearly not credible fit with an error of $187,320(\text{year})^{-2}$, over 197 times what it was in the 3-parameter fit! Thus, the modeling suggests that the value of f_s is substantially less than unity and that immunity to prostate cancer in the WM population exists.

Applying the compound model in equation (2.1) to the White Male prostate data yields the fit shown in Figure 5B/WM/CMPD with a fit error reduced by 11.3% to $840(\text{year})^{-2}$. The values of the parameters used in this fit are shown in Table 2. Although the compound model leads to a better fit to the WM data, it is really impossible to tell the difference between the fits in Figures (5A/WM) and Figure 5B/WM with the naked eye because the size of the hereditary prostate cancer WM population turns out to be relatively small. Since this result holds true when the above analysis is extended to the Black, Asian, and (American) Indian male prostate incidence data, there is no point in exhibiting the fits to these data using the incidence function in Equation (9.12).

Accordingly, the compound model fits to Black, Asian, and (American) Indian male prostate incidence data are shown in Figure 6B/BM/CMPD, Figure 7B/AM/CMPD, and Figure 8B/IM/CMPD,

Fig. (2A/BF). SEER 1992-2006 Black Female breast cancer incidence rate $IR(t)$ data as a function of age t together with ordered mutation model fit

respectively, and the values of the parameters for these three fits are also shown in Table 2.

From Table 2, natural immunity to prostate cancer is predicted by the modeling in 67.4% of White Males, 50.5% of Black Males, 77.7% of Asian Males, and 78.6% of (American) Indian Males. Thus, there is a greater racial disparity in immune percentage for prostate cancer than there is for breast cancer. From these results Black Males have an unusually high susceptibility to the development of prostate cancer.

The mean lifetime between consecutive ordered mutations for prostate cancer predicted by the modeling varies between 1.56-2.49 years, again indicating that these mutations are rare events; as with breast cancer, these periods constitute an upper bound on the lifetime of a prostate cell since all cancerous mutations occur during a cell turn-over.

The modeling predicts that the number of ordered mutations required to develop *sporadic* prostate cancer is 43 for Whites, 39 for Blacks, 51 for Asians, and 33 for Indians, indicating that sporadic prostate cancer can develop though significantly different pathways.

The modeling also leads to the prediction that only 0.40% of White Males are born with the 10th ordered mutation and require only 33 additional ordered mutations to develop *hereditary* prostate cancer. Similarly, the modeling predicts that only 1.65% of Black Males are born with the 6th ordered mutation and require only 33 additional ordered mutations to develop hereditary prostate cancer, that only 0.43% of Asian Males are born with the 12th ordered mutation and require only 39 additional ordered mutations to develop hereditary prostate cancer, and that 11.4% of Indian Males

are born with the first ordered mutation and require 32 additional ordered mutations to develop hereditary prostate cancer. Thus, with the exception of American Indian Males, a relatively small percentage of USA males are born with a susceptibility to the development of hereditary prostate cancer.

4. The Prevalence of Breast and Prostate Cancer with Aging

The fraction of a random sample of a risk population at age t that has developed a specific cancer, also known as the *prevalence* $P(t)$ of the cancer, is given by (9.23) in the appendix. For example, plugging the White Female breast cancer compound incidence function shown in Figure 1B/WF/CMPD into the right-hand side of equation (9.23) and integrating gives the prevalence function shown in Figure 9. This prevalence curve clearly saturates at the value of 0.2349 in total agreement with the results in Table 1 which predict that a fraction equal to $1 - 0.2349 = 0.7651$ of White Females is immune to developing breast cancer. The result in Figure 9 confirms that the White Female modeling results in Table 1 are correct. From Figure 9, by the age of 90 years old, only 18.2% of White Females will have developed breast cancer although 24.5% of them are susceptible to developing it. The SEER breast cancer incidence rates before the age of 20 years for all risk populations were found to be negligible, and the prevalence curve in Figure 9 confirms this. The modeling here predicts that the breast cancer prevalence for White Females is 0.4% at age 40 years, 1.96% at age 50 years, and 4.84% at age 60 years. Very similar results are obtained for the Black, Asian, and Indian Female breast cancer risk populations.

The prevalence $P(t)$ of prostate cancer in the male risk populations as a function of age t is obtained in exactly the same way. For example, plugging the White Male prostate cancer compound incidence function shown in Figure 5B/WM/CMPD into the right-hand side of equation (9.23) and integrating gives the prevalence curve shown in Figure 10. This prevalence curve clearly saturates at the value of 0.3251 in total agreement with the results in Table 2 which predict that a fraction equal to $1 - 0.3251 = 0.6749$ of White Males are immune to developing prostate cancer. The result in Figure 10 confirms that the White Male modeling results in Table 2 are indeed correct. From Figure 10, by the age of 80 years old, only 20.4% of White Males will have developed prostate cancer although 32.5% are susceptible to developing it. The SEER prostate cancer incidence rates before the age of 40 years for all risk populations were found to be negligible, and the prevalence curve in Figure 10 similarly confirms this. The modeling here predicts that the prostate cancer prevalence for White Males is 0.238% at age 60 years, 9.7% at age 70 years, and 20.4% at age 80 years. Similar results are obtained for the Black, Asian, and Indian Male prostate cancer risk populations. Clearly, the prevalence curve for prostate cancer in White Males is significantly different from the prevalence curve for breast cancer in White Females.

Attempts to improve the model fit results by increasing the number of distinct mutation rate constants have all failed—*increasing* the number of distinct (different valued) mutation rates in the fits *increases* the fit error! The fact that the best fit to the breast and prostate cancer incidence data for a risk population requires *only one* mutation rate constant k suggests that the mutations

leading to these cancers *are naturally and unavoidably generated at a constant rate* in those at risk, on the average. These modeling results imply that mutations (cancerous and non-cancerous) occur *at a constant rate* and that pre-cancerous cells are interacting with a constant mutagenic environment on the average. Moreover, the success of the modeling suggests that the generation of cancerous mutations *cannot be stopped*, at least in the normal environment! This implies that *external carcinogens have a minimal impact on the development of these cancers* that are caused by random, unavoidable, uncorrectable errors in the DNA copying process during cell turn-over.

The SEER incidence rate data curve for cervical cancer is clearly very different from those for breast and prostate cancers discussed in this paper. The incidence rate function for cervical cancer [caused by the Human Papillomavirus, HPV] must have more parameters to describe it than that for breast and prostate cancers because the incidence of HPV infection with age would have its own set of unique, independent parameters whose values must be put into the modelling by hand. **No analogous infection appears to be required in the development of breast and prostate cancers.**

Since cancerous mutations occur during the cell-turnover process, any substance that merely causes an increase in the cell turnover rate will be a risk factor for the disease. Such risk factors do not directly cause cancerous mutation themselves, they only increase the risk or probability that other factors will cause the mutation(s) over a given time period. It is important to note that eliminating or blocking these risk factors from the environment will **not** block cancer from developing, *it will merely reduce its incidence*. The *Helicobacter pylori* (*H. pylori*) bacterial infection may fall into this category along with infections that cause gastroenteritis.

Hormones that drive cell proliferation or turnover can increase the rate at which cancerous mutations accumulate in a cell and, therefore, increase the cancer incidence rate. Hormone-related cancers include breast, endometrium, ovary, prostate, testis, thyroid, and osteosarcoma. No specific initiators are required to generate these cancers which result from unavoidable (unrepaired) DNA copying errors that survive immune system surveillance. Interestingly, anti-hormone therapies have been effective in decreasing the cell turnover rate and, thereby, decreasing the cancer incidence rate at any age.

5. Genetic Models of Sporadic and Hereditary Cancer

The compound mutation model that leads to the incidence rate function in equation (2.1) is in complete agreement with Alfred Knudsons pioneering 1971 paper on retinoblastoma [4], a cancer caused when *both* copies of an allele of a specific gene (a *tumor suppressor* gene) in the retina are mutated. The *retinoblastoma* gene was discovered by Stephen Friend in 1986 [5], thereby confirming Knudsons model. There are two different groups of children born with a susceptibility to developing this disease. The first group has inherited the first cancerous mutation from a parent and, therefore, has a head-start in developing retinoblastoma (the *familial* type or *hereditary* retinoblastoma); of course, these children will develop the disease earlier, on the average, than those in the second susceptible group which are born with neither of the two mutations needed to

Fig. (2B/BF/CMPD). SEER 1992-2006 Black Female breast cancer incidence rate $IR(t)$ data as a function of age t together with COMPOUND ordered mutation model fit

cause it (the *sporadic* type). The Knudson model, verified by years of experiment, has exactly the form of the model presented in equation (2.1) that has been shown here to successfully describe the incidence of breast and prostate cancers.

Anthony T. Yeung, Alfred Knudson, et al., have recently shown that an inherited germline mutation in one allele of the adenomatous polyposis coli (APC) tumor suppressor gene causes the development of familial adenomatous polyposis (FAP, multiple polyps in the colon) that invariably leads to colon cancer [6]. The colon crypts of those with and without the inherited heterozygous APC gene were analyzed to determine the protein differences in the cells in the lining of the colon and rectum between these two groups. (Colon crypt cells renew the lining of the intestine and make mucus.) About 13% of the 1,695 identified proteins were abnormally expressed in the normal looking crypts of APC mutation carriers with FAP. These abnormally expressed proteins are proof that cells that have undergone cancerous mutations are identifiably different from their non-mutated counter-parts and that cancerous mutations can be inherited; these results again confirm the modeling in this paper which assumed that cancer incidence is made up of sporadic and heritable parts. Clearly, if a way could be found to selectively single out these mutated cells for destruction, the associated cancer could be prevented altogether.

The results in reference [6] implies that *susceptibility* to a cancer can be an inherited trait in that a set of proto-oncogenes determine whether a cell in a particular organ *in a particular environment of carcinogens* will undergo the ordered series of mutations necessary to become cancerous. The

Fig. (3A/AF). SEER 1992-2006 Asian Female breast cancer incidence rate $IR(t)$ data as a function of age t together with ordered mutation model fit

Knudson model requires *both* alleles of a proto-oncogene of a cell to become mutated in the same way in order for the *next* ordered mutation to become possible. Applying the Knudson model to the ordered mutation model developed in this paper leads to Table 3, which shows the status of the set of proto-oncogenes that lead to hereditary and sporadic cancer *at birth*.

The mutations of the proto-oncogenes occur in *ascending* order, first gene 1, then gene 2, etc., with allele 1 (either paternal or maternal) mutating first and allele 2 mutating second. An "O" in Table 3 indicates an un-mutated allele that is nonetheless susceptible to developing a cancerous mutation. An "X" in Table 3 indicates that a cancerous mutation has occurred in a given allele of a given proto-oncogene transforming it into an oncogene. Thus, as shown in Table 3, those that are susceptible to hereditary cancer are born with an inherited mutation in allele 1 in gene h and, therefore, have a head-start in developing cancer. An "I" in Table 3 indicates that a cancerous mutation is blocked in both alleles of any gene i , where i could be any positive number less than $m/2$. One possibility here is that those that are born with immunity to cancer are not susceptible to developing *any* of the set of ordered mutations necessary to cause it, in which case $i = 1$.

Fig. (3B/AF/CMPD). SEER 1992-2006 Asian Female breast cancer incidence rate $IR(t)$ data as a function of age t together with COMPOUND ordered mutation model fit

6. The Immune System's Role in Controlling Cancer Incidence

The fact that HIV-infected people experience increased incidences of Kaposi sarcoma, non-Hodgkins lymphoma, and invasive cervical cancer [7, 8, 9, 10, 11, 12] constitutes proof that the immune system can detect precancerous mutated cells and destroy them. 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 [13]. These data imply that the immune system can detect and destroy cells that have undergone particular cancerous mutations. Thus, for a cancerous mutation of a cell to survive, it must first evade the cells DNA repair mechanism and then the immune system.

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 [14], suggesting that the immunity to prostate cancer in the majority of men predicted here and in reference [1] 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 PSA levels were found to be common among HIV-infected men [15]. Thus, chronic inflammation of the prostate may be a prerequisite for the ordered mutations that eventually lead to cancer to occur [15, 16, 17].

Fig. (4A/IF). SEER 1992-2006 American Indian Female breast cancer incidence rate $IR(t)$ data as a function of age t together with ordered mutation model fit

The possibility that a normal, unsuppressed, functioning immune system can help eliminate malignant melanoma was confirmed by a case report by F. Steven Hodi, Scott Granter, and Joseph Antin [18]. Hodi et al. report on a patient with aplastic anemia who was treated with immuno-suppressive 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 tumor 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 [19].

7. Experimental Support of the Cancer Mutation Model

Is the saturated, ordered, mutation model described in this paper credible?

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 (white males, white females, etc.). Such studies

Fig. (4B/IF/CMPD). SEER 1992-2006 American Indian Female breast cancer incidence rate $IR(t)$ data as a function of age t together with COMPOUND ordered mutation model fit

have already been undertaken for the major cancers.

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* tumor 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 to cause melanoma in the White Male risk group [1]. The modeling results here for breast and prostate cancers certainly meet this requirement.

Colorectal cancer, one of the best understood types of human cancers, has been found to develop *in steps or stages*. The first step in this process is the loss of tumor-suppressor gene adenomatous polyposis coli (APC), which leads to the emergence of a polyp in the colon. The second step is the activation of *ras* oncogene (so named because it was originally isolated from rats with sarcoma). The third step is the loss of another tumor-suppressor gene called DCC for deleted in colorectal cancer which causes the benign polyp to grow into a larger adenoma. The fourth step is the loss of tumor-suppressor gene *p53* (named because this protein had an apparent molecular mass

Fig. (5A/WM). SEER 1992-2006 White Male prostate cancer incidence rate $IR(t)$ data as a function of age t together with ordered mutation model fit

of 53 kilodaltons) which, together with subsequent additional steps, leads to a malignant tumor. Each of these steps may require more than one mutation to cause it. **These results are completely compatible with the modeling results of this paper.**

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.

In 2002 it was discovered that the mutation of the BRAF gene is the key mutation causing most melanoma [20, 21]; 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.

Non-Hodgkin lymphomas (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 [22].

Direct proof that mutations in DNA can cause prostate cancer comes from an interesting controlled experiment by H. Li et al. [23] 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 tumor outgrowth and tumor 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 [24]. 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 3,090 sporadic prostate patients and controls was undertaken by R. K. Nam et al. [25]. 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 [26, 27, 28, 29].

Fig. (5B/WM/CMPD). SEER 1992-2006 White Male prostate cancer incidence rate $IR(t)$ data as a function of age t together with COMPOUND ordered mutation model fit

Fig. (6B/BM/CMPD). SEER 1992-2006 Black Male prostate cancer incidence rate $IR(t)$ data as a function of age t together with COMPOUND ordered mutation model fit

8. Blocking and/or Curing Cancer

Clearly, if mutated, *pre-cancerous* cells can be identified and destroyed, the associated cancer would be prevented from occurring in the first place. It should be easier to destroy a handful of mutated precancerous cells than 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.

Mobuto Yamamoto has discovered a drug that stimulates macrophage to recognize and destroy prostate and breast cancer cells [30, 31, 32], and studies to see if this or other drugs can identify and destroy mutated, *pre-cancerous* cells should become a research imperative.

Cells of some cancers survive the immune systems counterattack in part by secreting Nagalase (alpha-N-acetylgalactosaminidase), a substance that is highly immunosuppressive of macrophage activation. In a study of 16 prostate cancer and 16 breast cancer patients, Nobuto Yamamoto and his collaborators found that the Nagalase activity in these cancer patients varied between 3 and 10 times greater than in the studys healthy controls [31, 32].

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 [30, 31, 32, 33, 34, 35]. Since the number of

Fig. (7B/AM/CMPD). SEER 1992-2006 Asian Male prostate cancer incidence rate $IR(t)$ data as a function of age t together with COMPOUND ordered mutation model fit

cancerous cells in a patient is proportional to the patients 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 CAT-scans [30, 31, 32, 33, 34, 35].

Since significant incidence of breast and prostate cancers occurs only after the age of 50 years, routinely administering a drug like GcMAF to everyone who reaches this age could eliminate all mutated pre-cancerous cells in the body and block the development of these cancers for the over-whelming majority of those susceptible to developing it. **Experimenting with GcMAF as a prophylactic should become a research priority.**

The high prevalence of natural immunity to breast and prostate cancers predicted here and in [1] fits-in with the current concepts of cancer immunosurveillance, part of the more general theory of immunoediting [36, 37, 38, 39, 40, 41, 42].

From direct experiments with mice and humans, natural immunity or effective cancer immunosurveillance has been demonstrated [36]. CD4+ and CD8+ T cells have been shown to recognize tumor 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 tumor growth occurs) [36, 37, 38, 39, 40, 41, 42].

The availability of the human genome sequence has led to the discovery of cancer genes. Most recently, Christopher Greenman et al. have identified 158 cancer-promoting mutations in 119 genes [43] to be added to the approximately 350 cancer genes previously discovered. Thus, the total number of cancer genes discovered to date encompasses the number of mutations required to cause the cancers appearing in Tables (1) and (2) (see column four) resulting from the modeling in this paper.

9. Appendix

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

To construct the 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

Fig. (9) USA breast cancer prevalence $P(t)$ as a function of age t for the White Female risk population as predicted by the compound model fit to the SEER data

for the cell to become cancerous (*e.g.* a skin cell and melanoma). At any age t of a risk population, every cell of an organ of every person in the risk population is in one of the mutation states. A person's risk of developing cancer 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 p^{th} mutation state will be said to be in the p^{th} mutation state. Thus, once a cell of an organ of a person mutates into the m^{th} mutation state, the person develops cancer since it is assumed that only one cell is needed to cause it.

Now consider a number of N_s *newly born* infants that represent a random sample of a risk population 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 p^{th} mutation state by age t will be denoted by $N(p/m, t)$ or $N_p(t)$ for short, where $p = 0, 1, 2, \dots, m$. Assuming that all the infants are in the zeroth mutation state at birth, then at age $t = 0$, $N_p(0) = 0$ for $p = 1, 2, 3, \dots, m$, and $N_0(0) = N_0 = N_s$. A schematic representation of the sub-cohorts $N_p(t)$ at age t is shown in Figure 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 the cohort to

Fig. (10) USA prostate cancer prevalence $P(t)$ as a function of age t for the White Male risk population as predicted by the compound model fit to the SEER data

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 are reciprocals of each other so that $k_1 = T_1^{-1}$. Continuing in this way, the number of members of the cohort $N_p(t)$ who are in the p^{th} mutation state at age t depends on the values of a subset of the m mutation rates $k_1, k_2, \dots, k_m$ as depicted in Figure 1. It will be assumed that all the mutation rates are constants so that the mutations experienced by the cohort *occur randomly*.

If $p = 0, 1, 2, 3, \dots, m - 1$, 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 m quantities k_q , $q = 1, 2, 3, \dots, m$, will be called *mutation rates*. 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 $N_0(t), N_1(t), \dots, N_m(t)$ 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 $0 < f_s < 1$. 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

$$\frac{dN_0(t)}{dt} = -k_1 N_0(t) \quad (9.1)$$

$$\frac{dN_1(t)}{dt} = k_1 N_0(t) - k_2 N_1(t) \quad (9.2)$$

$$\frac{dN_r(t)}{dt} = k_r N_{r-1}(t) - k_{r+1} N_r(t), r = 1, 2, 3, \dots, m-1, \quad (9.3)$$

$$\frac{dN_m(t)}{dt} = k_m N_{m-1}(t). \quad (9.4)$$

Adding these equations together and integrating leads to

$$N_s = N_0(t) + N_1(t) + \dots + N_m(t) = \sum_{p=0}^m N_p(t) \quad (9.5)$$

since $N_p(0) = 0$ for $p = 1, 2, \dots, m$.

The solutions to the equations in (9.1-9.4) in order are

$$N_0(t) = N_s e^{-k_1 t} = f_s N_0 e^{-k_1 t}, \quad (9.6)$$

$$N_r(t) = e^{-k_{r+1} t} \int_0^t k_r N_{r-1}(t') e^{k_{r+1} t'} dt', r = 1, 2, \dots, m-1, \quad (9.7)$$

$$N_m(t) = k_m \int_0^t N_{m-1}(t') dt. \quad (9.8)$$

If $k_q t \ll 1$ for $q = 1, 2, \dots, m$, then the equations in (9.6-9.8) yield the approximate solution

$$N_m(t) = f_s N_0 k_1 k_2 k_3 \dots k_m \left(\frac{t^m}{m!} \right), k_q t \ll 1. \quad (9.9)$$

Thus, the value of m , the number of mutations necessary to cause a particular cancer, can be determined by fitting the function in (9.9) to the *early part* of the $N_m(t)$ data curve. If all the mutation rates k_q , $q = 1, 2, \dots, m$, are equal to the same constant k , then the solutions in (9.7) yield the particularly simple solutions

$$N_r(t) = f_s N_0 \frac{k^r}{r!} t^r e^{-kt}, r = 1, 2, \dots, m-1, \quad (9.10)$$

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

$$N_m(t) = f_s N_0 \left[1 - e^{-kt} \sum_{r=0}^{m-1} \frac{(kt)^r}{r!} \right]. \quad (9.11)$$

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 (9.8) and (9.10) or (9.11), 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

$$IR(m, k; t) \equiv \frac{dN_m(t)}{dt} = \frac{f_s N_0 k^m}{(m-1)!} t^{(m-1)} e^{-kt}, \quad (9.12)$$

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 (9.12) to cancer incidence data. Thus, the fit involves determining 3 independent parameters in general. If every member of the population is susceptible to developing a particular cancer, then $f_s = 1$ and fit involves determining the values of only two parameters, m and k .

Breast Cancer fits: SEER cohort & figure	Mutation rate k (in years ⁻¹)	Mutation lifetime $T = k^{-1}$ (in years)	Mutation numbers m_1 & m_2	Susceptible fractions f_{s1} & f_{s2}	Total percent Immune to cancer	Fit error (Chisq)* (years) ⁻²
White Females (1B).	0.360	2.77	$m_1 = 30$ $m_2 = 21$	$f_{s1} = 0.166$ $f_{s2} = 0.0686$	76.5%	150.7
Black Females (2B).	0.336	2.97	$m_1 = 28$ $m_2 = 19$	$f_{s1} = 0.141$ $f_{s2} = 0.0609$	79.7%	333.0
Asian Females (3B).	0.368	2.71	$m_1 = 30$ $m_2 = 21$	$f_{s1} = 0.0793$ $f_{s2} = 0.0549$	86.5%	226.7
Indian Females (5B).	0.385	2.59	$m_1 = 32$ $m_2 = 22$	$f_{s1} = 0.0892$ $f_{s2} = 0.0524$	85.8%	333.6

Table (1). **Female breast cancer** model parameters stemming from least-squares fits to SEER incidence data for White, Black, Asian, and Indian risk populations using the compound mutation model in equation (1).

* See equation (5f) in the appendix

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 (9.12) monotonically increases, peaks, and monotonically declines towards zero as the age of the risk population increases. The peak in the incidence function in (9.12) occurs at the age of $t_{max} = (m-1)/k$ with the value of

$$IR(m, k; t_{max}) = \frac{f_s N_0 k}{(m-2)!} (m-1)^{(m-2)} e^{-(m-1)}. \quad (9.13)$$

Prostate Cancer fits: SEER cohort & figure	Mutation rate k (in years ⁻¹)	Mutation lifetime $T = k^{-1}$ (in years)	Mutation numbers m_1 & m_2	Susceptible fractions f_{s1} & f_{s2}	Total percent Immune to cancer	Fit error (Chisq)* (years) ⁻²
White Males (5B).	0.559	1.71	$m_1 = 43$ $m_2 = 33$	$f_{s1} = 0.321$ $f_{s2} = 0.00406$	67.4%	840.7
Black Males (6B).	0.512	1.95	$m_1 = 39$ $m_2 = 33$	$f_{s1} = 0.478$ $f_{s2} = 0.0165$	50.5%	871.4
Asian Males (7B).	0.638	1.56	$m_1 = 51$ $m_2 = 39$	$f_{s1} = 0.218$ $f_{s2} = 0.00436$	77.7%	846.3
Indian Males (8B).	0.400	2.49	$m_1 = 33$ $m_2 = 32$	$f_{s1} = 0.0982$ $f_{s2} = 0.114$	78.6%	1699

Table (2). **Prostate cancer** model parameters stemming from least-squares fits to SEER incidence data for White, Black, Asian, and Indian risk populations using the compound mutation model in equation (1).

*See equation (5f) in the appendix

Cancer vulnerability	Cancer type	Allele	Gene 1	Gene 2	...	Gene $i < m/2$	...	Gene $h < m/2$	...	Gene $m/2$
Susceptible	Hereditary	1						X	O	O
		2						$O \rightarrow X$	O	O
Susceptible	Sporadic	1	$O \rightarrow X$	O	O	O	O	O	O	O
		2	O	O	O	O	O	O	O	O
Immune		1	$O \rightarrow X$	O	O	I				
		2	O	O	O	I				

Table (3). Status of proto-oncogenes **at birth** in ordered mutation model of hereditary and sporadic cancer in the Knudson carcinogenesis model. Mutations of genes occur in *ascending* order, first in allele 1 and subsequently in allele 2. An “O” indicates an un-mutated allele that is nonetheless susceptible to developing a cancerous mutation. An “X” indicates that a cancerous mutation has occurred in a given allele of a given gene, say gene h , and an “I” indicates that a cancerous mutation is blocked in a specific gene, say gene i . An “ $O \rightarrow X$ ” indicates the only un-mutated allele that can experience the next mutation.

If the set of n consecutive *data* points used in the fit are denoted by $[d_i]$ and if the corresponding model fit for these 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

$$\chi^2 \equiv \sum_{i=1}^n |x_i - d_i|^2. \quad (9.14)$$

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

If $kt \ll 1$ in (9.11), then the leading term in the solution there is

$$N_m(t) = f_s N_0 \frac{(kt)^m}{m!}, \quad (kt) \ll 1. \quad (9.15)$$

Equating (9.15) and (9.9) leads to the following relationship between the *average* mutation rate k and the set of m mutation rates $k_1, k_2, \dots, k_m$:

$$k^m = k_1 k_2 k_3 \dots k_m. \quad (9.16)$$

The *physical* model that produces the result in (9.12) 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 (9.6-9.8). For example, if $m = 1$, then using (9.6) and remembering that for $q = 1, 2, \dots, m$

$$IR(1, k_r; t) \equiv \frac{dN_1(t)}{dt} = k_1 N_0(t) = f_s N_0 k_1 e^{-k_1 t} \quad (9.17)$$

and

$$N_1(t) = f_s N_0 [1 - e^{-k_1 t}]. \quad (9.18)$$

Similarly, if $m = 2$, then

$$IR_2(2; k_r; t) \equiv \frac{dN_2(t)}{dt} = k_2 N_1(t) = f_s N_0 \frac{k_1 k_2}{(k_2 - k_1)} [e^{-k_1 t} - e^{-k_2 t}] \quad (9.19)$$

and

$$N_2(t) = f_s N_0 \frac{k_1}{(k_2 - k_1)} [e^{-k_1 t} - e^{-k_2 t}]. \quad (9.20)$$

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

$$IR_3(3; k_r; t) \equiv \frac{dN_3(t)}{dt} = f_s N_0 \frac{k_1 k_2 k_3}{(k_2 - k_1)} \left[\frac{(e^{-k_1 t} - e^{-k_3 t})}{(k_3 - k_1)} - \frac{(e^{-k_2 t} - e^{-k_3 t})}{(k_3 - k_2)} \right]. \quad (9.21)$$

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

Notice that the cancer incidence functions in (9.17), (9.19), and (9.21) 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 reach 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. 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 $N(p/m, t)$, $p = 0, 1, 2, \dots, m$, where m is the number of mutations necessary to cause a particular cancer in a characteristic cohort. If all the mutation rates, $q = 1, 2, \dots, m$, are equal to the same constant k , then the average time T_c it would take those that are susceptible to developing cancer to actually get it is given by

$$T_c = \frac{\int_0^\infty t dN_m(t)}{\int_0^\infty dN_m(t)} = \frac{m}{k} = mT, \quad (9.22)$$

where (9.12) was used to get this result.

Finally, the fraction of a random sample of a risk population at age t that has developed a specific cancer, also known as the prevalence $P(t)$ of the cancer, is given by

$$P(t) = \frac{N_m(t)}{N_0} = \int_0^t \frac{IR(t')}{N_0} dt', \quad (9.23)$$

where $IR(t')$ is the incidence rate in a shortened notation.

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 [44]. 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 [45]; 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 [46] could generate similar age-specific incidence curves to those in reference [43]. 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) [47, 48, 49]. One problem with this model is that not all of the parameters it must introduce can be determined from incidence data alone [50]. A good example of how this model is applied in practice is contained in a recent paper by Luebeck and Moolgavkar [51]. The appendix for reference [51] contains all of the mathematical details necessary to make this model work; clearly, to fit this models 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 (9.12) to cancer incidence data.

References

- [1] I. Kramer. *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*. Computational and Mathematical Methods in Medicine, 9 (2008), No. 1, 1-26.
- [2] I. Kramer. *What triggers transient AIDS in the acute phase of HIV infection and chronic AIDS at the end of the incubation period?: A model analysis of HIV infection from the acute phase to chronic AIDS stage*. Computational and Mathematical Methods in Medicine, 8 (2007), No. 2, 125-151.
- [3] I. Kramer. *Calculating the number of people with Alzheimer disease in any country using saturated mutation models of brain cell loss that also predict widespread natural immunity to the disease*. Computational and Mathematical Methods in Medicine, Oct. 5, 2009.
- [4] A. G. Knudson, Jr. *Mutation and cancer: statistical study of retinoblastoma*. Proc. Nat. Acad. Sci., USA, 68 (1971), No. 4, 820-823.
- [5] S. H. Friend, R. Bernards, S. Rogelj, R. A. Weinberg, J. M. Rapaport, D. M. Albert, T. P. Dryja. *A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma*. Nature, 323 (1986), 643-646.
- [6] A. T. Yeung, B. B. Patel, X.-M. Li, S. H. Seeholzer, R. A. Coudry, H. S. Cooper, A. Bellacosa, B. M. Boman, T. Zhang, S. Litwin, E. A. Ross, P. Conrad, J. A. Crowell, L. Kopelovich, A. Knudson. *One-hit effects in cancer: altered proteome of morphologically normal colon crypts in familial adenomatous polyposis*. Cancer Research, 68 (2008), 7579-7586.
- [7] S. Monfardini, E. Vaccher, G. Pizzocaro. *Unusual malignant tumors in 49 patients with HIV infection*. AIDS, 3 (1989), 449-452.
- [8] S.C. Remick. *Non-AIDS-defining cancers*. Hematol Oncol Clin North Am., 10 (1996), 1203-1213.
- [9] C. Smith, S. Lilly, K.P. Mann. *AIDS-related malignancies*. Ann. Med., 30 (1998), 323-344.
- [10] T.P. Cooley. *Non-AIDS-defining cancer in HIV-infected people*. Hematol Oncol Clin North Am., 17 (2003), 889-899.
- [11] S.M. Mbulaiteye, R.J. Biggar, J.J. Goedert. *Immune deficiency and risk for malignancy among persons with AIDS*. J Acquir Immune Defic Syndr., 32 (2003), 527-533.

- [12] M. Frish, R.J. Biggar, E.A. Engels. *AIDS-Cancer Match Registry Study Group Association of cancer with AIDS-related immunosuppression in adults*. JAMA, 285 (2001), 1736-1745.
- [13] F. Stephen Hodi, S. Granter, J. Antin. *Withdrawal of immunosuppression contributing to the remission of malignant melanoma: a case report*. Cancer Immunity, 5 (2005), 7.
- [14] N.F. Crum. *Increased risk of prostate cancer among HIV-infected men*. Contagion, Vol 2 (2005), No. 2, 66-70.
- [15] W.G. Nelson, A.M. De Maizo, W.B. Issacs. *Mechanisms of disease: prostate cancer*. N Engl J Med, 349 (2003), 366-381.
- [16] E.A. Platz, A.M. De Maizo. *Epidemiology of inflammation and prostate cancer*. J Urol, 171 (2004), S36-S40.
- [17] A.M. De Maizo, V.L. Marchi, J.L. Epstein. *Proliferative inflammatory atrophy of the prostate: implications for prostatic carcinogenesis*. Am J Pathol, 155 (1999), 1985-1992.
- [18] F. Steven Hodi, S. Granter, J. Antin. *Withdrawal of immunosuppression contributing to the remission of malignant melanoma: a case report*. Cancer Immunity, 5 (2005), 7.
- [19] J. Ruvinsky. *Are You Immune to Cancer?*. DISCOVER, 27 (2006), No. 8.
- [20] H. Davies. *Mutations of the BRAF gene in human cancer*. Nature, 417 (2002), 949-54.
- [21] M.S. Brose, P. Volpe, M. Feldman, M. Kumar, I. Rishi. *BRAF and RAS mutations in human lung cancer and melanoma*. Cancer Res., 62 (2002), No. 23, 6997-7000.
- [22] K.L. Novik, J.J. Spinelli, A.C. Macarthur, K. Shumansky, P. Sipahimalani, A. Lai, J.M. Connors, R.D. Gascoyne, R.P. Gallagher, A.B. Brooks-Wilson. *Genetic variation in H2AFX contributes to risk of non-Hodgkin lymphoma*. Cancer Epidemiol Biomarkers Prev, 16 (2007), No. 6, 1098-106.
- [23] H. Li, Y. Gu, B. Hukku, D.G. McLeod, T.K. Hei, J.S. Rhim. *Malignant transformation of human benign prostate epithelial cells by high linear energy transfer alpha-particles*. Int J. Oncol., 31 (2007), No. 3, 537-44.
- [24] P. Lichtenstein, N.V. Holm, P.K. Verkasalo, A. Iliadou, J. Kaprio, M. Koskenvuo, E. Pukkala, A. Skytthe, K. Hemminki,. *Environmental and heritable factors in the causation of cancer - analyses of cohorts of twins from Sweden, Denmark, and Finland*. N Engl J Med, 343 (2000), No. 2, 78-85.
- [25] R.K. Nam, W.W. Zhang, D.A. Loblaw, L.H. Klotz, J. Trachtenberg, M.A. Jewett, A. Stanimirovic, T.O. Davies, A. Toi, V. Venkateswaran, L. Sugar, K.A. Siminovich, S.A. Narod. *A genome-wide association screen identifies regions on chromosomes 1q25 and 7p21 as risk loci for sporadic prostate cancer*. Prostate Cancer Prostatic Dis., 2007 Sep 18, to be published.]

- [26] A. Vecchione, F. Gottardo, L.G. Gomella, B. Wildemore, M. Fassan, E. Bragantini, F. Pagano, R. Baffa. *Molecular genetics of prostate cancer: clinical translational opportunities*. J Exp Clin Cancer Res, 26 (2007), No. 1, 25-37.
- [27] T.M. Lane, J.C. Strefford, R.J. Yanez-Munoz, P. Purkis, E. Forsythe, T. Nia, J. Hines, Y.L. Lu, R.T. Oliver. *Identification of a recurrent t(4;6) chromosome translocation in prostate cancer*. J Urol, 177 (2007), No. 5, 1907-12.
- [28] O. Saramaki, T. Visakorpi. *Chromal aberrations in prostate cancer*. Front Biosci, 12 (2007), 3287-301.
- [29] I.M. Berguin, Y. Min, B. Wu, J. Wu, D. Perry, J.M. Cline, M.J. Thomas, T. Thornberg, G. Kulik, A. Smith, I.J. Edwards, R. D'Agostino, H. Zhang, H. Wu, J.X. Kang, Y.Q. Chewn. *Modulation of prostate cancer genetic risk by omega-3 and omega-6 fatty acids*. J Clin Invest, 117 (2007), No. 7, 1866-75.
- [30] N. Yamamoto, M. Ueda. *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, 2004.
- [31] N. Yamamoto, H. Suyama, N. Yamamoto, N. Ushijima. *Immunotherapy of metastatic breast cancer patients with vitamin D-binding protein-derived macrophage activating factor (GcMAF)*. Int. J. Cancer, 122 (2008), 461-467.
- [32] N. Yamamoto, H. Suyama, N. Yamamoto. *Immunotherapy for Prostate Cancer with Gc Protein-Derived Macrophage-Activating Factor, GcMAF*. Translational Oncology, 1 (2008), No. 2, 65-72.
- [33] N. Yamamoto. *Pathogenic significance of acetylgalactosaminidase Activity found in the Envelope Glycoprotein gp160 of Human Immunodeficiency Virus Type 1*. AIDS Research and Human Retroviruses, 22 (2006), No. 3, 262-271.
- [34] N. Yamamoto, M. Ueda. *Pathogenic significance of acetylgalactosaminidase activity found in the hemagglutinin of influenza virus*. Microbes and Infection, 7 (2005), No. 4, 674-681.
- [35] N. Yamamoto, M. Ueda. *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.
- [36] G.P. Dunn, L.J. Old, R.D. Schreiber. *The Immunobiology of Cancer Immunosurveillance and Immunoediting*. Immunity, 21 (2004), 137-148.
- [37] G.P. Dunn, A.T. Bruce, H. Ikeda, L.J. Old, R.D. Schreiber. *Cancer immunoediting: from immunosurveillance to tumor escape*. Nature immunology, 3 (2002), 991-998.

- [38] G.P. Dunn, C.M. Koebel, R.D. Schreiber. *Interferons, immunity, and cancer immunoediting*. Nat Rev Immunol., 6 (2006), No. 11, 836-48.
- [39] Q. Zhihai, T. Blankenstein. *A cancer immunosurveillance controversy*. Nature Immunology, 5 (2004), 3-4.
- [40] L. Zitvogel, A. Tesniere, G. Kroemer. *Cancer despite immunosurveillance: immunoselection and immunosubversion*. Nat Rev Immunol., 6 (2006), No. 10, 715-27. Epub 2006 Sep 15.
- [41] A.E. Germainis, V. Karanikas. *Immunoepigenetics: the unseen side of cancer immunoediting*. Immunol Cell Biol., 85 (2007), No. 1, 55-9, Epub 2006 Nov 28.
- [42] B.A. Inman, X. Frigola, H. Dong, E.D. Kwon. *Costimulation, coinhibition and cancer*. Curr Cancer Drug Targets, 7 (2007), No. 1, 15-30.
- [43] C. Greenman. *Patterns of somatic mutation in human cancer genomes*. Nature, 446 (2007), 153-158.
- [44] P. Armitage, R. Doll. *The age distribution of cancer and a multi-stage theory of carcinogenesis*. British Journal of Cancer, 8 (1954), No. 1, 1-12 .
- [45] F. Pompei, R. Wilson. *The age distribution of cancer; the turnover at old age*. Health and Environmental Risk Assessment, 7 (2001), No. 6, 1619-50.
- [46] P. Armitage, R. Doll. (1957). *A Two-Stage Theory of Carcinogenesis in Relation to the Age Distribution of Human Cancer*. Br. J. Cancer, 11 (1957), No. 2, 161-169.
- [47] S. Moolgavkar, A. Knudson. (1981). *Mutation and cancer: a model for human carcinogenesis*. J. Natl. Cancer Inst., 66, No. 6, 1037-1052.
- [48] S.H. Moolgavkar, A. Dewanji, D.J. Venzon. (1988). *A stochastic two-stage model for cancer risk assessment: I. The hazard function and the probability of tumor*. Risk Anal., 8, No. 3, 383-392.
- [49] S.H. Moolgavkar, E.G. Luebeck. (1990). *Two-event model for carcinogenesis: biological, mathematical, and statistical considerations*. Risk Anal., 10, No. 2, 323-341.
- [50] W.F. Heidenreich, E.G. Luebeck, S.H. Moolgavkar. (1997). *Some properties of the hazard function of the two-mutation clonal expansion model*. Risk Anal., 17, No. 3, 391-399.
- [51] E.G. Luebeck, S.H. Moolgavkar. *Multistage carcinogenesis and the incidence of colorectal cancer*. PNA, 99 (2002), No. 23, 15095-15100.