On substituting $r = 7$, we obtain the expression given earlier in the paper.

It is easy to show that $w(t_0)$ is a maximum when $t_0 = (s − 1) t/(r − 2)$. Substituting $r = 7$ and $s = 1$, the maximum weight occurs when $t_0 = 0$; this means that if p_1 changes with time, the incidence rate at any given age is proportional to the weighted mean of p_1 over the whole period of exposure, the weight being greatest near the beginning of the period. Similarly, substituting $r = 7$ and $s = 2, 3, 4, 5$ and 6, we find that the maximum weight occurs when $t_0 = t/5, 2t/5, 3t/5, 4t/5$ and t, respectively.

In applying this theory to the data of Lenowitz and Graham (1946) and Shrek and Lenowitz (1947), we must assume that the distribution of the ages at circumcision in the patients without penile cancer is representative of that in the population from which the penile cancer patients have been drawn; on this assumption the data of Table II provide valid estimates of the relative risks of penile cancer in the four groups. We shall ignore the variation in age amongst the patients, and consider the expected incidence of penile cancer at $t = 53$ years (the average age of all the patients referred to in Table II). We consider a theoretical model in which $r = 7$ and $s = 1$; and in which p_1 is constant for an uncircumcised person, but becomes zero as soon as circumcision is carried out. For each of the four age groups shown in Table II, we assume, for simplicity, that circumcision takes place at the mid-point of the group, i.e. at 3, 11.5, 26.5 and 44.5 years, respectively. From the theoretical result proved above, the incidence at $t = 53$ is proportional to the weighted mean of p_1 over the period of exposure, the weight corresponding to exposure at age t_0 being

$(53 − t_0)^5$. The incidences for the four groups are therefore proportional to

$$\int_0^T (53 − t_0)^5 \, dt_0 = \frac{1}{6} \{(53)^6 − (53 − T)^6\},$$

where T is the mid-point of each group. Using the values of T given above, the incidences are found to be proportional to 0.30, 0.77, 0.98 and 1.00.

A more realistic assumption would be that the men in the first group were almost all circumcised within the first few weeks of life; on this assumption the theoretical risk for this group would clearly be very small.

References

Armitage, P.—(1953) *J. R. statist. Soc.* Series B, **15**, 90.

Berenblum I., and Shubik P.—(1949) *Brit. J. Cancer*, **3**, 109.

Fisher J. C., and Hollomon J. H.—(1951) *Cancer*, **4**, 916

Kennaway, E. L.—(1947) *Brit. J. Cancer*, **1**, 335

Lenowitz H, and Graham AP—(1946) *J. Urol.*, **56**, 458

Lombard HL, and Potter EA—(1950) *Cancer*, **3**, 960

Nordling, C. O.—(1953) *Brit. J. Cancer*, **7**, 68

Registrar General of England and Wales.—(1952) 'Statistical Review of England and Wales for the year 1950'. Tables. Part I. Medical. London (H. M. Stationery Office). – (1953) 'Statistical Review of England and Wales for the year 1951.' Tables. Part I. Medical. London (H. M. Stationery Office).

Shrek R, and Lenowitz H.—(*Cancer Res.*, 1947) **7**, 180.

Armitage P, Doll R. The age distribution of cancer and a multistage theory of carcinogenosis. *Br J Cancer* 1954;**8**:1–12. Reprinted with permission.

Commentary: Mathematical models of cancer progression and epidemiology in the age of high throughput genomics

Steven A Frank

The major principles of genetics and evolution were developed before the structure of DNA was discovered in 1953. Darwin got pretty much everything right about evolution, despite his mistaken views on genetics. After the rediscovery of Mendel's theory, Fisher, Wright, and Haldane worked out the mathematical principles of mutation, selection, and evolutionary genetics during the first half of the 20th century. The spectacular accomplishments of modern molecular biology have greatly enriched understanding of genetics and evolution. However, the foundational principles of evolution and adaptation from the pre-molecular era remain the guidelines by which we interpret why biology appears as it does.

Perhaps Armitage and Doll's[1] paper marks the same sort of divide in cancer research. Their paper laid out foundational principles of cancer progression and epidemiology in

Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697-2525 USA. E-mail: safrank@uci.edu

mathematical form long before we knew about the molecular basis of somatic mutation and the key roles of genes such as *p53* and *APC*.

The puzzle faced by Armitage and Doll was set out by Fisher and Hollomon[2] and Nordling,[3] who used epidemiological data to infer that cancer incidence increases with approximately the sixth power of age. The pioneers of genetics and evolution faced the same sort of problem: how can one use easily observed patterns in populations to infer the underlying dynamical processes that give rise to those patterns? In the case of cancer, what can be said about the dynamical processes of progression within individuals that would explain the aggregate patterns of epidemiology observed in populations?

Fisher and Hollomon recognized that the rate of cancer occurrence would rise with the nth power of age if transformation required $n + 1$ independent steps. The argument is roughly as follows. Suppose each step happens at a rate of u per year, where u is a small rate. The probability of any step having happened after t years is ut. At age t, the probability that n of the steps has occurred is $(ut)^n$, and the rate at which the final step happens is u, so the rate of occurrence at time t is proportional to $u^{n+1}t^n$. For example, if $n + 1 = 2$, then the probability that one step has occurred at time t is ut, and the rate at which the final step happens is u, so the rate at time t is proportional to $u^2 t$.

Fisher and Hollomon suggested that about seven cells had to be transformed independently. This would give the observed rate of change with age. However, Armitage and Doll[1] pointed out that if transformation happens by one step in each of several independent cells, then tumour incidence should increase with about the sixth power of carcinogen dose. Instead, incidence rises approximately linearly with dose, rejecting the multiple cell hypothesis.

Nordling proposed that approximately seven successive steps must occur over the history of a transformed cell lineage. Armitage and Doll[1] noted that if the steps tend to occur in a particular order, this theory explains the rise in incidence with the six power of age, the linear relation between carcinogen dose and incidence, and the long time delay that usually occurs between carcinogen exposure and transformation.

Armitage and Doll followed this introduction with an analysis of incidence curves for cancers of various tissues and a discussion of how to think about the mathematical theory in terms of the biology. Several incidence curves matched the theory with about seven steps and constant rates of transition. Other curves failed to fit the simple theory, and may possibly be explained by fewer steps, with the rates of transition between steps increasing with age. Armitage and Doll emphasized that the steps represent stages in transformation, each step separated from the next stage by some rate-limiting event. Those events might be mutations, but other factors in transformation could also serve as rate-limiting events.

An extensive mathematical literature has refined this theory and fit various models to more detailed sets of data. Most discussions of cancer progression depend in some way on a multistage theory of progression, although opinions vary widely about the nature of those steps and which biological processes may be most important.

Returning to my opening theme, Armitage and Doll's multistage theory developed the major concepts for how to think about incidence, carcinogenesis, and progression. They
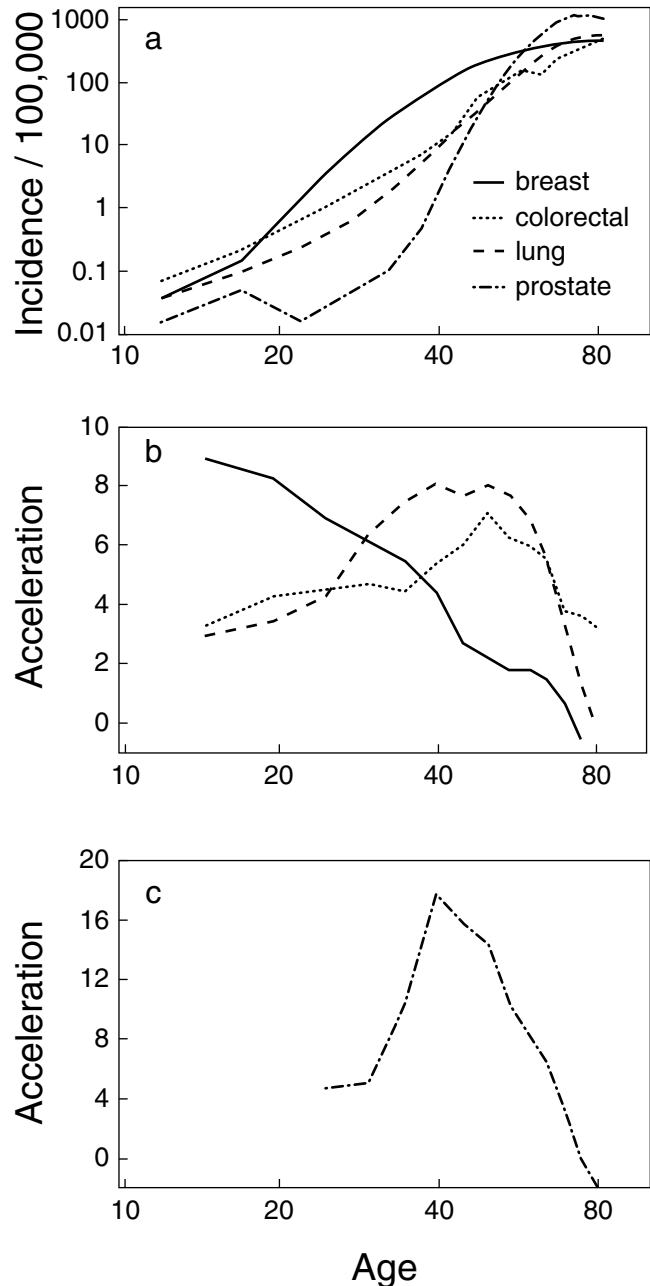

Figure 1 Age-specific incidence and acceleration of cancer for different tissues. (a) Age-specific incidence for four adult-onset epithelial cancers. (b, c) Age-specific log-log acceleration, which are slopes of the plots in (a) at different ages. All data from the SEER database (www.seer.cancer.gov) using the nine SEER registries, year of diagnosis 1992–2000. Breast cancer data for all females, colorectal cancer for all males, lung cancer for all males, and prostate cancer for white males. Reprinted from ref. 4.

did this while almost nothing was known about the genetic, molecular, and cellular mechanisms of progression. In this current age of high throughput genomics and the promise of soon knowing much about the mechanistic details of progression, will Armitage and Doll's insights continue to play a role in shaping the subject, or is this an historical footnote with little consequence for modern studies of cancer?

I believe that extensions of the Armitage-Doll mathematical theory will play an important role in the future of cancer studies. I give two examples of how this might occur.

The first example moves from detailed genetic observations up toward a full understanding of the dynamic processes that drive cancer progression. A typical tumour cell has a large number of genetic changes throughout its genome when compared with the normal, ancestral cell from which the tumour cell evolved. Many tumour lineages suffer periods of chromosomal instability that cause major karyotypic changes. Probably several of those karyotypic changes play an important role in transformation, whereas many other genetic changes have little or no effect on the success of tumour lineages.

Perhaps the most important goal of high throughput genetic analysis will be to determine which changes matter and which are less important. The definition of 'which changes matter' should be a quantitative, dynamical one. How do particular genetic changes accelerate or decelerate progression? How does the order of changes and the particular combination of changes affect rates of progression? Mathematical modelling will be important in connecting the genetic changes to the associated biochemical pathways and to the consequences for cellular birth and death rates. Those new mathematical models may become rather detailed and complex, but to be useful they must be brought back to the simple mathematical structure of Armitage and Doll—the stages of progression, the rates of change between stages, and the consequences for the rate of incidence in populations.

The second example works in the other direction, from simple mathematical models of progression like Armitage and Doll's to testable hypotheses about the nature of the underlying genetic and cellular mechanisms that drive progression. I recently examined the SEER database (www.seer.cancer.gov),[4] a more detailed data set on cancer incidence than was available to Armitage and Doll. Figure 1 shows four of the most common human cancers. The top panel plots the incidence data in the standard log-log style, for which the average slopes are about 5 or 6, corresponding to Armitage and Doll's model with six or seven rate-limiting steps. The bottom two panels plot the same data, showing the slope of the top graph at each point in time.

The top graph in Figure 1 is the cancer incidence, or rate, and thus the slope of the rate in the bottom panels is the acceleration of cancer incidence at each age. I plotted the prostate data in a separate panel at the bottom, because the scale differs from the other cancers. For all four cancers, the acceleration drops linearly in the later part of life, during which nearly all cases occur. This observation of a linear decline in acceleration sets a constraint that any mechanistic model of progression must be able to explain, just as Armitage and Doll used the pattern of incidence in response to carcinogens to reject Fisher and Hollomon's theory and support Nordling's idea about progression.

I do not know the explanation for the steady decline in acceleration in the latter half of life. However, I did develop a mathematical extension to Armitage and Doll's analysis that suggests a hypothesis about why the decline occurs.[4] Briefly, suppose that $n + 1$ rate-limiting steps must be passed before a cell lineage is transformed. At birth, one has a large number, N, of mostly pristine cells. As time passes to midlife, some of the those N cell lineages have accumulated, for example, m of the necessary changes for transformation. Those cells have $n + 1 - m$ steps remaining, and will be transformed at a rate that increases with the $(n - m)$th power of time instead of the nth power of time that described the transformation rate earlier in life. So, the hypothesis is that different cell lineages will be passing various steps independently, and in midlife a person without cancer will have progressed partway. That could be tested by high throughput genetic studies of normal cells at different times of life.

Figure 1 also shows a midlife rise in acceleration for three of the four cancers. I discuss that elsewhere.[4] The point here is that mathematical models can suggest new hypotheses about progression. In recent years, such top-down models have played relatively little role in molecular biology. That limitation probably occurred because most recent studies have focused on working out detailed aspects of molecular mechanisms, for which quantitative theories provide little insight. But now that interest has shifted to how various mechanisms combine to determine the behaviour of complex systems, a quantitative perspective such as the one provided by Armitage and Doll 50 years ago may become increasingly important for understanding how particular mechanisms contribute to cancer progression.

References

[1] Armitage P, Doll R. The age distribution of cancer and a multi-stage theory of carcinogenesis. *Brit J Cancer* 1954;**8**:1–12. (Reprinted *Int J Epidemiol* 2004;**33**:1174–79.)

[2] Fisher JC, Hollomon JH. A hypothesis for the origin of cancer foci. *Cancer* 1951;**4**:916–18.

[3] Nordling CO. A new theory on the cancer-inducing mechanism. *Brit J Cancer* 1953;**7**:68–72.

[4] Frank SA. Age-specific acceleration of cancer. *Curr Biol* 2004;**14**:242–46.