Spora: A Journal of Biomathematics

Volume 8 Article 3

2022

An Axiomatic and Contextual Review of the Armitage and Doll Model of Carcinogenesis

W. Zane Billings
Western Carolina University, wesley.billings@uga.edu

Justin Clifton
Western Carolina University, jdclifton2@catamount.wcu.edu

Josh Hiller
Adelphi University, johiller@adelphi.edu

Tommy Meek
Western Carolina University, thmeek1@catamount.wcu.edu

Andrew Penland
Western Carolina University, adpenland@email.wcu.edu

See next page for additional authors

Recommended Citation

Billings, W. Zane; Clifton, Justin; Hiller, Josh; Meek, Tommy; Penland, Andrew; Rogers, Wesley; Smokovich, Gabriella; Velasquez-Berroteran, Andrew; and Zamagias, Eleni (2022) "An Axiomatic and Contextual Review of the Armitage and Doll Model of Carcinogenesis," *Spora: A Journal of Biomathematics*: Vol. 8, 7–15.
DOI: https://doi.org/10.30707/SPORA8.1.1647886301.817127
Available at: https://ir.library.illinoisstate.edu/spora/vol8/iss1/3

An Axiomatic and Contextual Review of the Armitage and Doll Model of Carcinogenesis

Cover Page Footnote

The authors gratefully thank the Council for Undergraduate Research in Mathematics for support through NSF grant DMS-1722563

Authors

W. Zane Billings, Justin Clifton, Josh Hiller, Tommy Meek, Andrew Penland, Wesley Rogers, Gabriella Smokovich, Andrew Velasquez-Berroteran, and Eleni Zamagias

Intercollegiate Biomathematics Alliance

SPORA
A Journal of Biomathematics

ISSN 2473-3067 (Print)
ISSN 2473-5493 (Online)

An Axiomatic and Contextual Review of the Armitage and Doll Model of Carcinogenesis

W. Zane Billings[1], Justin Clifton[1], Josh Hiller[2,*], Tommy Meek[1], Andrew Penland[1], Wesley Rogers[1], Gabriella Smokovich [2], Andrew Velasquez-Berroteran[2], Eleni Zamagias[2]

[*]Correspondence:
Prof. Josh Hiller, Dept. of Mathematics and Computer Science, Adelphi University, One South Ave., Garden City, NY 11530, USA
johiller@adelphi.edu

Abstract

In 1954, Armitage and Doll published one of the most influential papers in the history of mathematical epidemiology. However, when one examines the literature one finds that there are in fact at least three distinct mathematical models attributed to the 1954 paper. In this study, we examine this important paper and the mathematical derivation of their model. We find, very surprisingly, that no stochastic process can account for all the assumptions of the model and that many of the models in the literature use a consistent subset of the assumptions used in Armitage and Doll's paper.

Keywords: multistage carcinogenesis, Armitage and Doll, Nordling, cancer, power-law

1 Introduction

In 1954, Peter Armitage and Richard Doll, two British epidemiologists, published a landmark paper [1]. The "multistage model of carcinogenesis," as their idea would come to be known, was the birth of mathematical oncology and more generally of mathematical epidemiology of noncommunicable disease [2].

The power of the multi-stage model is due in large part to its deliberate flexibility: Cancer is the end result of n discrete and irreversible changes which (might need to) occur in order. Yet from a mathematician's perspective, this flexibility translates as confusion: How does one model support such a broad hypothesis? Not surprisingly, many different answers to this question have appeared in the literature. In this paper, we consider the historical works of both Armitage and Doll to help determine their original model. Our ultimate conclusion is that there is no single stochastic process that satisfies all of the mathematical hypotheses that Armitage and Doll listed in their 1954 work. In the next section, we give a historical overview of the lives and contributions of both Armitage and Doll; our third section lists assumptions for the model with the historical context and justification for each assumption. The fourth section begins with the proof of the inconsistency of these assumptions, as well as a detailed overview of some of the mathematical models that are used to represent Armitage and Doll's framework of multistage carcinogenesis. A mathematical appendix with detailed proofs is included for those who are interested in seeing a rigorous justification for the results.

2 Nordling's Work: A Prelude to the Armitage and Doll Model

The Armitage and Doll model was a careful redefinition and derivation of a slightly earlier paper by Nordling [3]. Nordling wished to explain a phenomenon that had been found even earlier across several European countries [4, 5]: An inexplicable power law (a relationship of the form $y = ax^n$ for constants a, the scale parameter, and n, the degree of the power law) of degree 6 in cancer mortality with age. Later in life, Doll lamented that Nordling did not receive the recognition he was due for his truly groundbreaking idea [6]. We posit that Nordling's work did not receive the attention until much later for four basic reasons:

1. Nordling's data analysis methodology,
2. the specificity of his biological hypothesis,
3. the specificity of the phenomena that Nordling wished to explain,
4. a lack of mathematical transparency.

In the rest of this section, we will explain each of these assertions.

2.1 Data analysis methodology

Nordling's data analysis grouped mortality from all cancers together as if cancer were a single homogenous disease.

[1]Department of Mathematics and Computer Science, Western Carolina University, Cullowhee, NC, [2]Department of Mathematics and Computer Science, Adelphi University, Garden City, NY

This was not an uncommon practice of the time: It is exactly the methodology that had been employed by earlier researchers in Europe. However, physicians and several epidemiologists of the time understood that cancer was in fact several similar illnesses most easily characterised by their host tissue. To suggest that lung cancer and cancer of the jaw, for example, were essentially the same was very likely too far a stretch for clinicians to accept. This must have been understood by Armitage, Doll, and others who reported that the degree six power law held even when data was analyzed for many different cancers.

2.2 Biological hypothesis

Nordling's biological hypothesis was that cancer was caused by seven irreversible mutations. This hypothesis was also suggested independently by Herman Muller in 1951 [7]. Muller's work at that time, built off of his work from the 1920s where he established that radiation exposure can give rise to mutations [8]. The issue is that this seemed to contradict many decades of laboratory research. Even though the hypothesis that cancer was the result of genetic instability was not new (in fact this hypothesis dates back to at least the work of McCombs and McCombs in 1930 [9] and may have predated even that by more than a decade [6]), the number of mutations contradicted lab findings that showed that one or two mutations were sufficient [10]. These findings had made one and two stage models of carcinogenesis the established norm of the day [6, 10]. Besides this, Nordling's biological hypothesis makes no mention of order with respect to mutations. This seemed to discard experiments, dating back to the 1880s, showing two fundamentally different types of mutagens: initiators and promoters [11].

2.3 Specificity of the phenomena that Nordling wished to explain

Nordling only sought to explain a power law which seemed prevalent in data across multiple countries. However, he made no attempt to examine cancer incidence for cancers of the sexual organs, which Armitage and Doll addressed explicitly, nor did he try to explain how his theory might be modified for exposure to carcinogenic substances at different points in an individual's life.

2.4 Mathematical transparency

The last reason we posit that Nordling's work had little chance of gaining mainstream notoriety is because there was no mathematical justification in his paper—just a reference stating that "several successive mutations in the same cell, probably about seven in the case of human cancer, would be necessary" (page 69, [3, 69]). This is in sharp contrast to Armitage and Doll who included a simplified mathematical derivation in the text of the paper, a mathematical appendix with further details, and a reference to an earlier paper by Doll which further developed the mathematical concepts [1].

These specific limitations must have been at least partially understood and appreciated at the time of Nordling's publication. Within the same volume of the British Journal of Cancer (but a later issue) we see an attempt to correct these shortcomings by P. Stocks [12]. In his paper, Stocks gives a careful mathematical derivation for the incidence of stomach cancer and exchanges the term "mutation" for the term "encounter." However, the author is very clear in his closing paragraph to say that "with these [minor] reservations there appears to be no important conflict between the hypotheses [of Stocks and Nordling] themselves."

3 Historical Context of the Armitage and Doll Assumptions

For the purposes of this paper we use two different categories of assumptions involved in the Armitage and Doll model: explicit and implicit. The explicit assumptions are those that Armitage and Doll listed either in the main text of their paper or in its mathematical appendix. However, these assumptions alone are not enough to build a single mathematical model. Therefore, we will also appeal to implicit assumptions. These are the assumptions listed in either paper which are referenced as providing additional information in the mathematical appendix or information gleamed from papers published by Armitage or Doll later in life. The explicit assumptions listed are as follows:

1. Cancer is the result of several discrete changes with very low rates of occurrence,
2. Changes are stable (i.e irreversible),
3. Changes must proceed in a unique order,
4. Mortality is a good indicator for incidence (i.e. the waiting time from first cancer cell to death by cancer is short and deterministic),
5. Changes occur independently (i.e. can occur in any order but only cause cancer if they occur in a specific order),
6. The probability of a specific change occurring in a given interval is only a function of the length of the interval and not when the interval occurs (i.e. independent increments).

Armitage and Doll point to another paper [13] which adds the assumptions that the probability of the i^{th} change occurring in an interval of length Δ is $p_i\Delta + o(\Delta)$, where

$o(\Delta)$ is the standard little o-notation, and that the process was viewed as a pure-birth process with n stages. Thus if the full population is able to contract cancer (assuming an infinite life span), then risk should be asymptotically constant [2].

Of course, there is another possibility which would imply that risk should decrease after some period of time: if some proportion of the population is immune to cancer risk. However, in a later commentary on their model, Doll seemed to claim that he did not hold this view. In his words: "Whether an exposed subject does or does not develop a cancer is largely a matter of luck; bad luck if the several necessary changes all occur in the same stem cell when there are several thousand such cells at risk, good luck if they don't. Personally, I find that makes good sense, but many people apparently do not" [6] . The last comment from the quote may have come from the alternative assumption that risk heterogeneity accounts for departures from the power law late in life (see for example the work of [14, 15]). Thus from [13] and [6] we can add the following implicit assumptions:

7. The waiting time for the i^{th} stage is exponentially distributed,
8. Almost surely, every person develops cancer (assuming an infinite lifespan).

As the justifications for the implicit assumptions have already been provided, we will use the rest of this section to justify the explicit assumptions of the Armitage and Doll model.

3.1 Cancer is the result of several discrete changes

As stated previously, the main reason for this assumption was to explain the power law found in many populations around the world. However, the multi-stage theory of carcinogenesis was not the only theory put forth to account for this phenomenon. A competing theory stated that cancer is the result of a population of seven cells each of which would be activated by a single mutation. In essence, so the theory goes, seven cells would create a "critical mass" where cancer could thrive in the body. An idea first advanced by Fisher and Hollomon separately from Dahlberg's earlier work [4, 3]. While this theory definitely could explain the power law, it failed to explain a different important oncological observation: Long periods of dormancy between the application of a carcinogen and a clinically visible tumor. One could thus argue that the true intent of the Armitage and Doll paper was to see if the multiple mutation theory could explain a wider range of characteristic cancer behaviours.

3.2 Changes are irreversible

While there is some reason to believe that this assumption was at least partially biologically motivated (perhaps by relative risk reports that Armitage had compiled linking lung cancer with the early exposure to smoking and earlier work done by Muller on radiation and its effects on cancer risk), it seems at least partially that this was a mathematical simplification used to make the calculations more tractable. Later in life, Doll, who at the time was head of the Statistics department at Oxford University said that "I can see that an awful lot of people who would be able to understand the purposes of statistical methods, and would be able to use them sensibly, are wasting their time if they really try to go deeply into the maths. In my teaching, I tried to get across purposes, methods, philosophies, and reduce the maths to a minimum" [16]. In fact as the following theorem shows, even with reversible changes, one still obtains a power law.

Theorem 3.1. *Assume that we have a finite state time-homogeneous Markov chain with states $X_1, X_2, \ldots, X_n$ where the exponential rate of movement from X_i to X_{i+1} is given by λ, the rate from X_2 to X_1 is also λ and rate for any other situation is 0. Further assume that we begin at state X_1 at time $t = 0$, then the probability density function of being absorbed by state X_n follows a power law of degree $n - 1$.*

Proof. Under the assumptions of the above theorem, we see that if Y_T is our state at time $t = T$, then define the function

$$\begin{aligned} F(T) &= Pr(Y_T = X_n) \\ &= \sum_{i=0}^{\infty} \frac{1}{2^i} \left(1 - \sum_{k=0}^{n+2i-1} e^{-\lambda T} \left(\frac{\lambda^k T^k}{k!}\right)\right), \end{aligned}$$

then

$$\begin{aligned} F'(T) &= \sum_{i=0}^{\infty} \frac{1}{2^i} e^{-\lambda T} \frac{\lambda^{n+2i-1} T^{n+2i-1}}{(n+2i-1)!} \\ &= e^{-\lambda T} \sum_{i=0}^{\infty} \frac{1}{2^i} \frac{\lambda^{n+2i-1} T^{n+2i-1}}{(n+2i-1)!} \\ &= \left(1 - \lambda T + O(T^2)\right) \left(\frac{\lambda^n T^{n-1}}{(n-1)!2} + O(T^{n+1})\right) \\ &= \frac{\lambda^n T^{n-1}}{(n-1)!2} + O(T^n) \end{aligned}$$

It is worth noting that Theorem 3.1 is a very special case (as there is a single reversible transition); however, it shows that the assumption of irreversablility is unnecessary. There are more general theorems that would yield the same conclusion, such as the following, the proof of which is omitted here.

Theorem 3.2. *Assume that we have a finite state time-homogeneous Markov chain with states $X_1, X_2, \ldots, X_n$ where, if $i < n$, the wait in stage X_i has expected value $\frac{1}{\lambda_i+\alpha}$, that the probability of advancing to stage X_{i+1} is given by $\frac{\lambda_i}{\lambda_i+\alpha_i}$ and that the probability of returning to state X_1 is $\frac{\alpha}{\lambda_i+\alpha}$ and X_n is an absorbing state. Furthermore let τ be the hitting time for an individual starting at state X_1 at time $t = 0$. Then*

$$Pr(\tau < T) = pkT^n + O(T^{n+1}),$$

where

$$p = \prod_{i=1}^{n} \frac{\lambda_i}{\lambda_i + \alpha_i}, \text{ and } k = \prod_{i=1}^{n} \lambda_i.$$

3.3 Unique order of changes

The uniqueness of order was most likely not a mathematical simplification but a biologically consistent hypothesis. Laboratory experiments showed that, in mice, the order of exposure to mutagenic substances was immensely important. Also, accounting for a single order of precancerous stages, Armitage and Doll were able to explain why mutagenic exposure at different points in life would have varying effects on a population. As an example, if a young person were exposed to a mutagen that only affected the last precancerous stage, then the mutagen would have only a negligible effect on cancer development. However, someone later in life would have a much higher level of risk from the same level of exposure.

Beyond not being a mathematical simplification (in fact, more than a third of the their mathematical appendix is dedicated to dealing with the case of having exposure that affects a single stage), the assumption is not necessary to obtain a power law. In fact, even having a unique set of stages to develop cancer is not necessary to predict a power law, as the following theorem illustrates.

Theorem 3.3. *Assume that there are m possible sets of stages. We assume further that that an individual will develop cancer if every stage in any set is activated. Let us denote these sets of stages by $S_1, S_2, \ldots, S_m$ such that $|S_i \cap S_j| = 0$ whenever $i \neq j$. Furthermore, assume that the waiting time for stages are independent and that the set with minimum cardinality is S_1 with cardinality n, and that the expected waiting time for stage $s_i, j \in S_i$ is $\frac{1}{\lambda_{i,j}}$. Let τ be the hitting time for developing cancer. Then*

$$Pr(\tau < T) = aT^n + O(T^{n+1}),$$

where a is a constant depending only on the sets S_i of cardinality n.

Proof. Let S_1 and S_2 be such sets, τ_1, τ_2 be the hitting times for S_1 and S_2 respectively, then

$$\begin{aligned} Pr(\tau_{1(\text{or } 2)} < t) &= \prod_{i=1}^{|S_{1(\text{or } 2)}|} (1 - e^{\lambda_i t}) \\ &= \prod_{i=1}^{|S_{1(\text{or } 2)}|} \left(\lambda_i t + O(t^2)\right) \\ &= \left(\prod_{i=1}^{|S_{1(\text{or } 2)}|} \lambda_i \right) t^n + O(t^{n+1}). \end{aligned}$$

Thus, let $\tau = \min\{\tau_1, \tau_2\}$. Then $Pr(\tau < t) = Pr(\tau_1 < t \text{ or } \tau_2 < t)$. Since $|S_1 \cap S_2| = 0$ and the stages within each set are independent, τ_1 is independent of τ_2, so

$$\begin{aligned} Pr(\tau < t) = Pr(\tau_1 < t) + Pr(\tau_2 < t) \\ - Pr(\tau_1 < t)Pr(\tau_2 < t), \end{aligned}$$

and thus

$$\begin{aligned} Pr(\tau < t) = \left(\prod_{i=1}^{|S_1|} \lambda_{1,i} \right) t^{|S_1|} + \left(\prod_{i=1}^{|S_2|} \lambda_{2,i} \right) t^{|S_2|} \\ + o\left(t^{max\{|S_1|,|S_2|\}}\right). \end{aligned}$$

From here, a simple induction argument proves the theorem.

What the above theorem shows is that the power law shows the length of the shortest path in an Armitage and Doll type model.

Some readers may find interesting that what we have described here is a mathematical model called a *hypergraph.*

3.4 Mortality is a good indicator for cancer incidence

This assumption was largely pragmatic. Although cancer screening was still very much in its infancy, it was established as a commonly recorded cause of death in England and Wales [2]. Because of this, autopsy data was considered much more reliable, especially for individuals between the ages of 25 and 70. This methodology was used by many other epidemiologists of the period as well (see for example [17] and the references therein).

3.5 Changes occur independently

It is very likely that this was a mathematical simplification. As Armitage and Doll did not even list what exactly they meant by changes, it is impossible to ascribe a biological meaning to this assumption. However, the independence

is clearly something that Armitage and Doll considered. In their mathematical appendix they state that "there are $(r-1)!$ factorial possible orders in which [the r] changes could occur... Furthermore, any change is equally likely to occur at any instant in the interval $(0,t)$." However, this is by no means a necessary assumption to observe a power law in cancer incidence. To see this, one needs only consider the Taylor series for the stopping time in an Erlang random process [18].

4 Armitage and Doll Models of Carcinogenesis

Many different models attributed to Armitage and Doll appear throughout the mathematical literature. Given all of the assumptions in the previous section, it seems like there should be a "standard" Armitage and Doll model. However, as the following theorem shows this is actually not possible.

Theorem 4.1. *There is no stochastic process that satisfies all the explicit and implicit assumptions of the Armitage and Doll model of carcinogenesis with two or more stages.*

Proof. Assume that n stages must occur in order to develop cancer. Call these stages $X_1, X_2, \ldots, X_n$ where the random variable $X_i \in \{0,1\}$ is 0 if the change at stage i has not occurred and 1 if it has, thus at time $T=0$ all X_i are 0. Furthermore, assume that the waiting time for each X_i to go from 0 to 1 occurs almost surly in finite time. Then define a new random variable $C = (y_1, y_2, \ldots, y_n)$ where y_i is the index of the i^{th} change that took place. In order to develop cancer then, almost surely, $C = (1,2,\ldots,n-1,n)$. However, since changes can occur independently then with positive probability, C can take on the value $(2,1,\ldots,n)$ which is a contradiction. ∎

The intuition behind the formal proof is this: one needs only consider two of the assumptions which together form a contradiction. If our stochastic system has independent stages and a unique path to cancer then with a positive probability some part of the population will be stochastically protected from cancer. If this protection exists for some subset of the population, it can not also be the case that almost surely everyone gets cancer.

Since the full set of assumptions for the Armitage-Doll model is untenable, it is not surprising that many researchers have chosen to work with subsets of these assumptions. In this section, we will explore three common models attributed to [1], we will show which assumptions these models chose to embrace and discard and we will look at asymptotic properties of these three models.

4.1 Weibull random process

The Weibull distribution is ubiquitous in reliability engineering. This distribution falls within the much broader category of time-inhomogenous exponential distributions. While originally described in the late 1920s by M. Frechet [19], it is named for the Swedish mathematician Waloddi Weibull, who derived it independently in 1951 [20]. This distribution is defined as the unique distribution that generates an exact power law as its hazard function. The hazard function of a random variable, say T, is the limiting function

$$\lambda(t) = \lim_{\delta \to 0} \frac{Pr\left(t \leq T + \delta \mid T > t\right)}{\delta},$$

which represents the instantaneous potential per unit time of an event occurring, given survival up to time t. In particular,

$$\lambda(t) = \frac{f(t)}{1 - F(T)}$$

if T is a continuous random variable [21].

Here, $F(t)$ is a cumulative distribution function for a continuous random variable with probability density of $f(t)$. We therefore define the Weibull distribution as the unique solution to the equation

$$\lambda(t) = at^{n-1} = \frac{f(t)}{1 - F(t)}.$$

By solving this separable differential equation, one obtains the following theorem.

Theorem 4.2. *The cumulative distribution function for a Weibull random process has the form*

$$F(t) = 1 - e^{-\frac{at^n}{n}}.$$

Given that the defining epidemiological behaviour which Nordling and Armitage and Doll wished to explain is the presence of a power law, it seems clear why this distribution would be a clear front-runner to represent the Armitage and Doll model. Additionally this distribution satisfies the following property:

Theorem 4.3. *If $X_t \in \{0,1\}$ is a Weibull random process such that $X_0 = 0$ then almost surely there exists a $t = T$ such that $X_T = 1$ and for all $\tau > T$ $X_\tau = 1$.*

Thus if this distribution is chosen, we see that we obtain a power law for the duration of human life (and in fact eternally), almost surely every person would develop cancer, and a change would be irreversible. However, a Weibull distribution is really just a single event with a time inhomogenous rate, thus all the other assumptions of the Armitage and Doll model are discarded. It should be noted that while some authors treat this as the exact model, others (including experts on the multistage model) treat it as an approximation that is easy to work with [22].

4.2 Max of exponential random variables

Given exponential random variables $T_1, T_2, \ldots, T_n$ where the expected value of T_i is $\frac{1}{\lambda_i}$ Define the random variable $\tau = \max\{T_1, T_2, \ldots, T_n\}$, then we say that τ follows a max independent exponential distribution. The following Theorem describes important properties of a max independent exponential distribution. The proof is contained in the mathematical appendix of this paper.

Theorem 4.4. *If τ follows a max independent exponential distribution, then the following equality holds:*

$$Pr(\tau < t) = \prod_{i=1}^{n} 1 - e^{-\lambda_i}.$$

Furthermore, if we fix an order, then the probability that all the events occur by time t and that they occur in the correct order is

$$Pr(\tau < t) = \frac{1}{n!} \prod_{i=1}^{n} 1 - e^{-\lambda_i}.$$

Remark. It is important to highlight some unique properties of this ordered max-exponential distribution. Notice that some cells will be "fortunate" enough to have their mutations occur in an order which precludes the development of the cancer in the cell. Since each individual has a finite number of cells, each of which has a nonzero probability of never developing cancer, this means that some individuals will simply be immune to cancer through random chance.

If we assume that there are n stages and m cells in a person's body, the probability a person never develops cancer is

$$Pr\left(\text{cancer is never developed}\right) = \left(1 - \frac{1}{n!}\right)^m.$$

Hence, if τ follows a max-exponential distribution with a fixed order, then it is an *improper* random variable, i.e. $\lim_{t\to\infty} Pr(\tau \leq t) < 1$ (see e.g. [23]).

Given the exact derivation in the mathematical appendix of the Armitage and Doll paper, it seems likely that this was the original model that they envisioned. In fact the proof of this theorem found in our mathematical appendix is nearly identical to the one found in [1].

Given that this is the case, it is very tempting to count this model as "the" correct multistage model. However, it is impossible to reconcile some of the asymptotic properties of this model with the later papers of Doll. In particular, as will be shown later in the present paper, as time goes to infinity, the asymptotic risk goes to 0. This is because, under this model some (albeit a very, very small percentage) of an infinite population become stochastically immune to cancer.

Perhaps another reason why this model is not used as frequently is that calculating exact relative risk is not as easy as with the next model that we present.

4.3 Sum of exponential random variables

Given exponential random variables $T_1, T_2, \ldots, T_n$ where the expected value of T_i is $\frac{1}{\lambda_i}$ Define the random variable $\tau = \sum_{i=1}^{n} T_i$ then we say that τ follows a sum of independent exponential random variables. There is a beautiful interpretation to this distribution that coincides nicely with the multistage theory of carcinogenesis: Assume a pure birth process on the non-negative integers starting at 0 with transition rate from i to $i+1$ given by λ_i if $i < n$ and 0 if $i \geq n$. Then the time to reaching state n will be the sum of the time in each state from 0 to $n-1$ the waiting time at each stage will be independent and exponentially distributed (See Figure X).

In the case when $\lambda = \lambda_i = \lambda_j$ for all $i, j < n$ we call such a distribution an Erlang distribution. This distribution is particularly easy to work with because of its very simple formula. The derivation of this distribution can be proven using a stopping time on a Poisson point process with intensity λ. For completeness we list it now in the following theorem.

Theorem 4.5. *Let $X_t \in \{0, 1\}$ be an Erlang random process with $X_0 = 0$, then*

$$Pr(X_t = 1) = 1 - \sum_{i=0}^{n-1} \frac{(\lambda t)^i e^{-\lambda t}}{i!}.$$

For this model, all people develop cancer almost surely; however, we no longer have independence of stages as we cannot go to stage three if we have not arrived yet at stage two.

4.4 Asymptotic properties of the three models

Up to this point, there has been little concern expressed in the literature about the "correct" model. This is likely due to the fact that for small time intervals, all three models we have discussed are compatible with the same power law (possibly with appropriate substitution of constants). Therefore, all three models work equally well in providing useful descriptive statistics. However, as cancer surveillance and screening has improved, there has been a push to examine cancer in the elderly. This question was left open by Armitage and Doll who stated that "whether [the power law] persists in old age is conjecture" [1]. Thus the asymptotic properties for large values of time of the incidence (or hazard function) of these models may be an important distinguishing point as pointed out in [2]. The

following theorem describes the limiting behaviour of all three models.

Theorem 4.6. *Let $h_w(t)$ be the hazard function for a Weibull random process, let $h_m(t)$ be the hazard function for a max-exponential process which must occur in a specific order, and let $h_s(t)$ be the hazard function for the sum of independent exponential random variables with the same rate parameter (that is, the hazard function of an Erlang random process). Then the following hold:*

$$\lim_{t\to\infty} h_w(t) = \infty,$$
$$\lim_{t\to\infty} h_m(t) = 0,$$
$$0 < \lim_{t\to\infty} h_s(t) = k < \infty.$$

It is worth noting that even with though the large asymptotic behaviour of these models is qualitatively very different, none of them provide a good fit for existing data on cancer in the elderly (see, for example [2, 24] among others).

Armitage and Doll's work is important not only for the history of mathematical oncology, but also for the wider history of biomathematics. In this paper we have examined the logical and historical context of the model. Our hope is that examinations of this type will be of interest to both historians of science and modern day modelers.

5 Conclusion

The landmark 1954 paper by Armitage and Doll can be seen as the start of mathematical epidemiology of cancer specifically and noncommunicable diseases in general. To this day, their pioneering approach is still the backbone of mathematical oncology. In this paper, we have recounted the historical and biological context of their mathematical assumptions in hopes that it might shed some light on the model itself. From this type of project we see, not only that mathematics can help shed light on biologically motivated patterns in data, but also that careful mathematical derivation can help us create all new biological hypotheses. Hypotheses such as: could there be multiple sets of stages that could lead to cancer? Or, is it possible to reverse a precarcinogenic stage? In this case, mathematics and biology make for compelling mates.

Author Contributions

Authors are listed in alphabetical order. All authors contributed equally in this manuscript.

Mathematical Appendix

In this appendix we derive certain theorems from the main paper.

Proof of Theorem 3.3. Let $Pr_i(\tau < T)$ be the probability that every stage in set i has activated by time T. Then we find that

$$Pr_i(\tau < T) = \prod_{j=1}^{|S_i|}(1 - e^{\lambda_j}).$$

Thus if there is a unique set of stages the theorem clearly holds. If there are m sets of stages, we find that activation of between stages are independent events. Thus

$$\begin{aligned}
Pr(\tau < T) &= 1 - \prod_{i=1}^{m}(1 - Pr_i(\tau < T)) \\
&= 1 - \prod_{i=1}^{m}\left(1 - \prod_{j=1}^{|S_i|}(1 - e^{\lambda_j})\right) \\
&= \prod_{i=1}^{m}\left(1 - \left(\prod_{j=1}^{|S_i|}\lambda_i\right) t^{|S_i|} + o(t^{|S_i|+1})\right) \\
&= \sum_{j=\min_{i\le m}\{|S_i|\}}^{\max_{i\le m}\{S_i\}}\left(\sum_{S_i,|S_i|=n}\prod_{j=1}^{|S_i|}\lambda_i\right)t^n + o(t^{n+1}) \\
&= \left(\sum_{S_i,|S_i|=\min_{i\le m}|S_i|}\prod_{j=1}^{|S_i|}\lambda_i\right)t^n + o(t^{n+1}).
\end{aligned}$$

Proof of Theorem 4.4. Since there are n independent exponential random variables then the joint cumulative probability function is the product of the individual probability functions. Thus we see that $Pr(\tau < T) = \prod_{i=1}^{n}(1-e^{-\lambda_i t})$. However there are $n!$ possible orders in which changes can occur thus fixing an order gives us that the probability that all these changes occur by time T is order is given by $Pr(\tau < T) = \frac{1}{n!}\prod_{i=1}^{n}(1 - e^{-\lambda_i t})$.

Proof of Theorem 4.5. We view the Erlang random process as a stopping time on a Poisson point process $P_t^{\lambda} \in \{0, 1, 2, 3, ...\}$ where we define a Random variable $\tau = \min\{T|P_T \ge n\}$. Then we note that $X_t = 0$ if $t < \tau$ and $X_t = 1$ if $t \ge \tau$. Note then that

$$\begin{aligned}
Pr(X_t = 1) &= Pr(\tau < t) \\
&= Pr(P_t \ge n) \\
&= 1 - Pr(P_t < n) \\
&= 1 - \sum_{i=0}^{n-1}\frac{(\lambda t)^i e^{-\lambda t}}{i!}.
\end{aligned}$$

Proof of Theorem 4.6. Let $h_w(t)$ be the hazard function for a Weibull random process, let $h_m(t)$ be the hazard function for a max-exponential process which must occur in a specific order, and let $h_s(t)$ be the hazard function for an Erlang random process. Then, we have the following:

- Of course $h_w(t) = at^{n-1}$ thus proving the results.
- We now show that $\lim_{t\to\infty} h_m(t) = 0$. Recall that the ordered max-exponential distribution model of cancer follows an *improper* probability distribution (see the Remark in Section 4.2). From this, it follows that the survivor function approaches a positive number for large values of time. Meanwhile, the probability density function in the numerator of the hazard function still approaches 0, so the hazard function approaches 0 as well.
- Perhaps the easiest explanation for the fact that $0 < \lim_{t\to\infty} h_s(t) = k < \infty$ is that if we are waiting to make n steps and we have not yet made them after a very large amount of time then we are probably waiting to take the last step. Thus our function asymptotically approaches the risk function for an exponential random variable.

  Calling the parameters of the Erlang random process λ (the rate), and k (the shape), the hazard function is given by

$$h_s(t) = \frac{f_s(t)}{S_s(t)} = \frac{\frac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!}}{\sum_{n=0}^{k-1} \frac{e^{-\lambda x}(\lambda x)^n}{n!}} = \lambda r,$$

  where

$$r = \frac{\frac{1}{(k-1)!}(\lambda x)^{k-1}}{\sum_{n=0}^{k-1} \frac{1}{n!}(\lambda x)^n}.$$

  Then, r is a ratio of polynomials of x which are both of order $k-1$, and thus

$$\lim_{t\to\infty} h_s(t) = \lambda \lim_{t\to\infty} r = c,$$

  a constant. Since the rate parameter λ of an Erlang process is necessarily greater than zero,

$$0 < \lim_{t\to\infty} h_s(t) = c < \infty.$$

References

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

[2] Hiller, J., Vallejo, C., Betthauser, L., and Keesling, J. (2017). Characteristic patterns of cancer incidence: epidemiological data, biological theories, and multistage models. *Progress in Biophysics and Molecular Biology*, 124:41–48.

[3] Nording, C. (1951). A new theory on the Cancer-inducing Mechanism. *Brit. J. Cancer*, 7(1):68–72.

[4] Fisher, J. and Hollomon, J. (1951). A hypothesis for the origin of cancer foci. *Cancer*, 4(5):916–918.

[5] Nording, C. (1952). Cancerteorier och cancerstatistik. *Nord Med.*, 47(24):917–820.

[6] Doll, R. (2004). Commentary: The age distribution of cancer and a multistage theory of carcinogenesis. *International Journal of Epidemiology*, 33:1183–1184.

[7] Fisher, C. J. (1958). Multiple-mutation theory of carcinogenesis. *Nature*, 181:651–652.

[8] Mueller, H. J. (1927). Artificial transmutation of the gene. *Science*, 66:84–87.

[9] McCombs, R. S. and McCombs, R. P. (1930). A hypothesis on the causation of cancer. *Science*, 72:423–424.

[10] Weiss, R. (2004). Multistage carcinogenesis. *Brit. J. Cancer*, 91:1981–1982.

[11] Berenblum, I. and Shubik, P. (1947). A new, quantitative, approach to the study of the stages of chemical carcinogenesis in the mouse's skin. *Brit. J. Cancer*, 1:383–391.

[12] Stocks, P. (1953). A study of the age curve for cancer of the stomach in connection with a theory of the cancer producing mechanism. *Brit. J. Cancer*, 7:407–417.

[13] Armitage, P. (1953). A note on the time-homogeneous birth process. *Brit. J. Cancer*, 15:90–91.

[14] Gsteiger, S. and Morgenthaler, S. (2008). Heterogeneity in multistage carcinogenesis and mixture modeling. *Theoretical Biology and Medical Modeling*, 5:1–12.

[15] Mdzinarishvili, T. and Sherman, S. (2014). Heuristic modeling of carcinogenesis for the population with dichotomous susceptibility to cancer: a pancreatic cancer example. *PLoS ONE*, 9(6), doi: 10.1371/journal.pone.0100087.

[16] Armitage, P. (2004). Purposes, methods, philosophies. *Significance*, 1:170–173. doi: 10.1111/j.1740-9713.2004.00064.x.

[17] Iversen, S. (1954). Human cancer and age. *Brit. J. Cancer*, 8:575–583.

[18] Hiller, J. and Keesling, J. (2018). Asymptotic relative risk results from a simplified Armitage and Doll model of carcinogenesis. *Bull. Math Bio.*, 80:670–686.

[19] Frechet, M. (1927). Sur la loi de probabilité de l'écart maximum. *Annales de la Société Polonaise de Mathématique, Cracovie*, 6:93–116.

[20] Frechet, M. (1951). A statistical distribution function of wide applicability. *Journal of Applied Mechanics*, 18:293–297.

[21] Kleinbaum, D. G. and Klein, M. (1996). *Survival analysis.* Springer.

[22] Rittera, G., Wilson R., Pompeia, F., and Burmistrova, D. (2003). The multistage model of cancer development: some implications. *Toxicology and Industrial Health*, 19:125–145.

[23] Tsodikov, A. D., Ibrahim, J. G., and Yakovlev, A. Y. (2003). Estimating cure rates from survival data: an alternative to two-component mixture models. *J. Am. Stat. Assoc.*, 98:1063–1078.

[24] Tsodikov, A. D., Ibrahim, J. G., and Yakovlev, A. Y. (2009). A generalized beta model for the age distribution of cancers: application to pancreatic and kidney cancer. *Cancer Inform.*, doi: 10.4137/CIN.S3050.