

The Bad Mathematics of the Bad Luck Theory

Mariia Beliaeva¹

Abstract

The mathematics of the Bad Luck theory of carcinogenesis by Tomasetti and Vogelstein generated a great deal of controversy among cancer specialists but did not draw the mathematicians' attention. Thus the gross mathematical mistakes of the theory foundation did not get a proper critique and remained unnoticed. As a result, the sensational quantitative estimates of the role of Bad Luck in cancer occurrence, though being erroneous, have spread widely among researchers and the general public and got the unfair popularity. The present paper reviews the actual mathematical mistakes of Bad Luck theory.

1 Introduction

Almost five years have passed since Tomasetti and Vogelstein (hereafter T&V) attempted to explain the variation of cancer risk by the number of stem cell divisions [22]. Their idea, referred to as Bad Luck Theory, further elaborated in [25, 26, 27], generated a great deal of controversy. The discussion is still alive [17, 31, 35, 41], for more bibliography see [19] and [37].

T&V argue that intrinsic stochastic effects play the main role in tumor initiation, and specifically that the majority of cancers (22 of 31 considered) “appear to be mainly due to stochastic effects associated with DNA replication of the tissues' stem cells” [22]. In their opinion, it follows that the *primary prevention* measures—vaccination, altering lifestyle, environmental control—are not likely to be as effective as the *secondary prevention* that is early detection and treatment. The authors believe that their “results could have important public health implication” and conclude that “secondary prevention should be the major focus” [22]. If this result had been obtained in a correct way, it would be really significant as making it possible to optimize the research direction and funding.

Unfortunately for the authors and fortunately for all of us they were wrong. T&V claimed that their theory was based on a mathematical model—one of many cancer mathematical models which “are popping up everywhere in cancer research” [15]. The gross mathematical mistakes however make the Bad Luck model erroneous and its conclusions unfounded. Surprisingly, these mistakes were neither detected at the peer-review stage nor noticed after publication, maybe because professional mathematicians rarely read medical papers.

On the other hand, the statement that some conclusion is based on a mathematical model usually produces a magical effect on the general public. There are a great many

¹RHYTHM Engineering, Sofia, Bulgaria; maria.beljaeva29@gmail.com

ways though for a mathematical model to be incorrect - wrong assumptions, erroneous formulas, senseless criteria, or arithmetic mistakes. Thus each model has to pass a thorough examination until one can decide to trust it. Nevertheless some cancer specialists ([1, 6, 14, 21]) just believe that the Bad Luck theory is right and use it as a basis for the further research. The purpose of this paper is to stop that common misconception by demonstrating the crucial mathematical mistakes in the theory.

The rest of the paper is organised as follows. The three main mathematical mistakes of Bad Luck theory are described in Sections 2–4, Section 5 draws the reader’s attention to the unreasonable using of one fashionable method in the theory, in Section 6 we explain what is wrong with some reasoning of the theory fallacy, Section 7 contains the conclusions.

2 The First Mistake: Correlation between Logarithms treated as Correlation between their Arguments

T&V tried to calculate the coefficient R of correlation between the human lifetime number $lscd$ of stem cell divisions within different tissues and the lifetime cancer risk r for them [22, 23]. Having obtained $R = 0.804$, they argue that as the coefficient of determination $R^2 \approx 2/3$, then approximately $2/3$ of differences in cancer risk can be explained by the number of stem cells random divisions i.e., by bad luck. T&V concluded that “The stochastic effects of DNA replication appear to be the major contributor to cancer in humans” [22]. Actually, what T&V have really calculated was not the correlation coefficient between $lscd$ and r themselves but that between their logarithms. These two coefficients are not equal.

To know it immediately one only has to google *logarithms change correlation* and find out questions like “I don’t understand why researchers sometimes use logged versions of their variables and why correlation seems so much higher if you do so” [42] or “Before logs the correlation is 0.49 and after logs it is 0.9. I thought the logs only change the scale. How is this possible?” [43]. I failed to find any mathematical explanation why logarithms change correlation in the literature. The proof below fills the gap.

2.1 A Piece of Theory

The well-known formula for the variance D_y of a non-random function $y = \varphi(x_1, x_2, \dots, x_n)$ is [12, 39]

$$D_y \approx \sum_{i=1}^n \left(\frac{\partial \varphi}{\partial x_i} \right)_m^2 D_i + 2 \sum_{i < j} \left(\frac{\partial \varphi}{\partial x_i} \right)_m \left(\frac{\partial \varphi}{\partial x_j} \right)_m K_{ij}, \quad (2.1)$$

where x_1, x_2, \dots, x_n are random variables with means m_1, m_2, \dots, m_n and variances D_1, D_2, \dots, D_n correspondingly,

$$\left(\frac{\partial \varphi}{\partial x_i} \right)_m = \varphi'_{x_i}(m_1, m_2, \dots, m_n), \quad (2.2)$$

index m means that the derivatives are taken at the means of the argumens. Note that (2.1) is [12, 32]

- exact when function φ is linear,
- approximate when φ is non-linear but the range of the values of its arguments is small enough to make possible the linear approximation of φ ,
- unreliable otherwise.

Equation 2.1 was derived by means of the *error propagation method* [32]. Below we use the same method in order to get the approximate formula for the correlation coefficient between two nonlinear functions of random arguments and to show, as a consequence, that correlation between logarithms is not equal to that between the arguments. Let y_1 and y_2 be two functions of n and k random variables correspondingly:

$$\begin{aligned} y_1 &= \varphi_1(x_1, x_2, \dots, x_n), \\ y_2 &= \varphi_2(z_1, z_2, \dots, z_k). \end{aligned} \quad (2.3)$$

Their first-order Taylor approximations about the mean points (m_1, m_2, \dots, m_n) and $(\mu_1, \mu_2, \dots, \mu_k)$ of arguments are

$$\begin{aligned} \varphi_1(x_1, x_2, \dots, x_n) &\cong A_1 + \sum_{i=1}^n B_{1,i} x_i, \\ \varphi_2(z_1, z_2, \dots, z_k) &\cong A_2 + \sum_{j=1}^k B_{2,j} z_j, \end{aligned} \quad (2.4)$$

where

$$\begin{aligned} A_1 &= \varphi_1(m_1, m_2, \dots, m_n) - \sum_{i=1}^n \left(\frac{\partial \varphi_1}{\partial x_i} \right)_m m_i, \\ A_2 &= \varphi_2(\mu_1, \mu_2, \dots, \mu_k) - \sum_{j=1}^k \left(\frac{\partial \varphi_2}{\partial z_j} \right)_\mu \mu_j, \\ B_{1,i} &= \left(\frac{\partial \varphi_1}{\partial x_i} \right)_m, \quad i = 1, \dots, n, \\ B_{2,j} &= \left(\frac{\partial \varphi_2}{\partial z_j} \right)_\mu, \quad j = 1, \dots, k. \end{aligned} \quad (2.5)$$

The covariance between y_1 and y_2 is

$$K(y_1, y_2) = M(y_1 y_2) - M(y_1) M(y_2). \quad (2.6)$$

Using (2.4) we get

$$\begin{aligned} K(y_1, y_2) &\cong M \left[\left(A_1 + \sum_{i=1}^n B_{1,i} x_i \right) \left(A_2 + \sum_{j=1}^k B_{2,j} z_j \right) \right] - \\ &M \left(A_1 + \sum_{i=1}^n B_{1,i} x_i \right) M \left(A_2 + \sum_{j=1}^k B_{2,j} z_j \right). \end{aligned} \quad (2.7)$$

After opening parenthesis and substituting $A_1, A_2, B_{1,i}, B_{2,i}$ from (2.5) into (2.7) we obtain

$$K(y_1, y_2) \cong \sum_{\substack{i=1, \dots, n \\ j=1, \dots, k}} \left(\frac{\partial \varphi_1}{\partial x_i} \right)_m \left(\frac{\partial \varphi_2}{\partial z_j} \right)_\mu \left[M(x_i z_j) - M(x_i) M(z_j) \right], \quad (2.8)$$

which is equivalent to

$$K(y_1, y_2) \cong \sum_{\substack{i=1, \dots, n \\ j=1, \dots, k}} \left(\frac{\partial \varphi_1}{\partial x_i} \right)_m \left(\frac{\partial \varphi_2}{\partial z_j} \right)_\mu K(x_i, z_j), \quad (2.9)$$

where $K(x_i, z_i)$ is the covariance between x_i and z_i .

As well as (2.1) the formula (2.9) is

- exact when equalities (2.4) are exact that is when functions φ_1, φ_2 are linear,
- approximate when at least one of those functions is non-linear but the range of the values of its arguments is small enough to make possible the linear approximation 2.4,
- unreliable otherwise.

Finally, let us consider the particular case $n = k = 1$. Suppose φ_1 and φ_2 are univariate functions:

$$\begin{aligned} y_1 &= \varphi_1(x), \\ y_2 &= \varphi_2(z). \end{aligned} \quad (2.10)$$

Now (2.9) takes the form

$$K(y_1, y_2) \cong \left(\frac{\partial \varphi_1}{\partial x} \right)_m \left(\frac{\partial \varphi_2}{\partial z} \right)_\mu K(x, z). \quad (2.11)$$

Passing to correlation coefficients we get

$$r_{y_1, y_2} \sigma_{y_1} \sigma_{y_2} \cong \left(\frac{\partial \varphi_1}{\partial x} \right)_m \left(\frac{\partial \varphi_2}{\partial z} \right)_\mu r_{x, z} \sigma_x \sigma_z, \quad (2.12)$$

where $\sigma_{y_1} = \sqrt{D_{y_1}}, \sigma_{y_2} = \sqrt{D_{y_2}}$, and variances D_{y_1}, D_{y_2} are evaluated using (2.1):

$$\begin{aligned} D_{y_1} &\cong \left(\frac{\partial \varphi_1}{\partial x} \right)_m^2 D_x \\ D_{y_2} &\cong \left(\frac{\partial \varphi_2}{\partial z} \right)_\mu^2 D_z, \end{aligned}$$

hence

$$\begin{aligned} \sigma_{y_1} &\cong \left(\frac{\partial \varphi_1}{\partial x} \right)_m \sigma_x \\ \sigma_{y_2} &\cong \left(\frac{\partial \varphi_2}{\partial z} \right)_\mu \sigma_z. \end{aligned} \quad (2.13)$$

After substituting (2.13) into (2.12) we get

$$r_{y_1, y_2} \cong r_{x, z}$$

Being a special case of (2.9), the last equation is reliable if either both functions φ_1, φ_2 are linear, or if the scatter of arguments of both of them is small enough to make their linear approximation reasonable. None of these conditions is met in the considered case: the function $y = \log_{10}x$ is not linear while the interval from 10^5 to 10^{13} in which *lscd* varies [22] cannot be considered as small.

Therefore the coefficient of correlation between the number of stem cells divisions and cancer risk cannot be calculated as that of their logarithms, hence the estimate $2/3$ of the part of cancers occurring because of bad luck, is unfounded.

Indeed, having calculated the true value of the correlation coefficient between *lscd* and *r* for the data set from [22] we get $R = 0.96$ thus $R^2 = 0.93$ which differs substantially from the value obtained by T&V. Interestingly, the authors did calculate it in [23] but the 93% fraction of unpreventable cancers seemed too high to them: “The distribution of these cancer risks is extremely skewed in the original scale . . .” That’s why they “purposely chose the more conservative estimate, based on log-log transformation of the data” and claimed that “This is standard practice in statistics” [23].

It is not. Correlation coefficient is a meaningful characteristic and one cannot use non-linear functions of the original variables instead of variables themselves in order to get its desired value. That is valid not only for correlation but for other mathematical operations as well. When planning a car trip, we calculate its duration by dividing the distance *S* by the average speed *v* and get 10 hours for $S = 600$ km and $v = 60$ km/h. Using logarithms of *S* and *v* instead results in $\log_{10}600/\log_{10}60 \approx 1.56$ but that does not mean that the trip will take less than two hours.

3 The Second Mistake: Causation treated as a Consequence of Correlations

Knowing now the true *R* value and following the logic of Bad Luck theory one can conclude that not $2/3$ but 93% of cancers should occur due to bad luck. This result however is not consistent with statistical data which show the large influence of environmental, occupational, and lifestyle factors on cancer emergence [4, 9, 18].

The paradox is resolved by means of the well-known fact that even high correlation does not imply causation. If you don’t know the nature of real interdependence, neither correlation coefficient nor coefficient of determination is important. The famous statistician Sir Ronald Fisher wrote [7]: “. . . if two factors, A and B, are associated – clearly, positively, with statistical significance, as I say – it may be that A is an important cause of B, it may be that B is an important cause if A, it may be that something else, let us say X, is an important cause of both”. Moreover, in the same article he gives an example of strong correlation, statistically significant at a high level of significance, between the variables which are obviously unrelated with each other: “. . . in the years in which a large number of apples were imported into Green Britain, there were also a large number of divorces . . . But no one, fortunately, drew the conclusion that the apples caused the divorces

or that the divorces caused the apples to be imported”. To know more examples of high correlations of that sort, known today as *spurious*, one can visit [36] or listen to Harriett Hall aka Scepdoc [11].

“The value of a correlation r_{XY} shows only the extent to which X and Y are linearly associated. It does not by itself imply that any sort of causal relationship exists between X and Y . Such a false assumption has lead to erroneous conclusions in many occasions” [5]. And of course, if R knows nothing about the cause, then why R^2 would? In general, R^2 contains less information than R since it loses the sign. Obviously, neither R nor R^2 say anything about comparative significance of two or more factors, only their regression coefficients (not to be confused with coefficient of correlation) can. For example, if one has found that $r = \alpha + 100 \cdot lscd + 2 \cdot h + \varepsilon$, where r is risk, α is a constant, h is some factor characterizing heredity, ε is noise, then one could claim that one unit increase in $lscd$ is associated with a 100 unit increase in r (holding h constant), while a one unit increase in h is associated only with a 2 unit increase in r (holding $lscd$ constant). With a suitable choice of units one could claim that $lscd$ is 50-fold more significant than heredity.

Sornette and Favre [34] used the model of populations to demonstrate the fallacy of Bad Luck theory, Weinberg and Zaykin [38] as well as and Wu et al. [40] invented biological thought experiments to demonstrate that coefficients of determination and correlation between cancer risk and cell divisions number cannot differentiate between the contributions of intrinsic and extrinsic factors. The easier way to this conclusion is just to remember one of the eternal mathematical maxims which should be well understood by everybody before attempting to use statistics: *correlation does not imply causation* or, simply put, *correlation is not causation* [7].

4 The Mysterious ERS: Calculating the Distance from a Point to a Line

The “extra risk score” (ERS) was presented in [22] as a criterion for ranging cancers according to the role of Bad Luck in their occurrence. Without saying a word about ERS physical or mathematical sense T&V defined it as

$$ERS = \log_{10}r \times \log_{10}lscd \quad (4.1)$$

When Potter and Prentice [20] pointed out the weirdness of ERS, they got even more weird explanation from its authors: “It may seem intuitive to multiply rather than add logarithms” [25], though neither Potter and Prentice nor anybody else suggested adding logarithms.

Only in the unpublished “Musings” [23] T&V revealed the geometrical idea of ERS. Besides, if a mathematician took part in peer review process, the Bad Luck theory [22] certainly would not see the light of the day until its authors included therein the decisive criterion explanation.

According to “Musings”, ERS indicates how far is the point $(\log_{10}lscd, \log_{10}r)$ from the regression line. The authors defined that as the area of the rectangle formed by the point and the coordinate axes, see Figure 1. They claim that the smaller is this area the larger is the distance from the point to the regression line hence the greater is the role

of extrinsic factors in the occurrence of the corresponding cancer. Despite the fact that again using logarithms instead of the variables themselves, while discussing the latter correlation, is a mistake itself (see Section 2), the distance between a point and a line is measured by the length of the perpendicular from the former to the latter, not by any rectangle area. For example, the area of the rectangle corresponding to cancer F (Figure 1) is less than that of cancer A. According to T&V, point F is farther from the regression line than point A. That is obviously wrong.

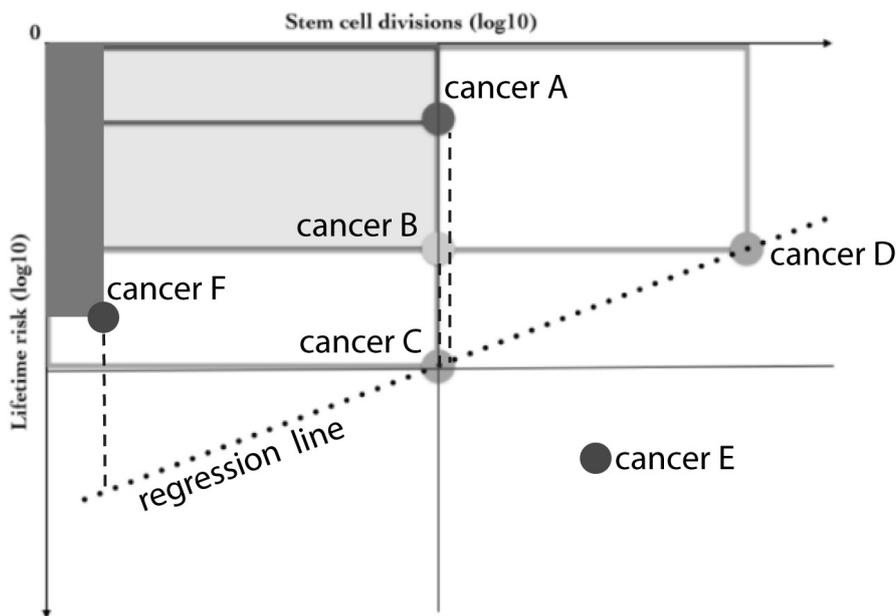


Figure 1: Ranking cancers according to ERS (Figure TR2 [23] with some additions)

To rank objects according to some characteristic I (distance from the regression line) which depends on an unknown parameter k (the slope of the line), one has to compare the average values I on the whole set of k values. T&V though simply throw away indicator I and take instead another one, J (the rectangle area) which does not correspond to the real relationship between the ranked objects.

One cannot form a working criterion by combining variables arbitrarily. One could multiply tangents of $lscd$ and r or divide their integrals by each other – it would be as meaningless as multiplying logarithms.

Consequently, dividing all cancers into two subsets with high and low ERS (rectangle areas) and state that to those with low ERS the “contribution of classic determinants (external environment and heredity) . . . is minimal” [22] does not make any sense.

The mistakes described in Sections 2–4 ruin the mathematical foundation of the Bad Luck theory completely. All of them would be detected immediately by any mathematician provided that one had examined the article [22] during the peer-review stage.

5 One funny Thing

T&V [22, 24] use no less than machine learning, specifically hierarchical clustering, in order to split a single set of thirty numbers into two parts. The method is efficient when complicated, predominantly multidimensional, sets are to be divided, or when there are a lot of sets. But to put once 30 figures in the increasing order and to divide this set into two clusters of relatively small and big values one needs no more than one's own eyes. Using machine learning in such a case is, putting it mildly, a bit redundant, and looks like an attempt to decorate medical research with fashionable mathematical terms.

6 Critiquing the Critique

6.1 The Assumption that was not made

The article by Stare [35] is the first one concerning the mathematical errors of Bad Luck theory, except for the aforementioned [20] and my own preprints [2],[3]. Stare is perfectly right that Bad Luck theory is erroneous but his reasoning is partly not correct. Specifically, he expresses the probability $p(i)$ of the particular cancer as the sum of two probabilities:

- the probability $s(i)$ of the cancer being random that is occurring due to stem cells divisions) and
- the probability $ns(i)$ of it being non-random that is occurring due to life style, environment, etc.

Such a decomposition means that the two causes of cancer are considered as incompatible random events. Thus the following assumption is made, though not formulated explicitly: *the particular cancer can occur only in one of two alternative ways - either randomly as a stochastic result of bad luck, or non-randomly (non-stochastically) - because of environment, behavior, etc; both events are considered random since this all is about probabilities. In doing so the author withdraws the very idea that cancer can occur as a result of the combination of bad luck and other factors.*

Then, Stare shows that the Bad Luck model does not give plausible results under that assumption, and claims that thus he proves the incorrectness of the model (Section 2 “A direct way of showing that R^2 cannot measure the randomness of cancer” [35]). That conclusion would be right if T&V had made that unnatural assumption. They however had not. Moreover, they argued that “Every individual cancer is the result of a possible combination of bad luck, “bad” genes, and “bad” environment”, referring to their favorite analogy of an auto accident which is a result of many factors sumaltaneously: “a long trip in a car with a drunk driver, bad tires, and poor brakes who was driving in torrential rain and hit by an out-of-control tractor-trailer” [23].

One cannot prove the fallacy of the theory by means of making assumptions which contradict the theory's foundation. Actually the conclusion that “There is no way one can claim any proportion of cancers being random, based on the analysis of Tomasetti and Vogelstein” [35] is absolutely correct, albeit for different reasons listed in the previous Sections.

6.2 One Interpretation

While criticizing the Bad Luck theory Kelly-Irving et al. [13] stated that “Even when stem-cell mutations occur as random, the initiation and development of a cancer cannot be viewed as random process”. The reason is named as the behaviour of the immune system has not been shown to be random.

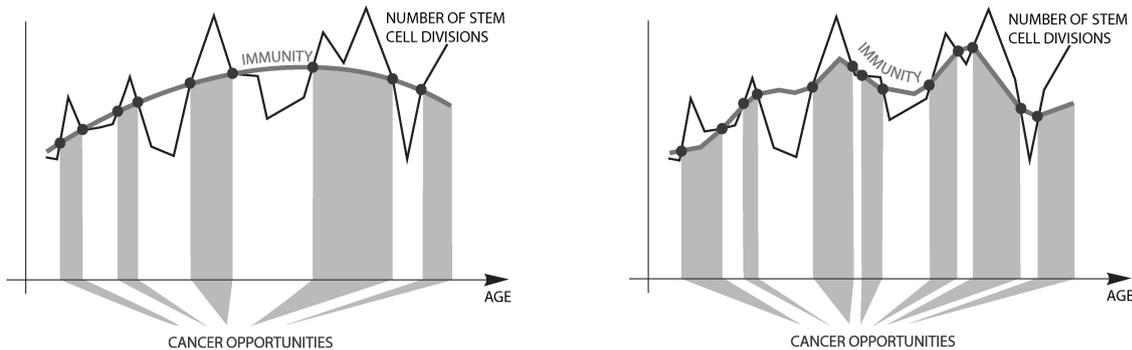


Figure 2: The intersection points (circles) of immunity (thick line) and number of stem cells divisions (thin line) are random when the immunity is nonrandom (left) as well as when it is random (right).

Figure 2 illustrates that this is wrong. If the T&V idea was correct, the number of stem cell mutations could sometimes exceed some threshold and the resistance (immunity) would fail thus giving start to cancer. If one of two processes is random and the other is not, then their intersections still form a random sequence (Figure 2a)). And of course immunity itself is a random process (Figure 2b)) as it is formed in the environment (the body) which has a lot of random characteristics and is influenced by a lot of random processes – heredity, external environment, accompanying diseases, age, the state of mood etc.

7 Conclusions

Confusing correlation between the variables with that of their non-linear functions, inferring causation just from correlation, using a senseless criterion for dividing cancers into sorts are mathematical mistakes that make the Bad Luck theory mathematically incorrect and its conclusions meaningless. That is not to say that the hypothesis about the prevailing role of bad luck in carcinogenesis is wrong. At the moment though, it is no more than mathematically unfounded assumption.

However, having been shared widely by mass-media ([16, 33] and many others) it became popular among cancer researchers and general public giving a false impression of having been mathematically proven. That can have, and maybe already has, serious consequences, on one hand allowing people to think that they don't have to worry about environmental conditions or lifestyle because a tremendous fraction of cancers are all the same unpreventable, and influencing the decision makers on the other hand. Meanwhile, in their later publications [27, 28, 26] T&V extend their erroneous method, formerly

confined to US population only, to the large set of countries, omitting now the very word *logarithm* as well as the definition of *extra risk score*. Thus those who begin their acquaintance with the Bad Luck theory from those later papers don't have a chance to recognize the fatal mistakes.

Besides, the authors themselves did a lot to make their theory as popular as possible: gave interviews, held a briefing [29], recorded a video [30]. Sorry for the pun, but this was a real *error propagation*. After getting extensive criticism, based mostly on biological reasoning, of their first Bad Luck article [22] T&V stopped announcing that early detection should be of major focus as being more effective than altering lifestyle for tumors with high ERS. Instead they stated that they were misunderstood and that their model is just explaining the differences between cancer rates in different tissues. This new interpretation certainly reduces the possible harmfulness of the Bad Luck theory making its results more of academic than of practical interest, but does not make them correct. The wrong mathematics proves nothing as *falsity implies anything* [8].

To conclude: the mathematician's assistance in developing mathematical models would help cancer specialists to avoid many mistakes and save them time and effort.

References

- [1] Belizário, J. E. (2018): Cancer risks linked to the Bad Luck hypothesis and epigenomic mutational signatures. *Epigenomes*, **2**(3), 13.
- [2] Beliaeva, M. (2016): Bad Luck theory from the mathematician's point of view. Retrieved from: <https://bit.ly/2X0Kya7>
- [3] Beliaeva, M. (2017): The Bad Luck theory mathematical mistakes for non-mathematicians. Retrieved from: <https://bit.ly/32DSkrs>
- [4] Clapp, R. W., Howe, G. K., and Jacobs, M. (2006): Environmental and occupational causes of cancer revisited. *Journal of Public Health Policy*, **27**(1), 61–76.
- [5] Draper, N. R. and Smith, H. (1998): *Applied Regression Analysis*. New York, NY: Wiley.
- [6] Euhus, D. M. (2019): Why breast cancer risk models fail: The case of lobular carcinoma in situ. *Annals Surgical Oncology*, 1–3.
- [7] Fisher, R. (1958): Cigarettes, cancer, and statistics. *The Centennial Review of Arts & Science*, **2**, 151–166.
- [8] Gensler, H. J. (2002): *Introduction to Logic*. London: Routledge Kegan Paul.
- [9] Golditz, G. A. (2016): The preventability of cancer. Stacking the deck. *JAMA Oncology*, **2**(9), 1131–1133.
- [10] Healy, M. (2017): 'Bad luck' with random DNA errors is responsible for two-thirds of cancer mutations, study says. Retrieved from <https://lat.ms/2K7ROeW>

- [11] Hall, H. (2015): Science-based medicine. Retrieved from: <https://bit.ly/2NCqEiD>
- [12] Jenkins, G. M. and Watts, D. G. (1968): *Spectral Analysis and its Applications*. San Francisco, CA: Holden-Day.
- [13] Kelly-Irving, M., Delpierre, C., and Vineis, P. (2017): Beyond bad luck: Induced mutations and hallmarks of cancer. *Lancet Oncology*, **18**(8), 999–1000.
- [14] Kokko, H. and Hochberg M.E. (2015): Towards cancer-aware life-history modelling. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **370**(1673), 20140234
- [15] Mackenzie, D. (2004): Mathematical modeling and cancer. *SIAM News*, **37**(1), 1–3.
- [16] Matthews, S. (2017): The most common cause of cancer? Simple bad luck: Random mistakes in genetics are to blame, scientists say. Retrieved from: <https://dailym.ai/36Wrjmr>
- [17] Palmer, S., Albergante, L., Blackburn, C. C., and Newman T. J. (2018): Further discussion on the immunological model of carcinogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, **115**(19), 1883–1888.
- [18] Parkin, D. M., Boyd, L., and Walker, L. C. (2011): The fraction of cancer attributable to lifestyle and environmental factors in the UK in 2010. *British Journal of Cancer*, **105**(S2), 577–581.
- [19] Perduca, V., Alexandrov, L. B., Kelly-Irving, M., Delpierre, C., Omichessan, H., Little, M. P. Vineis, P., and Severi, G. (2019): Stem cell replication, somatic mutations and role of randomness in the development of cancer. *European Journal of Epidemiology*, **34**(5), 439–445.
- [20] Potter, J. D. and Prentice, R. L. (2015): Cancer risk: Tumors excluded. *Science*, **347**(6223), 727.
- [21] Sarode, S. C., Sarode, G. S., and Patil, S. (2018): Decoding oral cancer conundrums from “bad luck” point of view. *Future Oncology*, **14**(5), 409–411.
- [22] Tomasetti, C. and Vogelstein, B. (2015): Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science*, **347**(6217), 78–81.
- [23] Tomasetti, C. and Vogelstein, B. (2015): Technical Report: Musing on the theory that variation in cancer risk among tissues can be explained by the number of divisions of the normal stem cells. Retrieved from: <http://arxiv.org/abs/1501.05035>
- [24] Tomasetti, C. and Vogelstein, B. Supplementary Materials for Variation in cancer risk among tissues can be explained by the number of stem cell divisions. Retrieved from: <https://bit.ly/2CJsNCZ>

- [25] Tomasetti, C. and Vogelstein, B. (2015) Cancer risk: role of environment-response. *Science*, **347**(6223), 729–731.
- [26] Tomasetti C., Li L., and Vogelstein, B. (2017): Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention, *Science*, **355**(6331), 1330–1334.
- [27] Tomasetti, C. and Vogelstein, B. (2015): On the slope of the regression between stem cell divisions and cancer risk, and the lack of correlation between stem cell divisions and environmental factors-associated cancer risk. *PLoS ONE*, **12**(5), e0175535.
- [28] Tomasetti, C., Durrett, R., Kimmel, M., Lambert, A., Parmigiani, G., Zauber, A., and Vogelstein, B. (2017): Role of stem-cell divisions in cancer risk. *Nature*, **548**(7666), E13–E14.
- [29] Tomasetti, C. and Vogelstein, B. (2017): Science Magazine press conference on Livestream. Retrieved from: <https://bit.ly/2CBby6o>
- [30] Tomasetti, C. and Vogelstein, B. (2017): What causes cancer? Cancer mutations and random DNA copying errors. Retrieved from: <https://bit.ly/33BPZib>
- [31] Trosko, J. E. and Carruba, G. (2017): “Bad Luck mutations”: DNA Mutations are not the whole answer to understanding cancer risk. *Dose Response*, **15**(2), 1559325817716585.
- [32] Tukey, J. W. (1958): *The propagation of errors, fluctuations and tolerances. Technical Reports 10-12*. Princeton, NJ: University of Princeton.
- [33] Scutti, S. (2017): ‘Bad luck’ mutations increase cancer risk more than behavior, study says. Retrieved from: <https://cnn.it/2NWDPK6>
- [34] Sornette D. and Favre M. (2015): Debunking mathematically the logical fallacy that cancer risk is just “bad luck”. *EPJ Nonlinear Biomedical Physics*, **3**(1), 10.
- [35] Stare, J. (2018): Bad luck of cancer – or misinterpreted statistics? *Metodološki zvezki*, **14**(2), 75–81.
- [36] Vigen, T. (2019): Spurious correlations. Retrieved from: <https://bit.ly/36VKI6Y>
- [37] Vassilev A. and DePamphilis M. L. (2017): Links between DNA replication, stem cells and cancer. *Genes*, **8**(2), 45.
- [38] Weinberg, C. R. and Zaykin, D. (2015): Is bad luck the main cause of cancer? *Journal of the National Cancer Institute*, **107**(7), 1–4.
- [39] Wentzel, E. (1962): *Theory of Probabilities*. Moscow: Fizmatgiz.
- [40] Wu, S., Powers, S., Zhu, W., and Hannan, Y. A. (2016): Substantial contribution of extrinsic risk factors to cancer development. *Nature*, **529**(7584), 43–47.

[41] Zhu, W., Wu, S., and Hannun, Y. A. (2017): Contributions of the intrinsic mutation process to cancer mutation and risk burdens. *EBioMedicine*, **24**(2017), 5–6.

[42] <https://bit.ly/2X4v98s>

[43] <https://bit.ly/2pRxeJ6>