Medical Hypotheses (2004) **x**, xxx–xxx

ELSEVIER

medical
hypotheses

http://intl.elsevierhealth.com/journals/mehy

Cancer death statistics: analogy between epidemiology and critical systems in physics

M. Filoche,[a] and L. Schwartz[b,*]

[a] *Laboratoire de Physique de la Matière Condensée, Ecole Polytechnique, 91128 Palaiseau Cedex, France*
[b] *Service de Radiothérapie, Hôpital de la Pitié-Salpêtrière, 47, Boulevard de l'Hôpital, 75013 Paris, France*

Received 28 May 2003; accepted 27 November 2003

Summary The determination of risk factors in carcinogenesis is said to be an essential step towards the understanding of this disease. Most mathematical models describing the evolution of mortality figures use the concept of death probability (or "force of mortality" or "hazard of death"). When summarizing the death statistics through this unique parameter, one implicitly makes the assumption that the death events are independent from one individual to another. In this paper, we show that this hypothesis has profound consequences as it implies a "gaussian" behavior of the death statistics fluctuations. In order to verify the validity of this assumption, French cancer death statistics between the years 1978–1996 are examined. Their fluctuations, for every age bracket, are computed and then compared to the expected gaussian fluctuations that should emerge from a model of death probability. We show that the observed fluctuations are in close agreement with a gaussian model up to 35–40 years. After 40 years, the fluctuations are much higher and cannot be explained by a model where every individual would have a given "probability of death". These observations may produce a new insight into old-age cancer mortality. It suggests that there could exist a major difference between cancers in young or older organisms: cancer developed in young organisms are the consequence of a specific attack against an organ (essentially originated from a single cause, like a virus or a genetic deficiency). On the other hand, older organism are closer to a "critical state" and, as such, the outcome of a cancer in a given organ could be the consequence of a chain of "malfunctions" (analogous to an avalanch in physical systems) in the entire organism.

Introduction

Most diseases (cancer, cardiovascular disease, dementia) are supposed to be the result of a complex interaction between the genes and the environment. In cancer, genes responsible for tumor formation have been cloned, as for example, Rb responsible for retinoblastoma, or P53 responsible for Li-Fraumeni syndrome, or B.C.R.A. responsible for early breast cancer [1].

The patients suffering from these hereditary tumors are young. So far, there is no human gene which appears to be clearly responsible for late onset carcinogenesis. Early or late onset breast cancer and melanoma does not appear to have the same risk factors [2–4]. Furthermore, we still do not know what causes most brain tumors, pancreatic carcinoma or leukaemia.

* Corresponding author.
E-mail addresses: marcel.filoche@polytechnique.fr (M. Filoche), laurent.schwartz@polytechnique.fr (L. Schwartz).

0306-9877/$ - see front matter © 2004 Published by Elsevier Ltd.
doi:10.1016/j.mehy.2003.11.021

In order to solve this paradox, we need to better understand the underlying biology, develop mechanistic hypotheses and test them in clinical trials in humans. Nevertheless, the concept of *universality*, that has emerged in the last 30 years in modern physics, can help us to extract some properties of the pathology directly from the death statistics, without going into the details of its mechanism.

Universality is a property that explains how very different systems that obey the same limited number of rules can exhibit identical macroscopic behavior, even though their microscopic components have nothing in common. For example, in physics of phase transition, a large class of systems exhibit at transition some properties known as *scale invariance*, *power law statistics* or *extreme sensitivity to external perturbations*. This is the case in very different domains, as for example in seismology (distribution of earthquakes), fracture in material sciences or "physics of the sand pile". One of the consequences of these properties is that the statistical fluctuations are much more important and more sensitive to any external perturbation than in classical systems in physics, which are at or near equilibrium.

Based on this reasoning, we will study in this paper the fluctuations of cancer death statistics and analyse them in order to determine whether they can fall in one or the other class of statistics.

From death statistics to death probability: the binomial law

It is very common to come across the expression "death probability". This number comprised between 0 and 1 is supposed to represent the chance for a given person to die in the next month or year. It is used as an indicator to compare the respective efficiencies of various types of cure, or the advantages and drawbacks of different ways of life. It is, for example, very common to read that smoking, or drinking increases the death probability by a certain percentage.

It seems clear that reducing the modelling of mortality to one parameter, simple to handle, can be of great interest for physicians. It allows direct comparisons between different populations and offers a clear picture of the situation. This simplicity can nevertheless be dangerous because this approach, although intuitive, contains implicitly a very important conceptual step: it models the death as a two-tier probability law, whose tiers are *alive* and *dead*. This law is entirely characterized by the probability p that a given individual dies during the next year.

If one considers a sample of N individuals who obey this law, and if one assumes that the death events are independent, then the probability that n of them would die during the year to come is given by

$$P(n) = C_N^n p^n (1-p)^{N-n}.$$

This law is called a *binomial law*. Its mean is equal to pN. This is in fact a practical way used to extract the probability p from the death statistics: it is the ratio between the average number of deaths a year (for a given cause and at a given age) divided by the average number of people of this age. This binomial law also possesses a standard deviation which represents the fluctuation of the law around its mean. As a consequence, the aggregation of a large number of independent events obeys a law whose standard deviations (or fluctuations) are also very precisely determined. In this paper, we will use this property to examine the validity of the "binomial" approach for death statistics.

Gaussian versus non gaussian fluctuations

By definition, death statistics are obtained by totaling the individual death events. It may seem reasonable to assume that these events are independent. Thus, in order to study the fluctuations of the death statistics, we first need to briefly recall the properties of the law of the sum of N independent events.

The main tool for analyzing such a law is called the "central limit theorem". This theorem states that, if an elementary law has a mean m and a standard deviation σ, then the sum of N independent events obeying this elementary law obeys an *aggregated* law which converges, in the limit of very large N, towards a gaussian law. A gaussian law with a mean M and a standard deviation Σ has the following form:

$$p(x) = \frac{1}{\sqrt{2\pi}\Sigma} \exp\left[-\frac{(x-M)^2}{2\Sigma^2}\right].$$

Within the central limit theorem, the values of M and Σ depend only on m, σ and N

$$M = Nm, \quad \Sigma = \sqrt{N}\sigma.$$

We have seen this previously in the case of the binomial law. We can see that the mean of the sum is proportional to the total number of tries, while the standard deviation is proportional to the square root of this number. Thus, the standard deviation

YMEHY 2097
DISK / 21/1/04
ARTICLE IN PRESS
No. of pages: 6
DTD 4.3.1 / SPS

of the sum increases as the square root of its mean. In other words, relative fluctuations (characterized by the standard deviation) decrease when the size of the sample (N) increases. This is a very important property that we will use later in this paper. Practically, this relations mean also means that in this model, once the average number of deaths is known for a given population, the fluctuation should also be precisely determined.

On the other hand, if the elementary law does not have a mean or a standard deviation (that is, if one of them is infinite), then one cannot apply the central limit theorem. These kinds of laws are characterized, among others properties, by the large amplitude of their fluctuations that do not decrease relatively when the size of the sample increases. Numerous examples of such laws can be found in modern physics, especially in the study of critical states. Thus, if one plots the distribution of earthquakes as a function of their magnitude, one finds a *power law* $P(x) \approx x^{-\alpha}$ with $\alpha \approx 2$ (Fig. 1) (5). The mean of such a law, when computed, is found to be infinite. This fact can be expressed by stating that "there is no average earthquake". One practical consequence is that any mean computed on a finite sample will be directly determined by the magnitude of the largest event in the sample. The fluctuations do not vary as the square root of the measured mean: whatever the number of events in the sample, the relative fluctuations have the same magnitude.

This type of behavior is one of the signatures of these probability laws that do not obey the central limit theorem, and therefore do not converge towards a gaussian law when aggregated. They appear especially in physics in the field of *phase transition*, in systems near a critical state. These systems are characterized by an extreme dependency on external perturbations, and are the center of events called "catastrophic", like avalanches, whose magnitudes are distributed according to a power law.

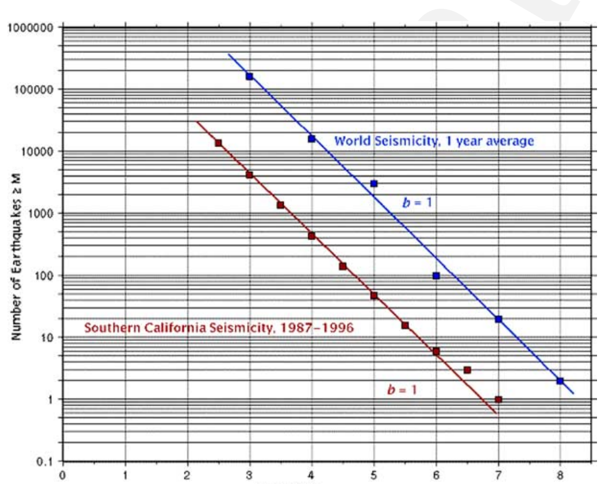

Figure 1 Distribution of earthquakes as a function of their magnitude. The red line represents data from California and the blue line data from the world (5).

In short, the analysis of the fluctuations of death statistics will allow to test the validity of the binomial approach used to describe mortality by cancer.

Analysis of French death statistics

We have gathered from INSERM (Institut National de le Santé et de la Recherche Médicale), the monthly figures of French mortality between the years 1978 and 1996, as well as the demographic data. At each death (French citizen or not), in order to get authorisation for burial, a certificate stating the primary and secondary causes of death must be filled and signed by a physician. The primary cause of death is the underlying disease, thought to be the main reason for death. For example a women dies from widely metastatic breast cancer. In the days before cardiac arrest, she was diagnosed with pneumonia. The primary cause of death is breast cancer, the secondary cause is pneumonia. These data are collected by the I.N.S.E.E. (Institut National des Statistiques et des Etudes Economiques) and transmitted to the I.N.S.E.R.M... They are freely available on the internet at http://sc8.vesinet.inserm.fr:1080/. The causes of death are filled in according to the *Classification Internationale des Maladies* (C.I.M 7 and 8) as well as sex, age at death, and calendar year of death.

The French population has been divided into men and women, and then in twenty age brackets of five years each. For each of these brackets, first the monthly death rate has been plotted as a function of time. For analysing the statistics, we have assumed two things:

1. The way of life in France during the period 1978—1996 has been constant enough so that the aggregation of the data is still valid. If anomalous fluctuations should be observed, then their magnitude will be far larger than the slight variations induced by changes in life style.
2. Secondly, it is assumed that, instead studying the same population along its aging, one can study different age brackets coexisting at the same time. Once again, the analysis of the data will confirm the validity of this assumption.

Fig. 2 presents the number of deaths for the age brackets 30—35, 60—65 and 90—95 years. One can see relative fluctuations strongly increase with age, even though the monthly number of deaths increases, for example between the green and the blue curves (one should recall that the relative fluctuation should decrease in a gaussian model). On the other hand, Fig. 3 presents the death rate deduced from the number of deaths and the population for the same age brackets.

First, we study the death rate (separately for men and women). It can be seen that the variation of this rate is very small (Fig. 3), for any age bracket, although the figures for these brackets can change drastically during the same period (Fig. 2): the blue curve (age bracket 60—65) presents a sharp increase at the beginning of the eighties, which corresponds in fact to the evolution of the birth rate during and right after the first world war. This constant value of the cancer death rate confirms our assumption that consists in studying the fluctuations during the period 1978—1996 and neglecting the possible changes in life style. If these changes exist, it is remarkable to notice that they did not heavily modify the average cancer death rate, nor annual or monthly fluctuations.

One can also notice a small periodicity on the red curve (age bracket 95—100). This periodicity corresponds to an increase of the cancer death rate in winter. One can say two things: first, after removing this fluctuation one would still have a larger fluctuation than the fluctuation predicted by a gaussian model. Second, it is also a sign that the death by cancer is in this case an event extremely sensitive to an external perturbation (pollution, temperature or light variation,...).

From these data, one can extract un average death rate for each age bracket over the whole period (Fig. 4). This rate varies greatly from one bracket to another, going from 1 to 10 000 for bracket 5—10, to 1—10 for bracket 95—100. In terms of probabilities $(p, 1-p)$, one can say that a very large range of p is explored.

The next figure presents the same type of data, but the means are computed on an annual basis instead of over the whole period 1978—1996 (Fig. 5). Indeed, for each age bracket, each cross represents one of the years. Thus, the vertical dispersion of the crosses gives an idea of the relative fluctuation. In order to represent more precisely this fluctuation, and to study its behavior, the ratio between the standard deviation extracted

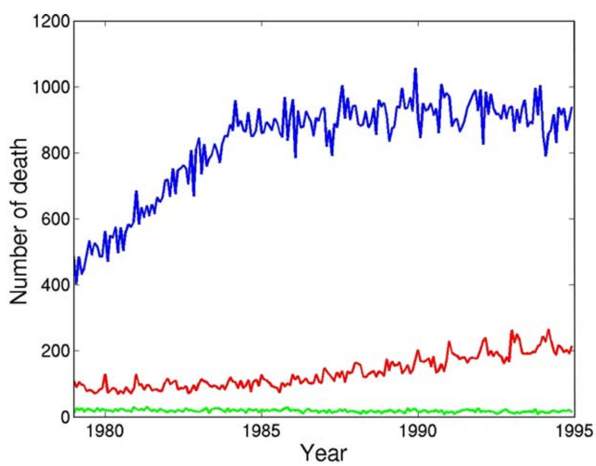

Figure 2 Monthly number of men deaths by cancer (between 1978 and 1995). Three age brackets are represented: in green, 30—35 years, in blue 60—65 and in red 90—95.

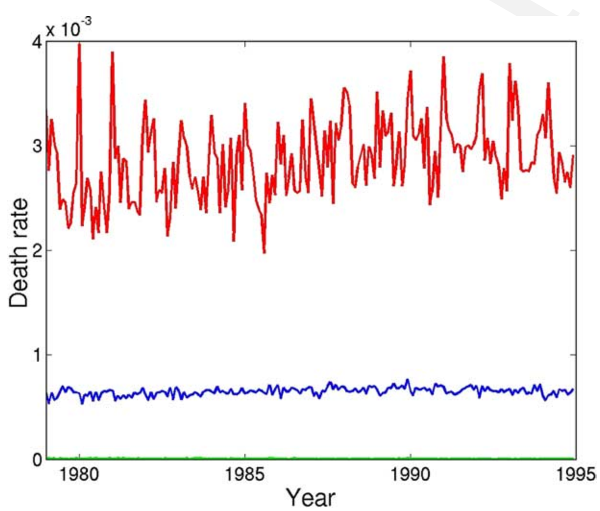

Figure 3 Monthly death rate by cancer for men (between 1978 and 1995). The age brackets are the same as in Fig. 2.

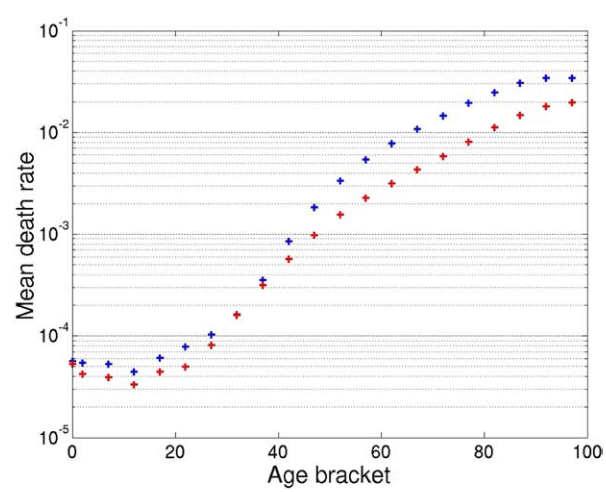

Figure 4 Average death rate for each age bracket during the period 1978—1996. The data are plotted in a semi-log. One can see the wide range of values explored for the death rate.

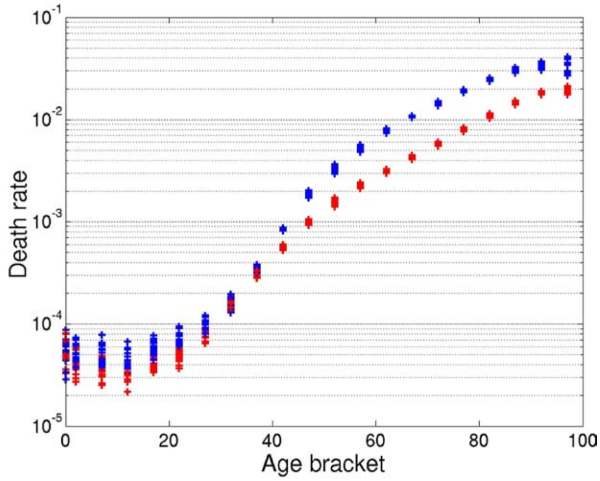

Figure 5 Annual death rate by cancer for each age bracket: each cross represents one year between 1978 and 1995. The blue crosses are the men and the red crosses the women.

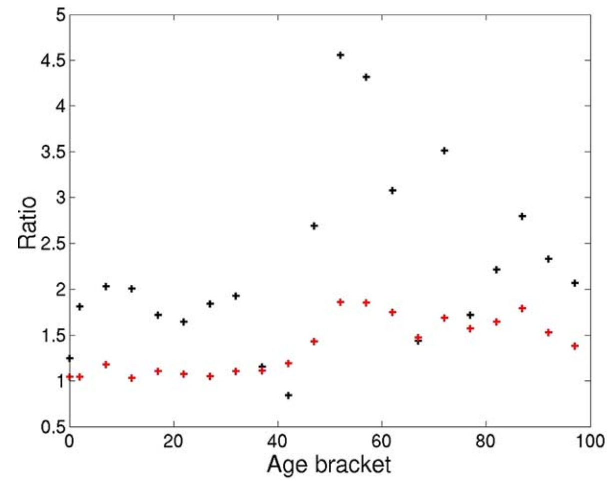

Figure 7 Ratio between the standard deviation measured from death rate by cancer and the standard deviation expected from gaussian behavior. One can see that this ratio deviates sharply from 1 after 40 years.

from the data and the standard deviation computed from the gaussian aggregration has been plotted, both for the number of deaths (Fig. 6) and the death rate (Fig. 7). One can notice immediately a striking change for men under 35 years of age. Indeed, this ratio stays very close to 1 up to 35 years, even though the death rate (or death probability) changes by a factor 50 between 5 and 35 years. Based on the study of these fluctuations, we can conclude that the binomial model $(p, 1 - p)$ is well fitted to describe cancer mortality up to 35 years.

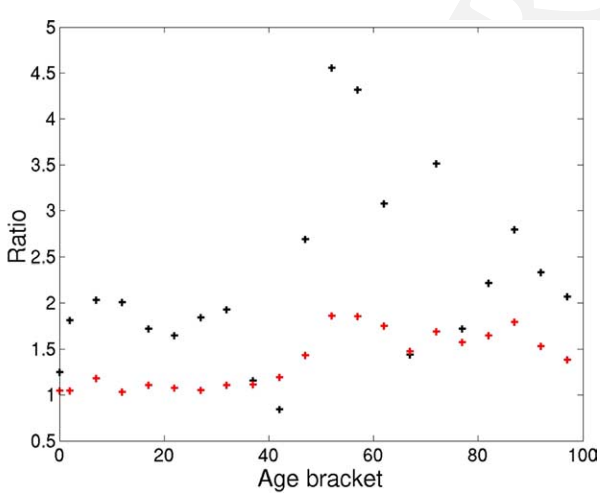

Figure 6 Ratio between the standard deviation measured from the number of deaths by cancer and the standard deviation expected from gaussian behavior. One can see that this ratio deviates sharply from 1 after 40 years.

On the other hand, from 35 years the ratio starts a dramatic increase, up to a value close to 20 (for men around 70 years). It then decreases between 70 and 100 years. This calculation gives for both men and women a kind of "bell curve", shifted 10 years later for the women. It appears then that, after 35 years, mortality by cancer cannot be summed up in one unique parameter p, as if it was a probabilistic factor based on a binomial law.

Variations of fluctuations versus the number of death

In order to analyse more precisely the variations of these fluctuations, one can plot the standard deviation extracted from the statistics versus the number of deaths, for each age bracket (Fig. 8). The lower dotted line represents the exact value of the standard deviation, as predicted by a binomial or gaussian model. The upper dotted line is simply a line of slope 1, in order to visualize what would be the slope if the variations were directly proportional to the number of deaths.

On this plot, one can see once again that the age brackets up to 35 years follow quite closely the gaussian model. After that age, the fluctuations increase and seem to be proportional to the number of deaths. Thus, if one compares the brackets 35—40 years and 90—95 years, one can see that the fluctuation is 10 times greater for the second bracket, while the number of deaths is similar for both brackets. The linear behavior of the fluctuations versus the number of events is very similar to what we can find in critical phenomena.

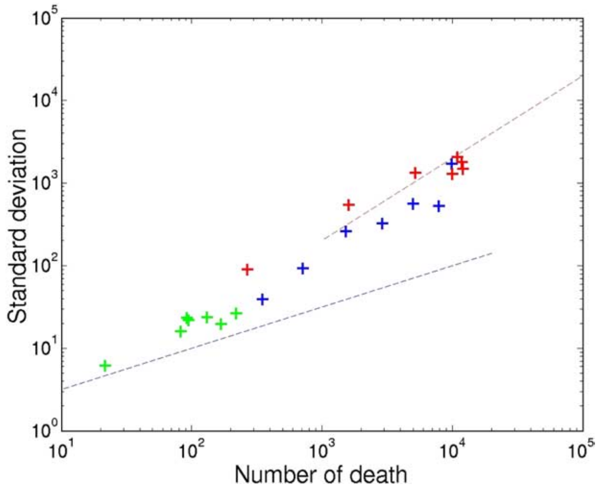

Figure 8 Log—log plot of the standard deviation versus the average annual number of deaths. Each cross represents an age bracket. The green crosses are for age brackets from 0 to 30 years, blue from 35 to 70 and red from 70 to 100. The dotted bottom line represents the theoretical standard deviation for the gaussian model.

Conclusion

The notion of death probability is inherently based on the assumption that the death of individuals are independent events that occur at a given rate. This intuitive idea has profound implications on the behaviour of the fluctuations of these events. By measuring the fluctuations (characterized in a first step by standard deviation) of death by cancer in the French population over 18 years, an anomalous behavior was exhibited for individuals older than 35 or 40 years. In fact, fluctuations of death by cancer for age brackets after 40 years are much stronger than what one should expect from a gaussian model based on independent events obeying a binomial law.

These fluctuations seem, on the other hand, increase at the same rate as the number of death events. This characteristic of large fluctuations can also be found, in modern physics, in the study of physical systems near critical states. This could be a strong indication that cancer of a young population and of an older population should be regarded as different diseases. In the former case, cancer would be more like a "classical" disease with "deterministic and simple" causes. In the latter, cancer acts on an aged organism that could be considered as a "system near a critical state". Thus, the global interaction between the disease and the organism could no longer be summarized in a single scalar quantity as the death probability, and the large fluctuations of death events observed in the statistics are a signature of this criticality. We wish to point out that this way of looking at cancer could have major consequences on the understanding of the behavior of the disease in general and also on the interpretation of therapeutic effects.

References

[1] De Vita V, Hellman S, Rosenberg S. Cancer: principles and practice of oncology. Lippincott: Philadelphia USA; 1993.
[2] Riggs JE. Biphasic Gompertzian malignant melanoma mortality rate dynamics: manifestation of genetic heterogeneity. Ann Human Biol 1995;22:515—23.
[3] Yashin AI, Vaupel JW, Iachine IA. A duality in aging: the equivalence of mortality based on radically different concepts. Mech Ageing Dev 1994;74:1—14.
[4] Mackaren G, Roche CA, Hughes K. The effectiveness of the gail model in estimating risk for development of breast cancer in women under 40 years of age. Breast J 2001;7:34—9.