Erlang distribution predicts the number of driver events for childhood and young adulthood cancers

Aleksey V. Belikov
Laboratory of Innovative Medicine and Agrobiotechnology,
School of Biological and Medical Physics,
Moscow Institute of Physics and Technology (MIPT),
Institutsky per., 9, 141701 Dolgoprudny, Moscow Region, Russia
Email: belikov.research@gmail.com

Abstract

It is assumed that cancers develop upon randomly acquiring (epi)mutations in driver genes, but their exact number for each cancer type is not known. I have recently shown that the age distribution of incidence for 20 most prevalent cancers of old age is closely approximated by the Erlang probability distribution. I then used it to predict the number of driver events for these cancer types, as it describes the probability of several successive random events occurring precisely by the given time. However, 4 other probability distributions out of 16 tested also provided acceptable fits to the incidence data, leaving some doubt about the validity of predictions. Here I show that the Erlang distribution is the only classical probability distribution that can adequately model the age distribution of incidence for all studied childhood/young adulthood cancers. This validates it as the universal law describing cancer development at any age and as a useful tool to predict the number of driver events for any cancer type.

Introduction

Since the discovery of the connection between cancer and mutations in DNA, in the middle of the XX century, there have been multiple attempts to deduce the number of driver mutations from the age distribution of cancer incidence or mortality[1]. The proposed models, however, suffer from several serious drawbacks. For example, early models assume that cancer mortality increases with age according to the power law[2-4], which at some advanced age would necessary lead to the mortality surpassing 100,000 people per 100,000 population. Moreover, already at that time it was known that many cancers display deceleration of mortality growth at advanced age, which is to be expected if the probability of death at a given age is to remain under 100%. Finally, when large-scale incidence data have accumulated, it became clear that cancer incidence not only ceases to increase with age but, for at least some cancers, starts to decrease[5,6]. More recent models of cancer progression are based on multiple biological assumptions, consist of complicated equations that incorporate many predetermined empirical parameters, and still have not been shown to describe the decrease in cancer incidence at an advanced age[7-12]. It is also clear that an infinite number of such mechanistic models can be created and custom tailored to fit any set of data, leading us to question their explanatory and predictive values.

Recently I have proposed that the age distribution of cancer incidence is, in fact, a statistical distribution of probabilities that a required number of driver events occurs precisely by the given age, i.e. a probability density function (PDF)[13]. I then tested the PDFs of 16 well-known continuous probability distributions for fits with the CDC WONDER data on the age distribution of incidence for 20 most prevalent cancers of old age. The best fits were observed for the gamma distribution and its special case – the Erlang distribution, with the average R^2 of 0.995[13]. Notably, these two distributions describe the probability of several independent random events occurring precisely by the given time. This allowed me to estimate the number of driver events, the average time interval between them and the

maximal populational susceptibility, for each cancer type. The results showed high heterogeneity in all three parameters amongst the cancer types but high reproducibility between the years of observation[13].

However, 4 other probability distributions – the extreme value, normal, logistic and Weibull – also showed good fits to the data, although inferior to the gamma and Erlang distributions. This leaves some uncertainty regarding the exceptionality of the gamma/Erlang distribution for the description of cancer incidence. Here I test these shortlisted distributions on the CDC WONDER data on childhood and young adulthood cancers. I show that the gamma and Erlang distributions are the only distributions that converge for all tested cancers and provide close fits. This result validates the gamma/Erlang distribution as the fundamental law describing the age distribution of cancer incidence for any cancer type, which also predicts important parameters of cancer development, including the number of driver events.

Results

To test the universality of the gamma/Erlang distribution, the publicly available data on the age distribution of incidence of childhood and young adulthood cancers were downloaded from the CDC WONDER database (see Methods). The PDFs for the general forms of the following continuous probability distributions were tested for fit with least squares non-weighted nonlinear regression analysis: extreme value, gamma, logistic, normal and Weibull (see Methods). Only the gamma distribution converged for all tested cancer types and provided good fits (Fig. 1, Table 1).

Importantly, the gamma distribution and the Erlang distribution derived from it are the only classical continuous probability distributions that describe the cumulative waiting time for k successive random events, with the Erlang distribution differing only in counting events as integer numbers. Because these properties suit excellently to describe the waiting time for real discrete random events such as driver mutations, the Erlang distribution provides the opportunity to get unique insights into the carcinogenesis process. I have previously proposed that the shape parameter k of the Erlang distribution indicates the average number of driver events that need to occur in order for a cancer to develop to a stage that can be detected during a clinical screening; the scale parameter b indicates the average time interval (in years) between such events; and the amplitude parameter A divided by 1000 estimates the maximal susceptibility (in percent) of a given population to a given type of cancer[13].

To obtain these parameter values, the Erlang distribution was fitted individually to incidence of each of 10 childhood/young adulthood cancer types (Fig. 2, Table 2). The goodness of fit varied from 0.9476, for extracranial and extragonadal germ cell tumors of young adulthood, to 0.9999, for extracranial and extragonadal germ cell tumors of childhood, with the average of 0.9846. The predicted number of driver events varied from 1, for extracranial and extragonadal germ cell tumors of childhood, neuroblastoma and ganglioneuroblastoma, retinoblastoma, intracranial and intraspinal embryonal tumors, and hepatoblastoma, to 9, for malignant gonadal germ cell tumors. The predicted average time between the events varied from 1 year, for extracranial and extragonadal germ cell tumors of childhood, to 14 years, for intracranial and intraspinal embryonal tumors. The predicted maximal populational susceptibility varied from 0.02%, for extracranial and extragonadal germ cell tumors of childhood, to 2%, for malignant gonadal germ cell tumors. Overall, the data confirm high heterogeneity in carcinogenesis patterns revealed in the previous study[13].

Discussion

I have previously shown that 5 probability distributions – the extreme value, gamma/Erlang, normal, logistic and Weibull – approximate the age distribution of incidence for 20 most prevalent cancers of old age[13]. The shape of those incidence distributions resembles the bell shape of the normal distribution, with some asymmetry, or at least the left

part of it. However, many cancers of childhood have a radically different shape of the incidence distribution, the shape of the exponential distribution (Fig. 2). Of the 5 shortlisted distributions, only the gamma/Erlang and Weibull distributions can assume that shape, i.e. reduce to the exponential distribution when the parameter *k* equals 1. Of the remaining 2 distributions, gamma/Erlang provides superior fit compared to Weibull. In fact, for cancers of old age, the average R^2 for the Weibull distribution is 0.9938, whereas for the gamma/Erlang distribution is 0.9954[13]. For cancers of childhood and young adulthood, the average R^2 for the Weibull distribution is 0.8781, whereas for the gamma/Erlang distribution is 0.9861 (Table 1). Moreover, the Weibull distribution failed to converge for extracranial and extragonadal germ cell tumors of childhood and for retinoblastoma, although this may be explained by too few data points. Thus, it appears that the gamma/Erlang distribution is the only classical probability distribution that fits universally well to cancers of childhood, young adulthood and old age.

I have proposed that the parameter *k* of the Erlang distribution indicates the average number of driver events that need to occur in order for a cancer to develop to a stage that can be detected during a clinical screening[13]. As mentioned above, the Erlang distribution reduces to the exponential distribution when *k* equals 1, because the exponential distribution describes the waiting time for a single random event. It would thus mean that cancers with the exponential shape of the age distribution of incidence require only a single driver event with random time of occurrence, most likely a somatic driver mutation[14] or epimutation[15]. This explains their maximal prevalence in the early childhood.

In his seminal paper[16], Alfred Knudson has proposed that hereditary retinoblastoma, a childhood cancer with the exponential age distribution of incidence, is caused by a single somatic mutation in addition to one heritable mutation. He also proposed that in the nonhereditary form of the disease, both mutations should occur in somatic cells. As hereditary form is estimated to represent about 45% of all cases[16,17], the number of driver mutations predicted from combined incidence data should be around 1.55. Interestingly, whilst the gamma distribution fits the incidence data excellently, with R^2=1.0, it predicts 1.336 driver events. This yields the estimate of the hereditary form prevalence at 66.4%. This higher value may point to the general underestimation of the hereditary component in unilateral retinoblastoma, perhaps due to limitations of routine genetic screening and the influence of genetic mosaicism[18].

In contrast to retinoblastoma, the hereditary form of neuroblastoma is estimated to comprise only 1-2% of all cases[19], hence the exponential age distribution of incidence would mean that only one somatic mutation is required. Indeed, the gamma distribution predicts 0.9616±0.0293 driver events (R^2=0.9998). Whilst the hereditary form is also nearly absent in hepatoblastoma, the gamma distribution predicts 1.444±0.211 driver events (R^2=0.9998). This could be interpreted in a way that 55.6% of the patients develop this cancer upon acquiring one somatic mutation, whereas the remaining 44.4% require 2 somatic mutations. Interestingly, it was shown that hepatoblastoma consists of two subtypes based on gene expression, with both subtypes presenting with a mutation in beta-catenin, and the minor subtype (36% of patients) exhibiting additional amplification of Myc[20]. An alternative explanation could be that one mutation is required for hepatoblastomas with epithelial histology (56% of patients[21]), whereas two mutations are required for hepatoblastomas with mixed epithelial/mesenchymal histology (44% of patients[21]).

The prediction of a single driver event in cancers with the exponential age distribution of incidence does not mean that only a single driver gene can be discovered in such cancer types. In fact, many driver genes are redundant or even mutually exclusive, e.g. when the corresponding proteins are components of the same signaling pathway[22]. Thus, each tumor in such cancer types is expected to have a mutation in one driver gene out of a set of several possible ones, in which all genes most likely encode members of the same pathway or are responsible for the same cellular function. For example, in each neuroblastoma tumor sample, a mutation was present in only one out of 5 putative driver genes – *ALK*, *ATRX*,

PTPN11, *MYCN* or *NRAS*[23]. In hepatoblastoma, driver mutations cluster in the Wnt pathway[20,24,25].

Another aspect to consider is that while one mutation is usually sufficient to activate an oncogene, two mutations are typically required to inactivate both alleles of a tumor suppressor gene. Therefore, cancers with the exponential age distribution of incidence are predicted to have either a single somatic mutation in an oncogene, or a single somatic mutation in a tumor suppressor gene plus an inherited mutation in the same gene. The former is the case for neuroblastoma, where an amplification or an activating point mutation in *ALK* is often present[26-28] and for hepatoblastoma, where an activating point mutation or an indel in *CTNNB1* is most common[24,29,30]. The latter is the case for retinoblastoma, where an inactivating mutation in one allele of *RB1* is usually inherited, whereas an inactivating mutation in the other *RB1* allele occurs in a somatic cell[31].

Finally, the number of driver events predicted by the Erlang distribution refers exclusively to rate-limiting events responsible for cancer progression. For example, it was shown that inactivation of both alleles of *RB1* leads first to retinoma, a benign tumor with genomic instability that easily transforms to retinoblastoma upon acquiring additional mutations[32]. In this case, two mutations in *RB1* are rate-limiting, whereas mutations in other genes are not, because genomic instability allows them to occur very quickly. In neuroblastoma, frequent *MYCN* amplification and chromosome 17q gain are found only in advanced stages of the disease[33,34], so they are unlikely to be the initiating rate-limiting events. Hepatoblastoma is sometimes associated with familial adenomatous polyposis coli syndrome[35,36], and inheriting a mutation in the *APC* gene increases the risk of hepatoblastoma[37]. However, unlike other heritable mutations, an *APC* mutation does not lead to the earlier onset of hepatoblastoma[37], which indicates that it is not the rate-limiting initiating event. This is confirmed by the rarity of germline *APC* mutations in hepatoblastoma[38].

Overall, application of the gamma/Erlang distribution to childhood and young adulthood cancers showed its exceptionality amongst other probability distributions. The fact that it can successfully describe the radically different age distributions of incidence for cancers of any age and any type allows to call it the universal law of cancer development. Because it is based on the Poisson process, it demonstrates the fundamentally random timing of driver events and their constant average rate[13]. Moreover, it allows to predict the number and rate of driver events in any cancer subtype for which the data on the age distribution of incidence are available. It therefore may help to optimize the algorithms for distinguishing between driver and passenger mutations[39], leading to the development of more effective targeted therapies.

Methods

Data acquisition

United States Cancer Statistics Public Information Data: Incidence 1999 - 2012 were downloaded via Centers for Disease Control and Prevention Wide-ranging OnLine Data for Epidemiologic Research (CDC WONDER) online database (http://wonder.cdc.gov/cancer-v2012.HTML). The United States Cancer Statistics (USCS) are the official federal statistics on cancer incidence from registries having high-quality data for 50 states and the District of Columbia. Data are provided by The Centers for Disease Control and Prevention National Program of Cancer Registries (NPCR) and The National Cancer Institute Surveillance, Epidemiology and End Results (SEER) program. Results were grouped by 5-year Age Groups, Crude Rates were selected as output and All Ages were selected in the Age Group box. All other parameters were left at default settings. Crude Rates are expressed as the number of cases reported each calendar year per 100,000 population. A single person with more than one primary cancer verified by a medical doctor is counted as a case report for each type of primary cancer reported. The population estimates for the denominators of incidence rates are a slight modification of the annual time series of July 1 county population estimates (by age, sex, race, and Hispanic origin) aggregated to the state or metropolitan area level and

produced by the Population Estimates Program of the U.S. Bureau of the Census (Census Bureau) with support from the National Cancer Institute (NCI) through an interagency agreement. These estimates are considered to reflect the average population of a defined geographic area for a calendar year. The data were downloaded separately for each specific cancer type, upon its selection in the Childhood Cancers tab.

Data selection and analysis

For analysis, the data were imported into GraphPad Prism 5. Only cancers that show childhood/young adulthood incidence peaks and do not show middle/old age incidence peaks were analyzed further. The middle age of each age group was used as the x value, e.g. 17.5 for the "15-19 years" age group. Data were analyzed with Nonlinear regression. The following User-defined equations were created for the statistical distributions:

Gamma:
Y=A*(x^(k-1))*(exp(-x/b))/((b^k)*gamma(k))
Extreme value:
Y=A*(exp(-((x-t)/b)))*(exp(-exp(-((x-t)/b))))/b
Logistic:
Y=A*(exp((x-t)/b))/(b*((1+exp((x-t)/b))^2))
Normal:
Y=A*(exp(-0.5*(((x-t)/b)^2)))/(b*((2*pi)^0.5))
Weibull:
Y=A*(k/(b^k))*(x^(k-1))*exp(-((x/b)^k))

The parameter *A* was constrained to "Must be between zero and 100000.0", parameter *t* to "Must be between zero and 150.0", parameters *b* and *k* to "Must be greater than 0.0". "Initial values, to be fit" for all parameters were set to 1.0. All other settings were left at default values, e.g. Least squares fit and No weighting.

For the Erlang distribution, the parameter *k* for each cancer type was estimated by the fitting of the Gamma distribution, rounded to the nearest integer and used as "Constant equal to" in the second round of the Gamma distribution fitting, which provided the final results.

Author contributions statement

A.V.B. conceived of and performed the analysis and wrote the manuscript.

Competing financial interests

The author declares no competing financial interests.

References

1 Hornsby, C., Page, K. M. & Tomlinson, I. P. What can we learn from the population incidence of cancer? Armitage and Doll revisited. *The Lancet. Oncology* **8**, 1030-1038, doi:10.1016/S1470-2045(07)70343-1 (2007).

2 Nordling, C. O. A new theory on cancer-inducing mechanism. *British journal of cancer* **7**, 68-72 (1953).

3 Armitage, P. & Doll, R. The age distribution of cancer and a multi-stage theory of carcinogenesis. *British journal of cancer* **91**, 1983-1989, doi:10.1038/sj.bjc.6602297 (2004).

4 Knudson, A. G. Two genetic hits (more or less) to cancer. *Nature reviews. Cancer* **1**, 157-162, doi:10.1038/35101031 (2001).

5 Saltzstein, S. L., Behling, C. A. & Baergen, R. N. Features of cancer in nonagenarians and centenarians. *Journal of the American Geriatrics Society* **46**, 994-998 (1998).

6 Harding, C., Pompei, F. & Wilson, R. Peak and decline in cancer incidence, mortality, and prevalence at old ages. *Cancer* **118**, 1371-1386, doi:10.1002/cncr.26376 (2012).
7 Luebeck, E. G. & Moolgavkar, S. H. Multistage carcinogenesis and the incidence of colorectal cancer. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 15095-15100, doi:10.1073/pnas.222118199 (2002).
8 Little, M. P. & Wright, E. G. A stochastic carcinogenesis model incorporating genomic instability fitted to colon cancer data. *Mathematical biosciences* **183**, 111-134 (2003).
9 Michor, F., Iwasa, Y. & Nowak, M. A. The age incidence of chronic myeloid leukemia can be explained by a one-mutation model. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 14931-14934, doi:10.1073/pnas.0607006103 (2006).
10 Meza, R., Jeon, J., Moolgavkar, S. H. & Luebeck, E. G. Age-specific incidence of cancer: Phases, transitions, and biological implications. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 16284-16289, doi:10.1073/pnas.0801151105 (2008).
11 Calabrese, P. & Shibata, D. A simple algebraic cancer equation: calculating how cancers may arise with normal mutation rates. *BMC cancer* **10**, 3, doi:10.1186/1471-2407-10-3 (2010).
12 Luebeck, E. G., Curtius, K., Jeon, J. & Hazelton, W. D. Impact of tumor progression on cancer incidence curves. *Cancer research* **73**, 1086-1096, doi:10.1158/0008-5472.CAN-12-2198 (2013).
13 Belikov, A. V. The number of key carcinogenic events can be predicted from cancer incidence. *Scientific reports* **7**, 12170, doi:10.1038/s41598-017-12448-7 (2017).
14 Pon, J. R. & Marra, M. A. Driver and passenger mutations in cancer. *Annual review of pathology* **10**, 25-50, doi:10.1146/annurev-pathol-012414-040312 (2015).
15 Roy, D. M., Walsh, L. A. & Chan, T. A. Driver mutations of cancer epigenomes. *Protein & cell* **5**, 265-296, doi:10.1007/s13238-014-0031-6 (2014).
16 Knudson, A. G., Jr. Mutation and cancer: statistical study of retinoblastoma. *Proceedings of the National Academy of Sciences of the United States of America* **68**, 820-823 (1971).
17 Dimaras, H. *et al.* Retinoblastoma. *Lancet* **379**, 1436-1446, doi:10.1016/S0140-6736(11)61137-9 (2012).
18 Chen, Z. *et al.* Enhanced sensitivity for detection of low-level germline mosaic RB1 mutations in sporadic retinoblastoma cases using deep semiconductor sequencing. *Human mutation* **35**, 384-391, doi:10.1002/humu.22488 (2014).
19 Tolbert, V. P., Coggins, G. E. & Maris, J. M. Genetic susceptibility to neuroblastoma. *Current opinion in genetics & development* **42**, 81-90, doi:10.1016/j.gde.2017.03.008 (2017).
20 Cairo, S. *et al.* Hepatic stem-like phenotype and interplay of Wnt/beta-catenin and Myc signaling in aggressive childhood liver cancer. *Cancer cell* **14**, 471-484, doi:10.1016/j.ccr.2008.11.002 (2008).
21 Herzog, C. E., Andrassy, R. J. & Eftekhari, F. Childhood cancers: hepatoblastoma. *The oncologist* **5**, 445-453 (2000).
22 Vandin, F., Upfal, E. & Raphael, B. J. De novo discovery of mutated driver pathways in cancer. *Genome research* **22**, 375-385, doi:10.1101/gr.120477.111 (2012).
23 Pugh, T. J. *et al.* The genetic landscape of high-risk neuroblastoma. *Nature genetics* **45**, 279-284, doi:10.1038/ng.2529 (2013).
24 Koch, A. *et al.* Childhood hepatoblastomas frequently carry a mutated degradation targeting box of the beta-catenin gene. *Cancer research* **59**, 269-273 (1999).
25 Jia, D. *et al.* Exome sequencing of hepatoblastoma reveals novel mutations and cancer genes in the Wnt pathway and ubiquitin ligase complex. *Hepatology* **60**, 1686-1696, doi:10.1002/hep.27243 (2014).
26 Janoueix-Lerosey, I. *et al.* Somatic and germline activating mutations of the ALK kinase receptor in neuroblastoma. *Nature* **455**, 967-970, doi:10.1038/nature07398 (2008).

27 George, R. E. *et al.* Activating mutations in ALK provide a therapeutic target in neuroblastoma. *Nature* **455**, 975-978, doi:10.1038/nature07397 (2008).
28 Chen, Y. *et al.* Oncogenic mutations of ALK kinase in neuroblastoma. *Nature* **455**, 971-974, doi:10.1038/nature07399 (2008).
29 Wei, Y. *et al.* Activation of beta-catenin in epithelial and mesenchymal hepatoblastomas. *Oncogene* **19**, 498-504, doi:10.1038/sj.onc.1203356 (2000).
30 Udatsu, Y., Kusafuka, T., Kuroda, S., Miao, J. & Okada, A. High frequency of beta-catenin mutations in hepatoblastoma. *Pediatric surgery international* **17**, 508-512 (2001).
31 Friend, S. H. *et al.* A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. *Nature* **323**, 643-646, doi:10.1038/323643a0 (1986).
32 Dimaras, H. *et al.* Loss of RB1 induces non-proliferative retinoma: increasing genomic instability correlates with progression to retinoblastoma. *Human molecular genetics* **17**, 1363-1372, doi:10.1093/hmg/ddn024 (2008).
33 Brodeur, G. M., Seeger, R. C., Schwab, M., Varmus, H. E. & Bishop, J. M. Amplification of N-myc in untreated human neuroblastomas correlates with advanced disease stage. *Science* **224**, 1121-1124 (1984).
34 Bown, N. *et al.* Gain of chromosome arm 17q and adverse outcome in patients with neuroblastoma. *The New England journal of medicine* **340**, 1954-1961, doi:10.1056/NEJM199906243402504 (1999).
35 Garber, J. E. *et al.* Hepatoblastoma and familial adenomatous polyposis. *Journal of the National Cancer Institute* **80**, 1626-1628 (1988).
36 Hughes, L. J. & Michels, V. V. Risk of hepatoblastoma in familial adenomatous polyposis. *American journal of medical genetics* **43**, 1023-1025, doi:10.1002/ajmg.1320430621 (1992).
37 Hirschman, B. A., Pollock, B. H. & Tomlinson, G. E. The spectrum of APC mutations in children with hepatoblastoma from familial adenomatous polyposis kindreds. *The Journal of pediatrics* **147**, 263-266, doi:10.1016/j.jpeds.2005.04.019 (2005).
38 Harvey, J. *et al.* Germline APC mutations are not commonly seen in children with sporadic hepatoblastoma. *Journal of pediatric gastroenterology and nutrition* **47**, 675-677, doi:10.1097/MPG.0b013e318174e808 (2008).
39 Raphael, B. J., Dobson, J. R., Oesper, L. & Vandin, F. Identifying driver mutations in sequenced cancer genomes: computational approaches to enable precision medicine. *Genome medicine* **6**, 5, doi:10.1186/gm524 (2014).

Figures

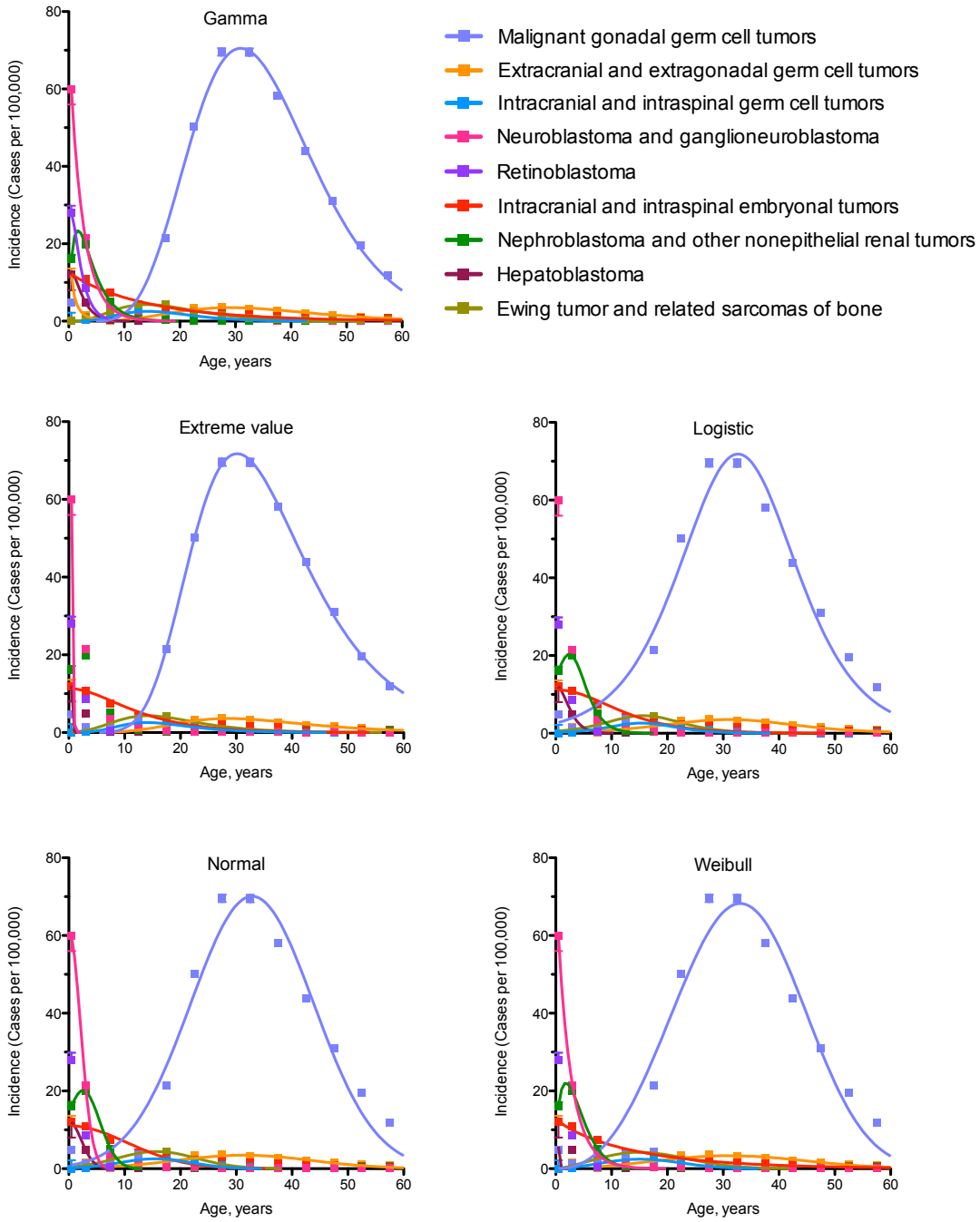

Figure 1. Comparison of different statistical distributions with actual distributions of childhood/young adulthood cancer incidence by age.

Dots indicate actual data for 5-year age intervals, curves indicate PDFs fitted to the data (see Table 1 for the R^2 comparison). The middle age of each age group is plotted. The fitting procedure was identical for all distributions and cancer types. Only cancers that show childhood/young adulthood incidence peaks and do not show middle/old age incidence peaks were selected.

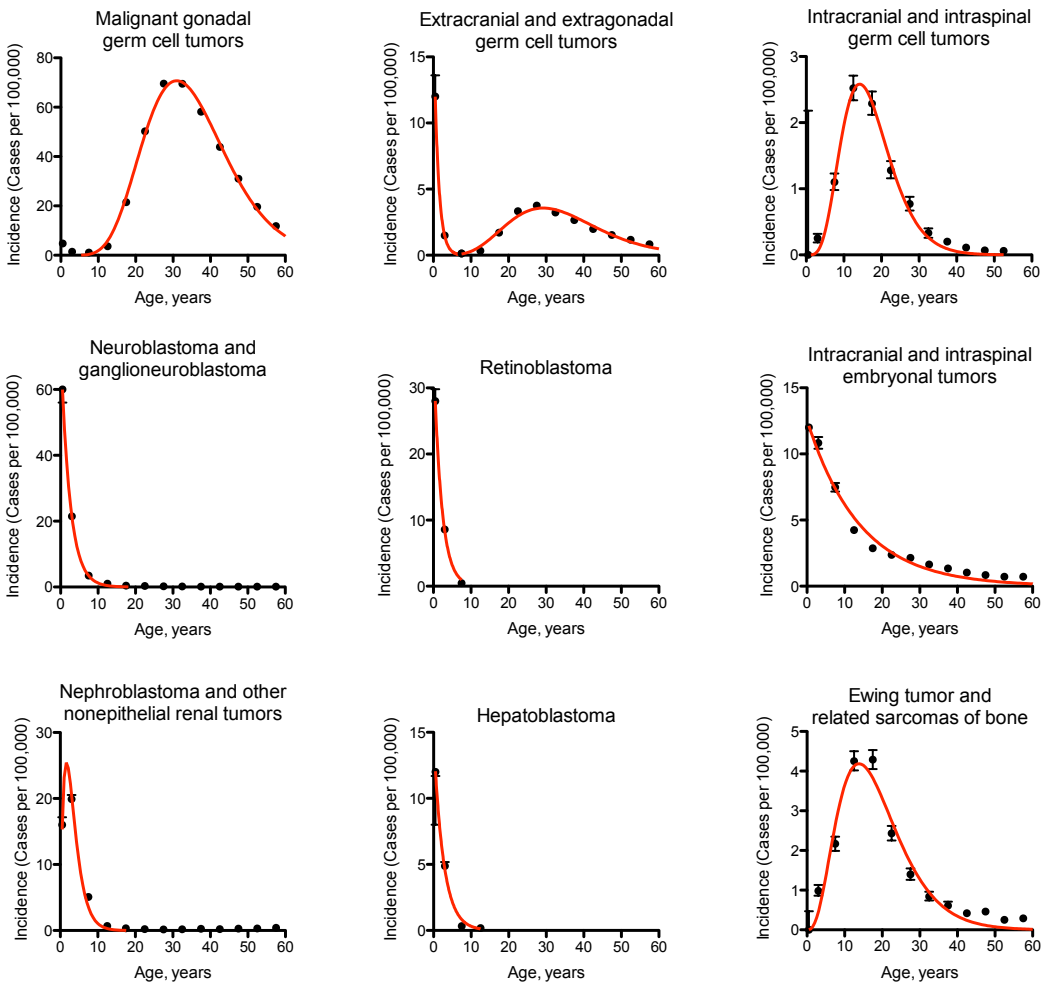

Figure 2. The Erlang distribution approximates cancer incidence by age for 10 childhood/young adulthood cancer types.

Dots indicate actual data for 5-year age intervals, curves indicate the PDF of the Erlang distribution fitted to the data (see Table 2 for the R^2 and estimated parameters). The middle age of each age group is plotted. The fitting procedure was identical for all cancer types. Only cancers that show childhood/young adulthood incidence peaks and do not show middle/old age incidence peaks were selected. Extracranial and extragonadal germ cell tumors of childhood and young adulthood are shown on the same plot.

Table 1. Comparison of the goodness of fit (R^2) of different statistical distributions to actual distributions of childhood/young adulthood cancer incidence by age.

Cancer type	Gamma	Extreme value	Logistic	Normal	Weibull
Malignant gonadal germ cell tumors	0.9922	**0.9967**	0.9606	0.9595	0.9547
Extracranial and extragonadal germ cell tumors of childhood*	**1.000**	NC	NC	NC	NC
Extracranial and extragonadal germ cell tumors of young adulthood	0.9495	**0.9680**	0.8725	0.8577	0.8689
Intracranial and intraspinal germ cell tumors	0.9898	**0.9948**	0.9609	0.9569	0.9698
Neuroblastoma and ganglioneuroblastoma*	**0.9998**	0.8696	NC	0.9964	**0.9998**
Retinoblastoma*	**1.000**	NC	NC	NC	NC
Intracranial and intraspinal embryonal tumors*	0.9779	0.9513	0.9405	0.9194	**0.9787**
Nephroblastoma and other nonepithelial renal tumors	**0.9970**	NC	0.9969	0.9963	0.9969
Hepatoblastoma*	**0.9997**	0.7406	0.9996	0.9987	0.3057
Ewing tumor and related sarcomas of bone	0.9553	**0.9679**	0.9471	0.9381	0.9501
Average (NC = 0)	**0.9861**	0.6489	0.6678	0.7623	0.7025
Average (NC not included)	**0.9861**	0.9270	0.9540	0.9529	0.8781

The best fit for each cancer type is highlighted in bold. NC – not converged. Cancer types with the exponential age distribution of incidence are marked by an asterisk. See Fig. 1 for graphical representation.

Table 2. Estimated carcinogenesis parameters for 10 childhood/young adulthood cancer types.

Cancer type	*k*	*b*	*A/1000*	R^2
	Number of driver events ± s.e.m.	**Average time between events, years ± s.e.m.**	**Maximal populational susceptibility, % ± s.e.m.**	**Goodness of fit**
Malignant gonadal germ cell tumors	9±0	3.87±0.03	1.96±0.03	0.9922
Extracranial and extragonadal germ cell tumors of childhood*	1±1	1.20±0.04	0.02±0.00	0.9999
Extracranial and extragonadal germ cell tumors of young adulthood	7±1	4.87±0.11	0.11±0.00	0.9476
Intracranial and intraspinal germ cell tumors	6±1	2.85±0.04	0.04±0.00	0.9888
Neuroblastoma and ganglioneuroblastoma*	1±0	2.44±0.03	0.18±0.00	0.9998
Retinoblastoma*	1±1	2.08±0.11	0.07±0.00	0.9993
Intracranial and intraspinal embryonal tumors*	1±0	14.11±0.91	0.18±0.01	0.9776
Nephroblastoma and other nonepithelial renal tumors	2±0	1.68±0.05	0.12±0.00	0.9904
Hepatoblastoma*	1±1	2.62±0.22	0.04±0.00	0.9966
Ewing tumor and related sarcomas of bone	4±1	4.60±0.15	0.09±0.00	0.9541

The parameters are determined for the Erlang distribution fitted to actual cancer incidence data (see Fig. 2). Cancer types with the exponential age distribution of incidence are marked by an asterisk.