Pathol. Oncol. Res. (2008) 14:199–204
DOI 10.1007/s12253-008-9050-z

ORIGINAL PAPER

A Mathematical Approach Predicting the Number of Events in Different Tumors

Ralf J. Rieker · Josef Hoegel · Michael A. Kern · Christina Steger · Sebastian Aulmann · Gunhild Mechtersheimer · Peter Schirmacher · Hendrik Blaeker

Received: 30 March 2008 / Accepted: 3 April 2008 / Published online: 23 April 2008

Abstract Supported by different investigations, multi-step models for tumorigenesis have been proposed for epithelial tumors. The age specific incidence of some cancers shows an exponential rise with increasing patient age. Yet, the onset and the slope of incidence curves varies between tumor types. One simple explanation for this disparity is that the number of mutations required for transformation differs in various tissues. We used a homogeneous Poisson process to estimate the number of events (N) and the intensity or event rate (λ) that might be needed for cancer development in various tissues (colon, prostate, oralpharynx, larynx). Estimations were performed, including 95% confidence intervals, for the male and female population. The expected number of events needed was higher in adenocarcinomas (colorectal carcinoma: $N\approx10$ for females and $N\approx11.0$ for males; prostatic cancer: $N\approx23$) than in squamous cell carcinomas (oropharynx: $N\approx5$–6 for females and $N\approx6$ for males; larynx: $N\approx7$ for females and $N\approx8$ males). Still, alternative models fixing N to values within the 95% confidence intervals determined, showed good coincidence with epidemiological data. Although the herein applied mathematical model neglects several biologic conditions, especially a presumed acceleration of mutation rates after tumor initiation it offers a plausible theory for the given epidemiologic data and matches with molecular biologic findings in the investigated cancers.

Keywords Incidence rates · Mutations · Poisson distribution · Probabilities · Tumorigenesis

R. J. Rieker · M. A. Kern · S. Aulmann · G. Mechtersheimer · P. Schirmacher · H. Blaeker (✉)
Department of Pathology,
University Hospital, University of Heidelberg,
INF 220/221,
69120 Heidelberg, Germany
e-mail: hendrik.blaeker@med.uni-heidelberg.de

J. Hoegel
Department of Human Genetics, University of Ulm,
Ulm, Germany

C. Steger
Department of Pathology, Medical University of Innsbruck,
Innsbruck, Austria

Present address:
R. J. Rieker
Department of Pathology, Medical University of Innsbruck,
Innsbruck, Austria

Introduction

Tumor development is considered to result from DNA mutations in tumor relevant genes and from epigenetic phenomenons. Mutations occur naturally in a somatic cell and have been associated with aging [1], since most functionally relevant mutations are likely to result in an impaired function of the coded protein. If an overall mutation rate of 5×10^{-9} per cell cycle is assumed, every cycle leads to approximately 25 mutations [2]. Following this calculation, a cell from highly regenerative tissue, for example the colon, acquires progressively several mutations in its life time [2]. If mutations unselectively target DNA, most of them will hit non-coding areas of the genome without consequence. Only very few mutations will alter the function of cell growth regulating genes (i.e. oncogenes and tumor suppressor genes) and only a combination of such mutations will cause cancer. The multi-step tumorigenesis with accumulation of mutations in relevant genes is molecularly established for cancer, best for colorectal carcinoma [3].

Common cancers of the western world/developed countries show a dramatic rise in incidence with increasing age. Graphical cancer statistics show exponential age specific incidence curves for which the onset and the degree of the slope varies between cancers of different tissues. It is a reasonable assumption that the variation of curves reflects differences in underlying mechanisms of tumor development. One such difference could be the tissue dependent number of mutations required for malignant transformation of a cell.

Drawing conclusions from incidence rates by using mathematical models has been established by several authors. Fisher and Holloman [4] as well as Nordling [5] found that the logarithm of death rate from cancer increased in direct to the logarithm of age. Nordling suggested, that the observed relationship was an end result of six or seven successive mutations. Armitage and Doll discussed a multi-stage theory [6], and a few years later they generated a two stage model which led to similar results concerning age specific incidence curves [7]. Knudson [8, 9] combined previous considerations with the idea of recessive oncogenesis, that led to the two-stage clonal expansion model [10].

In view of the improved knowledge of molecular genetics, the improved access to the epidemiological data and the as yet almost unexplained differences in incidence curves of different tumor types, we applied a statistical method to draw conclusions on the number of events/hits needed for the development of different cancers.

Materials and Methods

In the following we use the word event/hit instead of mutation, as in case of a tumor suppressor gene two mutations are needed for inactivation. To determine the number of necessary random occurrences of independent events from those observed 5 years incidence rates that grow exponentially with increasing age, we chose the homogeneous Poisson process as the underlying model. Within this model, the probability that exactly N events occur in the time interval $[0,t)$ is described as

$$f_{N,\lambda}(t) = \frac{(\lambda t)^N}{N!} \exp(-\lambda t),$$

denoting by exp the exponential function with base e and by $N!$ the product $N(N-1)\cdots 2\cdot 1$. The conditions under which this model is valid, include independence of the number of occurrences in nonoverlapping time intervals and dependence of the frequency in a single interval only on its length t (e.g. Breiman, p. 308) [11]. Then, the expected number of events in $[0,t)$ is λt, and for a short time interval Δt, the probability of one occurrence is about $\lambda \Delta t$. Hence, in this model, lambda (λ) is the intensity for the occurrence of one event, or an event rate. Using nonlinear regression analysis, the function $f_{N,\lambda}(t)$ was fitted to the observed probability of developing cancer before the age t calculated from the observed age-specific incidence curves. Optimal values for the number of events (N) and the intensity for the occurrence of one event (λ) were determined in order to minimize the residual sum of squares (RSS) within the regression model (a measure of deviation between observed and predicted frequencies). Calculations were performed using the NLIN procedure of SAS software, Version 8e. Further, 95% confidence intervals for the number N and the intensity λ were determined. In addition, alternative models other than those with the best fitting values were analyzed, starting with values for N below and above the 95% confidence limits output by NLIN. As a database, we used the 5 years age-specific incidence rates of the United States National Cancer Institute at http://seer.cancer.gov/csr/1975_2002/sections.html [12] and calculated the (cumulative) probability to have a cancer before age t. In case that incidence rates for young age cohorts were not given, we assumed these rates to be 0. The period analyzed started with the time interval 15–20 years and ended with the time interval 70–75 years. Based on their exponentially growing incidence rates, the following tumor sites were included: colon and rectum, larynx, oropharynx, prostate.

Results

Based on the Poisson distribution, we first investigated the estimated intensity λ and the number of events/hits N of the four investigated types of cancer (Table 1). Further, the 95% confidence interval for N and the RSS (i.e., residual sum of squares) as a measure of fit for the underlying model are given. Table 1 is divided into two subgroups: adenocarcinomas (Table 1) and squamous cell carcinomas (Table 1). In the group of the adenocarcinomas, prostate carcinomas had the highest number of calculated hits with an estimated N=23.2 in comparison to colorectal carcinomas with N=9.8 and N=11.0 calculated events, for the female and male population, respectively. Calculations in squamous cell carcinomas resulted in lower numbers of hits: Carcinomas of the larynx had N=6.7 events for the female and λ=7.7 events for the male population and carcinomas of the oropharynx had N=5.3 events for the female and N=5.8 events for the male population. The RSS, a measure of deviation between observed and predicted frequencies, assumed values less than 7.8×10^{-7}.

Then we calculated alternative models for colorectal tumors (Table 2). These alternative models were analyzed,

Table 1 Estimated values for the number of events with 95% confidence intervals and intensity values (λ), with the RSS for A and B

Location	Female		Male	
	Number of events (*N*) 95% Confidence interval	Lambda (λ) RSS	Number of events (*N*) 95% Confidence interval	Lambda (λ) RSS
A				
Colon, rectum	9.8	0.0529	11.0	0.0681
	9.4–10.1	1.3×10^{-8}	10.8–11.2	6.8×10^{-9}
Prostate	–	–	23.2	0.229
	–	–	21.3–25.1	7.8×10^{-7}
B				
Larynx	6.7	0.0153	7.7	0.0261
	6.0–7.4	4.2×10^{-10}	6.8–8.5	6.2×10^{-9}
Oropharynx	5.3	0.0116	5.8	0.0178
	5.2–5.4	2.5×10^{-10}	5.3–6.4	2.8×10^{-8}

A adenocarcinomas, *B* squamous cell carcinomas, *RSS* residual sum of squares

starting with values for *N* below and above the 95% confidence limits calculated under the best fitting model. These models also show comparable values for the goodness of fit.

Further, in Table 3 the observed incidence of cancer in the time interval (0,t] and the corresponding cumulative probabilities until age *t* predicted by the Poisson model are shown for colorectal cancer for the male and female population.

The transition from adenoma to carcinoma is considered to be the result of additional mutations within cells of the adenoma. Under the assumption, that adenomas of the colorectum require fewer hits than colorectal cancer, we finally calculated alternative cumulative probabilities (Tables 4 and 5). At the age of 75 the predicted cumulative probabilities for the female population at *N*=6/7 events were 0.103 and 0.058, respectively. For the male population the cumulative probabilities at *N*=7/8 events were 0.109 and 0.070, respectively.

Discussion

In previous publications several different models have so far been applied concerning carcinogenesis. Fisher and Holloman, Nordling as well as Armitage and Doll were the first to develop mathematical formulations concerning multistage carcinogenesis [4–7]. Further considerations by Armitage and Doll [6, 7] as well as Knudson [8, 9] led to the two stage expansion model. This stochastic model of carcinogenesis assumes that a malignant cancer cell arises following the occurrence of two critical mutations in a normal stem cell. The model provides some fits to a number of experimental and epidemiological data sets, however it is in some part too restrictive in assuming only two rate-limiting events.

Based on the Surveillance Epidemiology and End Results cancer statistics [12] of different tumor types we applied a mathematical model (homogeneous Poisson process) to draw conclusions on the number of relevant

Table 2 Alternative values for the numbers of events and the intensity value (λ), with the RSS, for colorectal carcinomas

	Female		Male	
	Events (*N*)		Events (*N*)	
	lambda (λ)	RSS	lambda (λ)	RSS
Colon, rectum	9.0	1.7×10^{-7}	10.0	2.5×10^{-7}
	0.046		0.058	
	9.5	1.0×10^{-7}	10.5	3.1×10^{-8}
	0.051		0.063	
	10.0	1.7×10^{-8}	11.0	8.4×10^{-8}
	0.055		0.068	
	10.5	1.5×10^{-7}	11.5	2.5×10^{-7}
	0.060		0.073	

RSS Residual sum of squares

Table 3 Cumulative probabilities to be diagnosed of colorectal cancer before a certain age, determined from the rates and predicted from the Poisson model

Age [years]	Cumulative probability ($*10^5$) of colorectal cancer in the female population	Predicted cumulative probability ($*10^5$) of colorectal cancer in the female population	Cumulative probability ($*10^5$) of colorectal cancer in the male population	Predicted cumulative probability ($*10^5$) of colorectal cancer in the male population
20	0.0	0.0	0.0	0.0
25	0.8	0.2	0.7	0.2
30	2.6	0.9	2.3	0.8
35	6.1	3.2	6.0	3.1
40	12	8.9	14	9.5
45	25	22	28	25
50	51	46	57	57
55	93	90	115	115
60	159	161	212	215
65	264	270	367	369
70	426	426	599	595
75	644	641	906	907

hits/events underlying tumor development. We fitted the model using only the age groups from 15–20 to 70–75 years since for patients younger than 15 age-specific incidence rates for the investigated carcinomas are not included in the database (and assumed to be 0 for our purposes) and the incidence rates beyond the age of 75 do not follow the same exponential rise as before the age of 75. The reason for the latter discontinuity is not clear but may reflect a positive selection of people beyond 75 with a different pattern of internal or external risk factors.

Calculations were performed for four different types of carcinomas that are known to be relatively homogenous in differentiation. They were not applied to cancer types that occur at the same location but naturally present with very different histologic features such as lung and stomach, since we presumed different types and numbers of molecular alterations underlying the malignant transformation process in the different subtypes (e.g. diffuse versus intestinal gastric carcinoma, small cell carcinoma versus adenocarcinoma of the lung).

The presented calculations are based on a mathematical model that naturally has to drastically simplify the complex process of tumorigenesis. The assumptions of the Poisson model [11] are that (1) the probability of a single mutation in a very short time interval is $\lambda\Delta\ t$ and that (2) the probability of more than one mutation during the time interval is negligible and that (3) the probability of a mutation during the time interval is independent of any occurrences in previous periods. Especially the last point is unlikely to be given under in vivo conditions, as some genes involved in tumor formation play a crucial role e.g. in the cell cycle or in apoptosis; increasing the likelihood of

Table 4 Predicted cumulative probabilities to have cancer before a certain age, for colorectal adenomas

Age [years]	Female population N=6	Female population N=7	Female population N=8	Female population N=9
20	0.0007	0.0001	0.0000	0.0000
25	0.002	0.0003	0.0001	0.0000
30	0.005	0.001	0.0002	0.0000
35	0.009	0.002	0.0005	0.0001
40	0.015	0.005	0.001	0.0003
45	0.023	0.008	0.002	0.0006
50	0.034	0.013	0.004	0.001
55	0.046	0.019	0.007	0.002
60	0.059	0.027	0.011	0.004
65	0.074	0.036	0.016	0.006
70	0.088	0.047	0.022	0.009
75	0.103	0.058	0.028	0.013

For the female (λ=0.0529) population with different number of events (N)

Table 5 Predicted cumulative probabilities to have cancer before a certain age, for colorectal adenomas

Age [years]	Male population N=7	Male population N=8	Male population N=9	Male population N=10
20	0.0004	0.0001	0.0000	0.0000
25	0.002	0.0003	0.0001	0.0000
30	0.004	0.001	0.0002	0.0000
35	0.008	0.002	0.0006	0.0002
40	0.014	0.005	0.001	0.0004
45	0.024	0.009	0.003	0.0009
50	0.035	0.015	0.006	0.002
55	0.048	0.023	0.009	0.004
60	0.063	0.032	0.015	0.006
65	0.079	0.044	0.021	0.009
70	0.094	0.056	0.030	0.014
75	0.109	0.070	0.039	0.020

For the male (λ=0.0681) population with different number of events (N)

subsequent mutations in tumor relevant genes. However, this mathematical model offers one conceptually simple explanation for the obvious differences in incidence curves of different tumors and could provide the basis for more sophisticated models which may include as yet unknown factors of mutational acceleration. It does not depend on the validity of multi-stage clonal expansion and simply assumes that cancer is the result of a series of independent events which need not be characterized in full depth.

At the molecular level, colorectal carcinoma is probably the best investigated and understood epithelial tumor [3]. The majority of colorectal carcinomas are sporadic tumors arising via adenomatosis polyposis coli (APC) inactivation and *k-ras* activation. A smaller subset of tumors displays microsatellite instability with members of the transforming growth factor ß family-growth suppressing signaling cascade typically affected by secondary mutational events. Given the molecular findings in microsatellite-stable colorectal carcinomas the model by Fearon and Vogelstein [3] predicts a carcinoma development involving monoallelic *k-ras* (one event), and biallelic mutations of the *APC, DCC, p53* (6 events). This would result in a total of at least seven events. Recent data, however, show that a variety of others genes is also frequently mutated in colorectal carcinoma. While *BRAF* mutations [13, 14] may substitute for *k-ras* mutations and therefore do not elevate the number of necessary events, recent studies stress the importance of oncogenic activation of *PIK3C* and inactivation of tyrosine phosphatases [15, 16]. Thus it is reasonable to assume more than seven (approximately ten) events/hits to be necessary for transformation of colonic epithelial cells. The number of about ten events proposed by our model is very much in keeping with this consideration.

Development of sporadic colorectal carcinoma typically follows an adenoma–carcinoma sequence. The transition from adenoma to carcinoma is considered to be the result of additional mutations within cells of the adenoma, which confer a growth advantage and enable the cells to invade the surrounding tissue. Presuming that the intensity of events in colorectal adenoma is the same as for carcinoma we calculated alternative models with fewer events and compared the data with the cumulative incidence rates of colorectal adenomas given by Loeve et al. (approximately 9% at 79 years) [17]. Our calculations revealed similar predicted cumulative probabilities of approximately six to seven events for females and seven to eight events for males necessary for adenoma development in the colon (Tables 4 and 5). This means that the transition from adenoma to carcinoma is the result of about three additional tumor relevant hits in an adenoma cell.

According to the applied mathematical model, other adenocarcinomas seem to need a higher number of events. The number of hits needed for carcinoma development in the prostate would be exceptionally high (N=24). Compared to colorectal carcinoma the molecular pathogenesis of prostate cancer is less understood and does not allow stringent comparison with known mutations in this neoplasm. Genomewide linkage analyses and loss of heterozygosity studies revealed 12 loci probably involved in prostate cancer [18]. Mutations have been observed in varying frequencies in at least 11 tumor suppressor genes (thus 22 events) and the existence of non-random deletions at specific chromosomal locations point to a number of as yet unknown tumor suppressor genes possibly involved in prostate carcinogensis. Further, at least ten oncogenes have been described to be mutationally activated in prostate cancer [19]. In a review by Karayi et al. more than 15 different alterations are mentioned in localized prostate cancer (including alterations in normal prostate cells) [19]. Whether the diversity of tumour suppressor gene mutations

in prostatic carcinomas reflects a variety of alternative tumor-developmental pathways or indicates a higher number of mutations needed for malignant transformation remains to be determined. The conclusions in our model strengthen the latter hypothesis. The large number of mentioned genes offers a logical explanation for the peculiar late onset and slope of incidence in this tumor type, which may also a hint for hormonal influences.

In contrast to adenocarcinomas of the colorectum and prostate our calculations revealed lower numbers between five and eight relevant events for squamous cell carcinoma in the oropharynx and the larynx. Again a comparison of our calculated number with molecular data is more difficult than for colorectal cancer, since a clear concept of molecular carcinogenesis does not exist in these entities. The number of genes described to be mutated in squamous cell carcinomas is higher than the number we calculated [20]. However, for some cases of squamous cell carcinoma an oligostep genetic pathway has been described [21]. Basically, a combination of mutations in *p53* and *p16*, deletions at chromosome *9p* and amplification of cyclin D1 seem to be sufficient for development of squamous cell carcinoma in the larynx and the pharynx. Other frequent alterations, like overexpression of epidermal growth factor receptor and overexpression of Her2/neu are epigenetic and are not associated with genetic events in the corresponding genes [20, 21]. One clinical aspect that may strengthen the idea of fewer mutations in the development of squamous cell carcinoma in larynx and pharynx is the phenomenon of field cancerisation, which indicates several independent transformed cell clones. Squamous cell carcinomas of the larynx and the oropharynx are strongly associated with smoking and alcohol intake. If only a small number of mutations is needed for transformation the likeliness of simultaneous tumor occurrence under exposure towards strong carcinogens should be higher.

In conclusion, we present a mathematical analysis designed to draw conclusions on the number of relevant events underlying the development of four common cancers. Although the model does not take into account the complexity of interdependence of mutations, the results match well with molecular findings of some of the investigated tumors and offer explanations for the obvious discrepancies in incidence curves between the investigated tumor types.

References

1. Hasty P, Campisi J, Hoeijmakers J et al (2003) Aging and genome maintenance: lessons from the mouse? Science 299:1355–1359
2. Tomlinson I, Sasieni P, Bodmer W (2002) How many mutations in a cancer? Am J Pathol 160:755–758
3. Fearon ER, Vogelstein B (1990) A genetic model for colorectal tumorigenesis. Cell 61:759–767
4. Fischer JC, Hollomon JH (1951) A hypothesis for the origin of cancer foci. Cancer 4:916–918
5. Nordling CO (1953) A new theory on the cancer-inducing mechanism. Br J Cancer 7:68–72
6. Armitage P, Doll R (1954) The age distribution of cancer and a multi-stage theory of carcinogenesis. Br J Cancer 8:1–12
7. Armitage P, Doll R (1957) A two-stage theory of carcinogenesis in relation to the age distribution of human cancer. Br J Cancer 11:161–169
8. Knudson AG (1971) Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci USA 68:820–823
9. Knudson AG (1973) Mutation and human cancer. Adv Cancer Res 17:317–352
10. Hazelton WD, Clements MS, Moolgavkar SH (2005) Multistage carcinogenesis and lung cancer mortality in three cohorts. Cancer Epidemiol Biomarkers Prev 14:1171–1181
11. Breiman L (ed) (1968) Probability. Wesley, Reading
12. Ries LAG, Eisner MP, Kosary CL et al (eds) (2005) SEER Cancer Statistics Review, 1975–2002. National Cancer Institute, Bethesda. Accessed at http://seer.cancer.gov/csr/1975_2002/, based on November 2004 SEER data submission, posted to the SEER web site 2005. Cited 15 Jan 2007
13. Davies H, Bignell GR, Cox C et al (2002) Mutations of the BRAF gene in human cancer. Nature 417:949–954
14. Yuen ST, Davies H, Chan TL et al (2002) Similarity of the phenotypic patterns associated with BRAF and KRAS mutations in colorectal neoplasia. Cancer Res 62:6451–6455
15. Samuels Y, Velculescu VE (2004) Oncogenic mutations of PIK3CA in human cancers. Cell Cycle 3:1221–1224
16. Wang Z, Shen D, Parsons DW, Bardelli A et al (2004) Mutational analysis of the tyrosine phosphatome in colorectal cancers. Science 304:1164–1166
17. Loeve F, Boer R, Zauber AG et al (2004) National Polyp Study data: evidence for regression of adenomas. Int J Cancer 111:633–639
18. Nwosu V, Carpten J, Trent JM et al (2001) Heterogeneity of genetic alterations in prostate cancer: evidence of the complex nature of the disease. Hum Mol Genet 10:2313–2318
19. Karayi MK, Markham AF (2004) Molecular biology of prostate cancer. Prostate Cancer Prostatic Dis 7:6–20
20. Kim MM, Califano JA (2004) Molecular pathology of head-and-neck cancer. Int J Cancer 112:545–553
21. Pai SI, Wu GS, Ozoren N et al (1998) Rare loss-of-function mutation of a death receptor gene in head and neck cancer. Cancer Res 58:3513–3518