Genomic Instability and Radiation Risk in Molecular Pathways to Colon Cancer

Jan Christian Kaiser*, Reinhard Meckbach, Peter Jacob

Institute of Radiation Protection, Helmholtz Zentrum München, Deutsches Forschungszentrum für Gesundheit und Umwelt (GmbH), Neuherberg, Germany

Abstract

Colon cancer is caused by multiple genomic alterations which lead to genomic instability (GI). GI appears in molecular pathways of microsatellite instability (MSI) and chromosomal instability (CIN) with clinically observed case shares of about 15–20% and 80–85%. Radiation enhances the colon cancer risk by inducing GI, but little is known about different outcomes for MSI and CIN. Computer-based modelling can facilitate the understanding of the phenomena named above. Comprehensive biological models, which combine the two main molecular pathways to colon cancer, are fitted to incidence data of Japanese a-bomb survivors. The preferred model is selected according to statistical criteria and biological plausibility. Imprints of cell-based processes in the succession from adenoma to carcinoma are identified by the model from age dependences and secular trends of the incidence data. Model parameters show remarkable compliance with mutation rates and growth rates for adenoma, which has been reported over the last fifteen years. Model results suggest that CIN begins during fission of intestinal crypts. Chromosomal aberrations are generated at a markedly elevated rate which favors the accelerated growth of premalignant adenoma. Possibly driven by a trend of Westernization in the Japanese diet, incidence rates for the CIN pathway increased notably in subsequent birth cohorts, whereas rates pertaining to MSI remained constant. An imbalance between number of CIN and MSI cases began to emerge in the 1980s, whereas in previous decades the number of cases was almost equal. The CIN pathway exhibits a strong radio-sensitivity, probably more intensive in men. Among young birth cohorts of both sexes the excess absolute radiation risk related to CIN is larger by an order of magnitude compared to the MSI-related risk. Observation of pathway-specific risks improves the determination of the probability of causation for radiation-induced colon cancer in individual patients, if their exposure histories are known.

Introduction

Cancer is caused by multiple genomic alterations which lead to genomic instability. Two principal molecular forms of genomic instability have been observed in tissue from colorectal tumors. High-level microsatellite instability (MSI) appears if DNA mismatch repair (MMR) genes are defective. MSI tumors exhibit frequent mutations in short repeated DNA sequences called microsatellites. About 15–20% of sporadic cases are related to MSI which often starts with silencing of the MMR gene MLH1 by promoter methylation. Chromosomal instability (CIN) constitutes the second form of genomic instability which is less clearly defined. CIN tumors show a large heterogeneity in chromosomal copy number and structure (named aneuploidy), whereas MSI tumors are near-diploid with few karyotypic abnormalities. CIN is associated with the loss of wild-type copies of tumor suppressor genes (TSGs) such as APC, TP53 or SMAD4 which regulate the growth and death of cells with tumorigenic mutations. CIN tumors are micro-satellite stable through effective mismatch repair. About 80–85% of colorectal tumors are of the CIN (or low-level MSI) type. However, CIN and MSI can share molecular properties such as mutations in the BRAF gene or the CpG island methylator phenotype (CIMP). The pathways are not mutually exclusive and a more refined pathway classification has been suggested [1–8].

About 20–30% of patients with colorectal cancer possess a familial risk with two or more first or second degree relatives having colorectal cancer but only 5–10% of all patients develop the disease in a strictly inherited manner [9]. The two main phenotypes are hereditary nonpolyposis colorectal cancer (HNPCC or Lynch syndrome) and familial adenomatous polyposis (FAP) [10]. FAP constitutes the hereditary form of the CIN pathway. A germline mutation in the APC gene has been detected in about 80% of FAP patients [9]. FAP and attenuated FAP are also related to bi-allelic inherited mutations of the MutYH gene without showing APC mutations [11]. HNPCC is associated with the MSI pathway and causes about 3% of colorectal tumors [10].

Environmental factors and lifestyle can also influence the formation of colorectal cancer but the impact may vary in colon and rectum [12]. For the present analysis, which is focused on the life span study (LSS) of a-bomb survivors, the Japanese setting is of special interest. A westernized diet has been identified as an important cause for the increase of colon cancer incidence in Japan [13]. Incidence rates converged to those in the US population from the late 1950s to the mid-1990s [14].

A number of mathematical models have been developed to represent the most important biological processes of colon carcinogenesis. A model with a single path to colon cancer has been fitted to incidence data of the Surveillance, Epidemiology
and End Results (SEER) registry [15]. APC loss was included as the promoter of genetic chaos consistent with Knudson's two-hit paradigm of oncogenesis. At the same time a conceptual model has been proposed which considered CIN as an early event [16]. Cells carrying a CIN property together with silenced APC genes are expected to show markedly elevated mutation rates compared to cells with silenced APC genes alone. Both models were included in a comparative analysis of five mechanistic models and a descriptive model which have been fitted to the SEER data [17]. Based on goodness-of-fit criteria no clear winner emerged from this exercise. An explorative study of models which explicitly aimed to identify traits of MSI and CIN in the SEER data did not produce compelling evidence [18]. Typical time scales for the development of precancerous lesions and tumor growth in different sites of the gastro-intestinal tract (including the colon) have been detected by applying multi-stage clonal expansion (MSCE) models to SEER data [19,20]. MSI tumors are rarely found in the distal colon [2]. Compared to the proximal colon slightly faster adenoma growth has been observed in the distal colon, possibly caused by differential oncogenic dynamics of the CIN and MSI pathways [21]. None of the precursor models unambiguously discovered imprints of distinct molecular pathways for incidence data of colorectal cancer in the SEER cohort.

Mechanisms that relate radiation to genomic instability are still not fully explained [22–24]. Radiation-induced genomic instability and other molecular radiation effects have been mimicked in biologically-based models of carcinogenesis for several organs [25]. To estimate radiation risks simple two-stage models of initiation and promotion have been applied to incidence data from the LSS cohort [26,27]. These mechanistic models with radiation effects relied on a uniform description of tumorigenic processes which did not address organ-specific peculiarities.

For many organs, including the colon, estimates of the excess absolute risk (EAR) and excess relative risk (ERR) are derived from the LSS data with descriptive models [28]. Such estimates are considered as the accepted standard by committees BEIR VII [29], ICRP [30] and UNSCEAR [31], which issue recommendations for radiation protection. They are applied in compensation claims from nuclear workers and US army veterans [32]. Particularly for colon cancer, reliable risk coefficients are needed for a risk-benefit analysis of mass screening by computer tomography colonography (CTC) [33].

In the present study biologically-based modelling is applied to detect imprints of pertinent tumorigenic processes for colon cancer in the LSS incidence data. It is aimed to reproduce the share of clinically observed cases in the MSI and CIN pathways. For both sexes the total risk and pathway-specific risks are compared to standard risk coefficients from descriptive models.

Materials and Methods

LSS dataset of colon cancer incidence

In August 1945, residents of Hiroshima and Nagasaki were acutely exposed to a mixed field of γ-radiation and neutrons from two a-bomb explosions. Individual radiation doses are represented in the latest dosimetry system DS02 [34]. For the neutron contribution to the total colon dose a weight of ten is used which is motivated by higher biological effectiveness. Incidence data for solid cancers were collected from 1956 onwards for 120,321 members of the LSS cohort to assess late health effects. Subjects came from all age groups and were not selected for pre-existing illness.

The LSS cohort was created as a stratified random sample of the entire available population of such survivors, including all of the available survivors who had been exposed to the bombs at proximal distances. In addition to extensive collection of demographic and exposure data in the early years after cohort inception, the cohort has been followed for mortality using nationwide data in Japan and for cancer incidence by tumor registries established in both Hiroshima and Nagasaki. Operation of the Hiroshima and Nagasaki tumor registries is reviewed regularly by the institutional review boards of the Radiation Effects Research Foundation (RERF) and the registries. Protocols used for vital status and cause of death ascertainment in the LSS are regularly reviewed by the RERF board. The protocols include the assurance that patient information would be kept confidential and grant permission to access the stored information. With the approval of the tumor registries, the RERF cohorts are routinely linked with the registries to identify tumors among cohort members. The full LSS data set is publicly available in file lssinc07.csv from the RERF website. The set consists of 24,205 completely anonymised Poisson records in grouped form which prevents the identification of individual patient information.

Although carcinogenesis acts very similar in colon (ICD10:C18) and rectum (ICD10:C20), the rectum data are discarded in the present study. The radiation risk for the rectum is negligible in the LSS [28]. Person years (PY) and cases in excess of 4 Gy shielded air kerma have been excluded to avoid modelling of deterministic radiation effects. These exclusions reduce the number of colon cancer cases by 8 (5 male/3 female) to 1508 (Table A4 in ref. [28]). A summary of the LSS data for colon cancer incidence is given in Table 1.

The person-year weighted mean dose of about 0.081 Gy for both sexes combined is very similar to the subject-weighted mean dose of 0.083 Gy (0.085 Gy male, 0.081 Gy female). The case-weighted mean dose for both sexes combined is 0.12 Gy. A higher value for the case-weighted mean indicates an association of colon cancer and radiation. But the association appears notably weaker in women than in men. To confirm this observation, relative risks of groups with low to moderate (0.005–0.25 Gy) dose and moderate to high (>0.25 Gy) doses have been calculated compared to the unexposed (<0.005 Gy) population. The crude data show a significant relative risk in the group with moderate to high doses only for men or for both sexes combined (Table 2). Relative risks from crude data are only indicative and cannot replace a proper risk assessment study.

Mechanistic model

The present cell-based model for the two main molecular pathways to colon cancer (in short two path (TP) model) relies on the concept of growth control for precancerous lesions by caretaker and gatekeeper genes [35,36]. Although cell alignment and spatial movement play a role in tumorigenesis [37], the two path model is only concerned with the kinetics of mutations and cell growth.

Colon epithelium consists of a single cell layer organized in finger-shaped crypts. Each of the many million crypts houses a small stem cell population in a niche at the bottom. The total population of healthy stem cells is in homeostasis and can reproduce all intestinal cell types by asymmetric division [37]. This process generally creates one altered daughter cell and leaves the other daughter cell unchanged. In the simplified conceptual model of Figure 1 the pathways to cancer are initiated either by bi-allelic mutation of the APC gene (CIN) or by bi-allelic methylation of the MLH1 gene (MSI) [1]. Both genetic alterations are generated in asymmetric cell division. The baseline rates \( \lambda_{1} \) and \( \lambda_{2} \) for genetic alterations in the first and second hit cannot be determined independently [19]. These two successive rates have been set equal
but differences between pathways were allowed. No further assumptions for the cell kinetics in the two path model were made.

Clonal expansion of cells with tumorigenic mutations generates neoplastic lesions, which undergo further transitions on the way to cancer. Clonal growth of initiated cells is a stochastic process, in the early stage clones may die out or survive. Growth of adenoma starts with surviving clones in separate crypts (monocryptal adenoma). It is assumed that a cycle of crypt fission and extinction dominates clonal expansion in of premalignant cells at this early stage [38]. Crypt fission is a very slow process which occurs on average once in 2–3 decades [37]. In the two path model initiated cells either divide symmetrically with rate \( r_I \) or are inactivated (i.e. by apoptosis or extinction) with rate \( b_I \). A direct functional relation between the net growth rate \( \gamma = r_I - b_I \) and the rate of crypt fission is not obvious, since \( \gamma \) belongs to events for single cells and crypt fission involves many cells. However, both rates depend on the same underlying cell kinetics and similar numerical values for such rates seem plausible. As an effective rate parameter \( \gamma \) describes the growth dynamics together in monocryptal adenoma and in the crypt cycle equally for both pathways. Crypt fission at a normal rate is the mechanism which spreads inactivated TSGs such as \( APC \) or \( MLH1 \) in the human colon [39]. During growth of early adenoma the transient patterns of MSI and CIN diverge possibly due to different effects of silenced TSGs in both pathways. In the model the pathways are treated as independent so that incidence rates for different pathways can be added to obtain the total incidence. However, in reality some molecular processes such as deregulated WNT signalling are found in both pathways [4–6,8].

A single transforming mutation \( v_{MSI} \) concludes the MSI path by creation of at least one malignant cell which leads to a tumor. Although the MSI path exhibits a higher degree of complexity, a simplification is justified by the small number of expected cancer cases from MSI [2].

The CIN pathway continues with a destabilizing event of rate \( \lambda_{CIN} \) which precedes clonal growth in larger adenoma. To account for lifestyle trends, \( \lambda_{CIN} \) is scaled by an exponential factor \( \exp[\beta_s/(1915.6-b)] \) which increases with birth year \( b \). The net rate of stochastic clonal growth \( \gamma_{CIN} = \gamma_{CIN} - b_{CIN} \) for CIN cells is determined by the difference between symmetric cell division \( s_{CIN} \) and inactivation \( b_{CIN} \). Transformation of CIN cells with mutation rate \( \mu_{CIN} \) to at least one malignant cell, which leads to a tumor, is considered as the final rare event of tumorigenesis in the CIN pathway. In both pathways a fixed lag time \( t_{lag} = 5 \) yr is chosen for the duration until the first malignant cell grows into a clinically relevant tumor.

Radiation action has been assumed to increase the rate \( v_{T2} \) of the second hit in initial mutations or in hypermethylation. Reduction of the inactivation rate \( b_{CIN} \) for CIN cells was applied as a second radiation effect (Figure 1). Reduced cell inactivation is a plausible mechanism to promote clonal growth [40]. Combined radiation action on cell division and inactivation or on division alone might also be considered but different radiation effects in promotion have negligible influence on the fit results. Radiation action on the destabilizing CIN event and other radiation targets (results not reported) have been tested as well. The statistical quality of model fits has been measured by the Akaike Information Criterion [\( AIC = \text{deviance} + 2 \times \text{no. of model parameters} N_{param} \)] [41].

### Numerical solution of the two path model

The two path model fits into the mathematical framework of Little and Wright [42] who have generalized the two-step clonal expansion (TSCE) model introduced by Moolgavkar and Knudson [43]. The TSCE model relies on two rate-limiting mutations which are separated by clonal expansion of initiated cells. Mutation rates and rates of cell division or inactivation are treated as transient Poisson point processes of cell birth and death which are expressed in a set of master equations [44]. The approach to solve the TSCE model for piecewise constant model parameters has been extended to the larger set of master equations for the two

<table>
<thead>
<tr>
<th>Dose group</th>
<th>RR Men</th>
<th>RR Women</th>
<th>RR Both sexes</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.005–0.25 Gy</td>
<td>1.05 (0.89; 1.24)</td>
<td>0.99 (0.85; 1.14)</td>
<td>1.01 (0.90; 1.13)</td>
</tr>
<tr>
<td>&gt;0.25 Gy</td>
<td>1.77 (1.42; 2.21)</td>
<td>1.03 (0.81; 1.31)</td>
<td>1.35 (1.14; 1.58)</td>
</tr>
</tbody>
</table>

**Table 2.** Relative risk (RR) with 95% CI in colon dose groups 0.005–0.25 Gy and >0.25 Gy compared to dose group <0.005 Gy.

**Table 1.** Summary of colon cancer incidence data in the LSS cohort from 1958–1998 for dose groups with shielded air kerma < 4 Gy, 95% percentiles of frequency distribution for person years in brackets.
path model [45]. This set has been transformed into a system of coupled differential equation of the Ricatti type which is solved efficiently by an approximate iterative algorithm for calculating the survival function. The hazard is obtained by numerical differentiation of the survival function. The total hazard of the two path model is given by the sum of the hazard for the separate MSI and CIN models. Mathematical derivations of equivalent models have been given in ref. [19] (MSI without tlag) and ref. [20] (CIN without tlag) in a notation which is applied in the present study in a similar way.

Identification of model parameters

Eight different parameters for biological transition rates are shown in Figure 1. These rates should be at least in principle accessible for experimental investigation. But the differential equations for the two path model are couched in terms of less intuitive identifiable parameters. The identifiability problem follows from the mathematical model structure and cannot be removed by increasing statistical power [46]. In the so-called deterministic versions of the MSI and CIN models, Mathematical derivations of equivalent models have been given in ref. [19] (MSI without tlag) and ref. [20] (CIN without tlag) in a notation which is applied in the present study in a similar way.

Parameter estimation and uncertainty analysis

The MECAN software package has been used for pre-processing of the grouped data, regression, comparison of observed and expected cases, and simulation of uncertainty intervals [47]. The package is written in the C++ programming language. Its object-oriented design is based on separate libraries for processing of epidemiological data sets and for the introduction of new mechanistic or descriptive risk models. The libraries are linked to the computational core which performs the standard tasks of likelihood minimisation and simulation of uncertainties for risk estimates. Thanks to a high degree of standardisation, new projects of radio-epidemiological analysis can be set up with little programming effort. Parallelisation has been achieved by linking the code to functions of the OpenMP library (www.openmp.org).

MECAN includes the C++ library Minuit2 from CERN which is used for the minimization of $-2 \ln L$, where $L$ denotes the
Poisson likelihood [48]. The Poisson deviance is given by the minimum of \(-2 \ln L\), which is reached with the maximum likelihood estimates (MLE) of the model parameters. It is assumed that a parabolic approximation of the region around the minimum is valid. In this case Wald-based standard errors (SE), confidence intervals (CI), and a correlation matrix can be computed for the model parameters. Results of MECAN were found to be in good agreement with the EPICURE package which is a standard software for the analysis of radio-epidemiological data [49].

For the conceptual model of Figure 1 with different identifiable baseline parameters for both pathways and both sexes. But different parameters have been kept in the model only if the fit was improved with a probability of at least 95% (or the deviance was lowered by at least 3.8 points) in a likelihood ratio test (LRT). Radiation dependent parameters have been added only if they passed the same LRT. More details on the statistical analysis approach of model parameter selection are given in ref. [50].

Results

Goodness-of-fit

In Table 3 Poisson deviance and AIC for the tested models are shown. In the mechanistic TP models the fixed lag time \(t_{lag} = 5 \text{yr}\) was counted as an additional model parameter, while the remaining identifiable parameters have been determined by a fit. Two path model, TP0 without a radiation effect provides the benchmark for models TP1 to TP4, which show similar goodness-of-fit for different radiation targets. Replacing radiation action for men only in model TP3 by unisex radiation action on the destabilizing CIN event did not improve the fit compared to model TP1. Model TP4 yielded the lowest Poisson deviance and AIC, and is preferred for risk assessment in the present study. MLE, SE and \(\sigma_{\text{CI},p}\) from the likelihood profile are given in Table 4 for the identifiable parameters. Radiation-dependent versions of the mechanistic models by Meza et al. [19] (M1, Figure S1 in File S1), and by Little and Li [17] (M2, Figure S2 in File S1) yielded AIC values which came out higher by 13 points and 8 points, respectively. Fitting a radiation-dependent version of the full MSCE model by Luebeck et al. [20] was not successful. Parameter estimates for models M1 and M2 are shown in Tables S2 and S3 in File S1. The baseline parameters of model M1 roughly agree with those of the three-stage model in ref. [19]. For model M2 a low deviance was achieved, but parameter estimates are notably different for both sexes. For men acceleration in subsequent phases of clonal growth was found but the opposite trend for women is biologically plausible. The initiating rates of the first and second hit were set equal \(v_{10} = v_{12} = v_{1}\) since a pathway-specific treatment was rejected by appropriate LRTs. The rates of early clonal growth \(\gamma_1\) also came out very similar in both pathways and they have been set equal as well. Fitting two path models for both sexes separately produced similar mutation rates (including birth cohort dependences) and rates of early clonal growth. A distinction between sexes was not necessary for these parameters on the basis of LRTs. However, the relatively small difference for the sex-specific rates of clonal growth \(\gamma_{\text{CIN}}\) in late adenoma was highly significant. The deviance was increased by more than hundred points if the growth rates were set equal for both sexes.

From the estimates of identifiable parameters (Table 4) the biological baseline rates of the two path model TP4 can be derived, if assumptions on the total number of susceptible stem cells \(N\), and the rates of symmetric cell division for initiated cells \(\gamma_2\) and for destabilized CIN cells \(\gamma_{\text{CIN}}\) are made. The number of stem cells has been estimated to approx. 10^4 with an accuracy of an order of magnitude [51,52]. MSI tumors appear mainly in the proximal colon so that a (by a factor of 2–3) lower number of susceptible stem cells might be considered for the MSI path [2]. However, the biological parameter \(N\) alone is not identifiable and the uncertainties in the estimates for parameters including \(N\) are too large to prove effects of different values in the LSS data. Thus, the same value for \(N\) has been applied in the MSI and CIN pathways to derive the inactivation rate \(\gamma_1\). Cell division rates of 9 yr\(^{-1}\) in adenoma and 29 yr\(^{-1}\) in early carcinoma have been reported [53]. If these values are assigned to \(\gamma_2\) and \(\gamma_{\text{CIN}}\), rates for the transforming mutations \(\nu_{\text{MSI},i}\), \(\nu_{\text{CIN},i}\), and the cell inactivation rates \(\beta_i\), \(\beta_{\text{CIN}}\) can be calculated. Values for biological baseline parameters, which describe the cell kinetics of the preferred two path model TP4, are summarized in Table 5.

Case shares and radiation risks in molecular pathways

The ability to reproduce the case shares of 15–20% in the MSI pathway and 80–85% in the CIN pathway is an important test for the biological plausibility of the two path model TP4. In Table 6 the computed MSI case shares are listed for the complete follow-up period and for cases recorded before and after 1980. In the early period the shares of MSI cases and CIN cases are about equal. For the later period the predicted MSI share of 17% (men 11%, women 21%) agrees remarkably well with the clinically observed data [4]. For the full period radiation generated 64 (MSI: 10) additional cases in both sexes. For women the values are 19 (MSI: 7) and for men 45 (MSI: 3). Figure 2 shows that especially for women MSI cases appear earlier than CIN cases. Also in good agreement model M3 predicted 22% MSI cases for the full period and 15% after 1980. Whereas in the CIN path models M3 and TP4 exhibit a similar radiation risk, the relative risk in the MSI path is reduced by more than a factor of two for model M3.

Models DERR and DEAR are considered as the quasi-standard for radiation risk assessment. In general, the estimates for the EAR and the ERR are predicted lower by two path model TP4 compared to the descriptive models DERR and DEAR (Figures 3 and 4, Table 7). In the calculation of pathway-specific excess risks only the contribution of a single path is used.
The estimated rate of 0.057 yr^{-1} for clonal growth in early carcinogenesis implies a rate of about one event in 18 years (or a doubling time of 12 years). Crypt fission dominates the growth dynamics in the healthy colon with a similar rate, suggesting that crypt cycle dynamics is reflected in the incidence data as the first unicyryptal LOH in MLH1 has been estimated to 2×10^{-5} yr^{-1} per stem cell from data of HNPCC patients [53]. Hence, the assumption of similar rates for early events in the MSI and CIN pathways appears justified on both biological and statistical grounds.

### Discussion

Biological plausibility of the two path model

Loss of heterozygosity (LOH) in the APC gene and silencing of the MLH1 gene evolve on a similar time scale [3]. The estimate of the initial mutation rate \( \nu_I \) (Table 5) agrees well with a recent estimate of about 10^{-5} yr^{-1} per stem cell for the somatic mutation rate in the APC gene [54], but exceeds older estimates [15,55] by an order of magnitude. Germline mutations (i.e. from \( APC^{+/+} \) to \( APC^{-/-} \) prior to CIN) can occur in both pathways but have not been considered explicitly in the two path model. The rate of unicryptal LOH in MLH1 has been estimated to 2×10^{-5} yr^{-1} per stem cell from data of HNPCC patients [53]. Hence, the assumption of similar rates for early events in the MSI and CIN pathways appears justified on both biological and statistical grounds.

The estimated rate of 0.057 yr^{-1} for clonal growth in early carcinogenesis implies a rate of about one event in 18 years (or a doubling time of 12 years). Crypt fission dominates the growth dynamics in the healthy colon with a similar rate, suggesting that crypt cycle dynamics is reflected in the incidence data as the first

### Table 3. Deviance and \( \Delta AIC \) values for the descriptive and mechanistic models of the present study, \( \Delta AIC \) is defined as the difference in AIC to the preferred two path model TP4.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Model specification</th>
<th>Radiation response parameters</th>
<th>Deviance</th>
<th>( N_{par} )</th>
<th>( \Delta AIC )</th>
</tr>
</thead>
<tbody>
<tr>
<td>DERR(^a)</td>
<td>Descriptive ERR</td>
<td>Sex-specific ERR</td>
<td>4485.5</td>
<td>20</td>
<td>29.4</td>
</tr>
<tr>
<td>DEAR(^a)</td>
<td>Descriptive EAR</td>
<td>Sex-specific EAR</td>
<td>4489.4</td>
<td>20</td>
<td>33.3</td>
</tr>
<tr>
<td>M1(^b)</td>
<td>Two initial mutations, promotion and malignant transformation</td>
<td>Sex-specific, lifelong on promotion</td>
<td>4495.5</td>
<td>7</td>
<td>13.4</td>
</tr>
<tr>
<td>M2(^c)</td>
<td>Early genomic instability</td>
<td>Sex-specific on 2nd destabilizing mutation</td>
<td>4476.0</td>
<td>14</td>
<td>7.9</td>
</tr>
<tr>
<td>M3(^d)</td>
<td>Deterministic MSI path and stochastic CIN path, no early clonal expansion in CIN</td>
<td>Unisex on 2nd hit of initiation ( \nu_C ) for men only on CIN promotion ( \lambda_C ) (like TP4)</td>
<td>4479.0</td>
<td>10</td>
<td>2.9</td>
</tr>
<tr>
<td>TP0</td>
<td>Two path</td>
<td>No radiation effect</td>
<td>4507.9</td>
<td>9</td>
<td>29.8</td>
</tr>
<tr>
<td>TP1</td>
<td>Two path</td>
<td>Unisex on 2nd hit of initiation ( \nu_I )</td>
<td>4481.9</td>
<td>10</td>
<td>5.8</td>
</tr>
<tr>
<td>TP2</td>
<td>Two path</td>
<td>Unisex on 2nd hit of initiation ( \nu_C ) unisex on CIN promotion ( \lambda_C )</td>
<td>4477.1</td>
<td>11</td>
<td>3.0</td>
</tr>
<tr>
<td>TP3</td>
<td>Two path</td>
<td>Unisex on 2nd hit of initiation, for men only on destabilizing CIN event ( \lambda_C )</td>
<td>4476.0</td>
<td>11</td>
<td>1.9</td>
</tr>
<tr>
<td>TP4(^e)</td>
<td>Two path</td>
<td>Unisex on 2nd hit of initiation ( \nu_C ) for men only on CIN promotion ( \lambda_C ) (like M3)</td>
<td>4474.1</td>
<td>11</td>
<td>0.0</td>
</tr>
</tbody>
</table>

\(^a\)from Preston et al. [28].

\(^b\)baseline model equivalent to the three-stage model of Meza et al. [19], see SI, Table S2 and Figure S1 in File S1.

\(^c\)Inspired by Nowak et al. [16], derived from Little and Li [17] (their Figure 2), see SI, Table S3 and Figure S2 in File S1.

\(^d\)from Meza et al. [19], see SI, Table S2 and Figure S1 in File S1.

\(^e\)preferred model of the present study.

doi:10.1371/journal.pone.0111024.t003

### Table 4. MLE, SE from a parabolic approximation around the minimum of the likelihood function, and \( \Delta CLP \) from the actual likelihood profile in the standard \( \sigma \) range for the identifiable parameters of the two path model TP4 with relation to biological parameters, superscript m,f indicates sex-dependence, radiation-response parameters \( r_I \) and \( r^{m, CIN} \) on dose \( D \) are given for an exposure duration of 1 week, rate \( \lambda_C \) of destabilizing events in CIN increases exponentially with birth year \( b = 1945.6–e \) (age at exposure).

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Unit</th>
<th>MLE</th>
<th>SE</th>
<th>( \Delta CLP )</th>
<th>Relation to biological parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>( R_{MSI,0} )</td>
<td>yr^{-1}</td>
<td>-17.56(^a)</td>
<td>0.33</td>
<td>-0.31; 0.30</td>
<td>( \nu_Y ) ( Y_{MSI} ) r(0)</td>
</tr>
<tr>
<td>( r_I )</td>
<td>week Gy^{-1}</td>
<td>6.89(^a)</td>
<td>0.33</td>
<td>-0.38; 0.30</td>
<td>( r(D) = 1+\nu_I ) D/week</td>
</tr>
<tr>
<td>( \nu_I )</td>
<td>yr^{-1}</td>
<td>0.0568</td>
<td>0.013</td>
<td>-0.012; 0.012</td>
<td>( \nu_C ) ( Y_{MSI} )</td>
</tr>
<tr>
<td>( \delta_I )</td>
<td>yr^{-2}</td>
<td>-11.1(^a)</td>
<td>1.1</td>
<td>-1.0; 0.9</td>
<td>( \nu_C ) ( Y_{MSI} )</td>
</tr>
<tr>
<td>( R_{CIN,0} )</td>
<td>yr^{-1}</td>
<td>-26.6(^a)</td>
<td>2.0</td>
<td>not calc.</td>
<td>( \nu_Y ) ( Y_{CIN} ) (1915.6) ( \lambda_C ) r(0)</td>
</tr>
<tr>
<td>( k_0 )</td>
<td>yr^{-1}</td>
<td>-0.109</td>
<td>0.023</td>
<td>-0.021; 0.015</td>
<td>( \lambda_C ) ( (b) ) = ( \lambda_C ) (1915.6) exp[( (1915.6-b) )]</td>
</tr>
<tr>
<td>( \gamma_C )</td>
<td>yr^{-1}</td>
<td>0.234</td>
<td>0.043</td>
<td>-0.022; 0.026</td>
<td>( \lambda_C ) = ( \lambda_C ) - ( Y_{CIN} )</td>
</tr>
<tr>
<td>( r_C )</td>
<td>yr^{-1}</td>
<td>0.266</td>
<td>0.040</td>
<td>-0.020; 0.026</td>
<td>( \lambda_C ) = ( \lambda_C ) - ( Y_{CIN} ) g(0) - ( Y_{CIN} )</td>
</tr>
<tr>
<td>( f^{m, CIN} )</td>
<td>week Gy^{-1}</td>
<td>5.68(^a)</td>
<td>0.34</td>
<td>-0.40; 0.30</td>
<td>( g(D) = 1+\nu_C ) D/week</td>
</tr>
<tr>
<td>( \nu_C )</td>
<td>yr^{-2}</td>
<td>-16.9(^a)</td>
<td>2.4</td>
<td>-2.0; 1.4</td>
<td>( \nu_C ) ( Y_{CIN} )</td>
</tr>
</tbody>
</table>

\(^a\)log-transformed.

doi:10.1371/journal.pone.0111024.t004
round of clonal expansion [37]. Crypt fission is not influenced by carcinogenesis but distributes inactivated TSGs in the colon [39]. Therefore, the same rate of early clonal expansion must pertain to the MSI and CIN pathways. The biological argument is supported by a statistical criterion. Allowing different rates of early clonal expansion in both pathways did not improve the fit significantly.

The mean sojourn time from the birth of a non-extinct premalignant clone to the appearance of at least one malignant cell is given by \( T_I = -\ln(b_1/\gamma_1) / \gamma_1 \) in the MSI pathway [19]. Inserting the MLEs of Table 4 for the identifiable parameters \( \delta_1 \) and \( \gamma_1 \) yields \( T_I = 94 \) yr which exceeds mean human lifetime. The growth dynamics of clones in crypt fission is too slow to produce tumors of clinically relevant size. This conflict of time scales demands involvement of at least one more round of clonal expansion in cancer induction [56]. With MSCE models it has been shown that the introduction of a lag time substitutes the explicit modelling of tumor growth with a second round of clonal expansion in proper approximation [20]. The corresponding lag time has been estimated to 5–6 yr in the SEER cohort. A study of interval cancers, which develop in the time interval between serial colonoscopy, showed that MSI cancers are found four times more likely among interval cancers than among non-interval cancers [57]. These observations suggest rapid growth of malignant MSI clones after a slow development of benign adenoma. Therefore, a shorter lag time of about 3 yr should be expected in the MSI model. But models fit showed almost no dependence on lag time in the LSS cohort so that a fixed value of 3 yr was kept for both pathways.

The ratio of proliferating cells to apoptotic cells is smaller in early adenoma compared to late adenoma with high-grade dysplasia [58]. Other studies report an increase of both apoptosis and proliferation in colonic neoplasms which results in enhanced turnover rates [59]. These observations are compatible with higher estimates of about 0.23 yr\(^{-1}\) (women) and 0.27 yr\(^{-1}\) (men) for clonal growth in adenoma with CIN cells compared to model M3 with just one phase of clonal growth. The small sex difference in these promotion rates is highly significant, possibly due to different causes of the disease in men and women [12]. But also the altered oestrogen status after menopause might explain slower growth of adenoma which appear in women later in life [60]. A similar hormonal effect has been observed for neoplastic lesions which precede breast cancer in female a-bomb survivors [61]. Note, that the difference in a single biological model parameter fully explains the difference in the baseline incidence of men and women.

The larger number of cases in the CIN pathway allows a better resolution of the tumorigenic processes so that a second phase of clonal expansion in late adenoma could be detected. The sojourn time from the development of a small adenoma (containing CIN properties and surviving extinction) to the birth of the first malignant cell is expressed by \( T_{CIN} = -\ln(\delta_{CIN}/\gamma_{CIN}) / \gamma_{CIN} \) [20]. With the sex-specific MLEs from Table 4 sojourn times of \( T_{CIN} = 52 \) yr (men) and 58 yr (women) are calculated. In the SEER cohort a similar sojourn time \( T_I \) (in Table 1 of ref. [20]) of 51 yr has been found for men, but for women the value of 49 yr is lower than in the LSS cohort.

The last event before the appearance of the first malignant cells is often associated with inactivation of gene TP53 [1]. It has been suggested that this event is very rare, so that the estimated small value of \( 1.5 \times 10^{-5} \) yr\(^{-1}\) for the transforming mutation rate \( \gamma_{CIN} \) could pertain to inactivation of TP53 (or of another gene downstream in the CIN pathway) [15].

### Lifestyle trends

In many mechanistic models of the SEER cohort the hazard functions have been adjusted descriptively for secular trends in calendar year or birth year [15,17,19–21,62]. A statistical interpretation of this adjustment is based on the separation of secular trends from the “natural” carcinogenic mechanism. There is, however, a biological interpretation of this correction which often (depending on the mathematical model structure) amounts to

<table>
<thead>
<tr>
<th>Table 5. Estimates for baseline rates (unit yr(^{-1}) per cell) of cell-kinetic processes in the two path model TP4.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cell-kinetic process</strong></td>
</tr>
<tr>
<td>APC mutation or MLH1 hypermethylation (1(^{st}) and 2(^{nd}) hit)</td>
</tr>
<tr>
<td>Cell inactivation in early adenoma (crypt cycle)</td>
</tr>
<tr>
<td>Transforming mutation in MSI</td>
</tr>
<tr>
<td>Rate of destabilizing events in CIN (chromosomal gain or loss) for birth years ( b = 1895, 1915, 1935 )</td>
</tr>
<tr>
<td>Cell inactivation in adenoma with CIN cells (male)</td>
</tr>
<tr>
<td>Cell inactivation in adenoma with CIN cells (female)</td>
</tr>
<tr>
<td>Transforming mutation in CIN</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Follow-up period</th>
<th><strong>Men</strong></th>
<th><strong>Women</strong></th>
<th><strong>Both sexes</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total cases</td>
<td>MSI share [%]</td>
<td>Total cases</td>
</tr>
<tr>
<td>1958–1998</td>
<td>688</td>
<td>18</td>
<td>820</td>
</tr>
<tr>
<td>1958–1980</td>
<td>161</td>
<td>41</td>
<td>206</td>
</tr>
</tbody>
</table>

[10.1371/journal.pone.0111024.t005](https://doi.org/10.1371/journal.pone.0111024.t005)

[10.1371/journal.pone.0111024.t006](https://doi.org/10.1371/journal.pone.0111024.t006)
assuming an impact on early TSG mutations. Since lifestyle-dependent health risks appear later in life, a biological effect on later carcinogenic events has also been tested in the present study. Firstly, an exponential dependence on birth year was applied to the initiating mutations or to the MSI mutation in the two path model. This adjustment did not improve the fit significantly (i.e. by more than 3.8 deviance points). However, the rate of CIN of the destabilizing CIN event (related to chromosomal gain or loss) grew exponentially in subsequent birth cohorts, the deviance fell by some hundred points. Rising red meat intake in a westernized diet may have increased the risk of colon cancer in Japan after the Second World War [13]. In European patients red meat intake was inversely associated with high-level MSI tumors when compared to population-based controls and positively associated with low-level MSI/CIN tumors [63]. Feeding a western diet of high fat content and low levels of calcium and vitamin D to APC knock-out mice resulted in a higher risk of cancer in the CIN pathway [64]. If westernized diet affected molecular sub-types similarly in Japan, it could explain the pronounced increase of CIN cases after 1980 (Table 6).

Hallmarks of CIN in the two path model

The rate of early clonal expansion in the two path model occurs on a time scale which is reported for crypt fission [37]. This process is capable of spreading inactivated TSGs in the normal human colon and of amplifying their number [39]. If the number of cells with mutated APC genes increases in the CIN path, WNT signalling can no longer be suppressed effectively [65]. Up-regulation of this signalling network may precede or even cause the generation of cells carrying a CIN property. Chromosomal gain or loss compared to the normal karyotype has been identified as a prominent property of CIN cells, and the rate of gain or loss of a single chromosome has been measured to one per five stem cell divisions in vitro [35]. Although a comparison with in vitro data must be applied with caution, this measured value falls in the range of large estimates for the rate of CIN of destabilizing events in the two path model (Table 5). Other modelling studies also find a markedly enhanced “fast” third event with a four-stage model of colon carcinogenesis [15,19]. Although the four-stage model does not consider crypt fission explicitly, the high rate of asymmetric stem cell divisions has been interpreted as the amplification of TSG-/-/- cells in a crypt. CIN occurs as the consequential effect of this amplification. The present results support this interpretation and add the consideration of crypt fission in the two path model as another piece to the puzzle. But they are at variance with suggestions that the CIN property is conferred early to cells with inactivated APC genes [16].

Radiation risk

Estimates of the radiation response parameters $r_{1} = 19 \ (\sigma_{CLP} \ 13; 26) \ yr \ Gy^{-1}$ for both sexes in initiation and $r_{2} = 5.6 \ (\sigma_{CLP} \ 3.8; 7.6) \ yr \ Gy^{-1}$ for men in clonal growth of CIN cells are close to estimates of a simpler mechanistic model for a joint set of nine cancer sites including colon [26]. Crude data of the LSS cohort (Tables 1 and 2) and estimates from standard descriptive models suggest a markedly lower radiation risk for women compared to men [28]. In the preferred two path model TP4 this fact is explained by a negligible radiosensitivity late in the female CIN path. Independent biological evidence for this conjecture is not known to the authors of the present study. Based on statistical criteria, a radiation response on the destabilizing event in CIN for men or a unisex radiation response in the two path model cannot be ruled out (Table 3). Probably radiation affects more if not all stages of carcinogenesis but a thorough analysis of radiation targets is beyond the scope here. If models with different radiation targets describe the data almost equally well, an approach of multi-model inference, which combines all plausible models for risk estimation, might be applied [61].

Observance of pathway-specific risks has implications for risk assessment. In combination with molecular sub-type ascertainment, consideration of pathway-specific risks will improve the
accuracy of expert opinion in compensation claims [32]. If the status of MSI or CIN were testable in adenoma, the application of diagnostic (or therapeutic) radiation could be optimized with a more targeted risk-benefit analysis [33].

**Limitations of the two path model**

Some 200 genes are mutated in colorectal cancer [8,66]. Eleven mutations have been counted on average per tumor but only a few mutations are common to most tumors [2]. The two path model cannot consider the effects of this complex genomic structure in detail. A one-to-one association of model parameters to mutations in specific genes such as \(APC\) or to bi-allelic silencing of \(MLH1\) should be regarded as tentative. Since a model (M3) without early clonal expansion in the CIN path yielded only a slightly inferior description of the data, the interpretation of early clonal expansion as crypt fission should also be regarded as tentative. Mutation rates of TSGs and the rates of clonal growth during crypt fission and or in premalignant adenoma should be understood as effective phenomenological parameters.

Up to now, simplifications of tumorigenic processes are inevitable as a consequence of general issues with parameter identifiability, limited statistical power and biological insight. But remarkable agreement with molecular data for a number of processes, which have been reported in the literature over the last fifteen years, emphasizes the biological plausibility of the two path model.

**Supporting Information**

**File S1** Contains the following files: **Table S1.** Identifiable baseline parameters in the deterministic and stochastic versions of the two path model with MSI and CIN paths (Figure 1), biological parameters \(N, \nu_1, \nu_2\) and \(\gamma_1\) are set equal in both pathways. **Table S2.** MLE, SE from a parabolic approximation around the minimum of the likelihood function, and \(\Delta CLF_L\) from the actual likelihood profile in the standard \(\sigma\) range for the identifiable parameters of model M1 (Figure S1) with relation to biological parameters, superscript \(m,f\) indicates sex-dependence, radiation-

---

Figure 3. MLE with 95% CI of the excess absolute risk (EAR) per \(10^4\) PY and of the excess relative risk (ERR) for women (panels A, C) and men (panels B, D), exposed to 1 Gy at age 30 (born in 1915) for the descriptive models DERR and DEAR [28] (black), and the two path model TP4 (red). Only MLE are shown for pathway-specific excess risks pertaining to MSI (blue) and CIN (green). doi:10.1371/journal.pone.0111024.g003
response parameters \( r_{i}^{m,f} \) on dose \( D \) are given for lifelong radiation effect on clonal expansion, one initial mutation rate \( v_{i} \) increases exponentially with birth year \( b = 1945.6 - e \), the baseline version of model M1 is mathematically equivalent to the three stage model by Meza et al. [1]. Table S3. MLE and SE from a parabolic approximation around the minimum of the likelihood function for the identifiable parameters of model M2 (Figure S2) with relation to biological parameters, superscript m,f indicates sex-dependence,

**Table 7.** MLE (95% CI in brackets) of the excess relative risk (ERR) and the excess absolute risk (EAR) per \( 10^4 \) PY for persons of attained age 70, exposed to 1 Gy at age 30 (born in 1915) from the descriptive models DEAR and DERR [28] and the two path model TP4.

<table>
<thead>
<tr>
<th>Model</th>
<th>ERR (95% CI)</th>
<th>EAR per ( 10^4 ) PY (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Men</td>
<td>Women</td>
</tr>
<tr>
<td>Descriptive</td>
<td>0.77 (0.29; 1.4)</td>
<td>0.33 (0.066; 0.70)</td>
</tr>
<tr>
<td>Two path</td>
<td>0.54 (0.32; 0.95)</td>
<td>0.13 (0.073; 0.35)</td>
</tr>
<tr>
<td>MSI path</td>
<td>0.39 (0.21; 0.68)</td>
<td>0.94 (0.42; 1.7)</td>
</tr>
<tr>
<td>CIN path</td>
<td>0.56 (0.31; 1.1)</td>
<td>0.047 (0.0095; 0.27)</td>
</tr>
</tbody>
</table>

doi:10.1371/journal.pone.0111024.t007
radiation-response parameters r on dose D given for direct radiation effect on mutation rate \( \lambda_1 \) with exposure duration of 1 week, mutation rates \( \nu_{0a}, \lambda_0 \) and \( \nu_{0b} \) increase exponentially with birth year \( b = 1945.6 - r \), \( \Delta C I P \) from the actual likelihood profile in the standard \( \sigma \) range could not be computed by MINUTT, the harmonic mean of the \( LMLE \) is inspired by Nowak et al. [2] and derived from Little and Li [3] (their Figure 2). Table S4. MLE, SE from a parabolic approximation around the minimum of the likelihood function, and \( \Delta C I P \) from the actual likelihood profile in the standard \( \sigma \) range for the identifiable parameters of model M3 (Figure S3) with relation to biological parameters, deterministic MSI model without dependence on \( \delta_{MSI} \), superscript \( m,f \) indicates sex-dependence, radiation-response parameters \( r_1 \) and \( \gamma_{CIN} \) on dose D are given for an exposure duration of 1 week, one initial mutation rate \( \nu_{CIN} \) increases exponentially with birth year \( b = 1945.6 - r \) (age at exposure). Figure S1. Parametrisation of model M1 with lifelong radiation action (jagged bolt) on cell inactivation \( \beta \) (TSG: tumor suppressor gene), one initial mutation rate \( \nu_1 \) increases exponentially with birth year b. Figure S2. Parametrisation of model M2 with radiation action (jagged bolt) on mutation rate \( \lambda_1 \), mutation rates \( \nu_{0a}, \lambda_0 \) and \( \nu_{0b} \) increase exponentially with birth year b. Figure S3. Parametrisation of model M3 with deterministic MSI path (no dependence on \( \delta_{MSI} \)) and stochastic CIN path, radiation action (jagged bolt) on second initial mutation rate \( \nu_I \) equal in MSI and CIN paths and on cell inactivation \( \nu_{CIN} \) for men only, one initial mutation rate \( \nu_{CIN} \) increases exponentially with birth year b. (DOCX)

Acknowledgments

This report makes use of data obtained from the Radiation Effects Research Foundation (RERF) in Hiroshima and Nagasaki, Japan. RERF is a private, non-profit foundation funded by the Japanese Ministry of Health, Labour and Welfare (MHLW) and the U.S. Department of Energy (DOE), the latter through the National Academy of Sciences. The data include information obtained from the Hiroshima City, Hiroshima Prefecture, Nagasaki City, and Nagasaki Prefecture Tumor Registries and the Hiroshima and Nagasaki Tissue Registries. The conclusions in this report are those of the authors and do not necessarily reflect the scientific judgment of RERF or its funding agencies.

The authors have benefited from discussions with D. J. Brenner, L. Sahatie, M. Eidenmüller, who are members of the EpiRadBle consortium or advisory board, and with C. Staudt.

Author Contributions

Conceived and designed the experiments: PJ JCK. Performed the experiments: RM JCK. Analyzed the data: JCK. Contributed reagents/materials/analysis tools: RM JCK. Wrote the paper: JCK.

References