A New Stochastic Model of Carcinogenesis Induced by Ionizing Radiation and the Concept of Breaking Barrier Cell Mechanisms

Igor Akushevicha, Galina Veremeyevab, Svetlana Ukraintseva, Konstantin Arbeevea, Alexander Akleyevb, and Anatoli Yashinb

aCenter for Population Health and Aging, Duke University, Durham 27708-0904, North Carolina, USA
bUral Research Center for Radiation Medicine, 68-A Vorovsky Street, Chelyabinsk 454076, Russia.

Abstract. In this report we present a new multiple-pathway dynamic stochastic model of leukemia development, which combines and further extends ideas used in: i) models of hematopoietic system; ii) multistage mechanistic models of carcinogenesis and leukemia, and iii) population models of aging, cancer morbidity, and mortality. The model represents three levels of the organism’s vital organization: i) a cellular level, where dynamics of the processes of cell kinetics, repair, and apoptosis, are defined, ii) a level of human organism, where covariates describing health states are measured, and iii) a population, where such characteristics as incidence and mortality rates associated with cancer are predicted. The new approach considers breaking the barrier mechanisms of a cell as key feature of carcinogenesis. The barrier mechanisms (e.g., antioxidant defense, repair, apoptosis) represent the complex of cell responses to primary cell damages caused by exogenous and endogenous factors. Detrimental phenotypic changes in the cells result from insufficiency of respective barrier mechanisms. The advantage of suggested modeling approach is in the possibility of natural combining of different measurements including age-specific hazard rate and measures, characterizing fractions of cells with breaking barriers. Another advantage is in the opportunity to study effects of protracted low-dose Ionizing Radiation (IR) where such barrier mechanisms play an important role. Simulation studies show that the model parameters can be identified from joint analyses of epidemiological data and the results of individual biomolecular measurements of barrier states. Application of the results to the analyses of SEER data on leukemia risks demonstrates the strong predictive power of the model. Further generalizations of the model, possible applications to available data on subjects chronically exposed to IR, and incorporating IR-induced genomic instability are discussed.

KEYWORDS mechanisms of carcinogenesis, multi-stage models, cell barrier mechanisms, stochastic modeling, leukemia risk, ionizing radiation

1. Introduction

There exist many situations in biomedical research when mathematical models constructed on the basis of biological theories and concepts are more preferable than standard empirical analyses. Studying carcinogenesis is one of them.

Carcinogenesis modeling has a long history that started with papers in the fifties of the last century [1,2]. These models were suggested to explain the observation that age-specific rates of many common carcinomas increased roughly with a power of age. Then many interesting biological ideas and ingenious mathematical methods were developed to understand spontaneous and radiation carcinogenesis. The set of these models is the first methodological background for our approach. Consequently, Section 2 is devoted to a brief review of these models focusing on those features which will be compared with the model developed in this paper.

The further progress in modeling carcinogenesis can be achieved by the development of a new formalism capable of including additional information, e.g., measurements of respective risk factors at an individual level or auxiliary information coming from other sources. Models with required properties are known in demography and population studies as stochastic process models [3,4]. Such models allow for inclusion of measurements of covariates (risk factors) and combined description of their dynamics and survival. Besides, these models have very useful and well-established mathematical properties. Such population models constitute the second part of the methodological background of our approach and, therefore, are also reviewed in Section 2. The linkage between population and mechanistic (i.e., models of carcinogenesis discussed above) formalisms is provided by the theorem proved by Tan [5], which explains the connection between them.
Biological background of the approach to carcinogenesis modeling is based on the consideration of carcinogenesis as the dynamic trade-off between two antagonistic forces or processes, promoting or hindering carcinogenesis at its different stages (initiation, promotion, conversion). Processes promoting the cell malignization are represented by mutations or adverse epigenetic events. Antagonistic processes preventing the neoplastic transformation of the cell and forthcoming its fixation in next cell generations are represented by barrier mechanisms. We present the barrier breaking mechanism (BBM) model in Section 3. The discussion in this section includes description of the concept of barrier mechanism, its application to the SEER data [6] on leukemia risks, a possibility to measure states of the barrier mechanisms and analyses of possible biases in such measurements, results of simulation studies Discussion of different aspects of the models is presented in Section 4.

2. Background Methodologies
The current state of the art of modeling efforts in tumorigenesis relies on a multi-stage hypothesis that is implemented in multi-stage models of carcinogenesis. Usually three stages of the process of tumor development are considered: i) formation of initiated cells, ii) promotion and neoplastic conversion of initiated cells resulting in appearance of the first malignant clonogenic cell, and iii) subsequent growth and progression of a malignant tumor. The duration of each stage of carcinogenesis (initiation, promotion, conversion) is thought of as a random variable. The underlying idea of these models is the concept of sequential, interacting gene mutations as the driving force of tumorigenesis.

From the mathematical point of view, the problem is to predict a hazard function (e.g., incidence, mortality, survival, etc.) in terms of age and biologically interpretable parameters. The sequence of steps of mathematical modeling is typical for a majority of the developed models. They include i) defining cell processes to be taken into account, e.g., specifying a compartment structure, assumptions on proliferation, apoptosis, differentiation, and mutation rates; ii) defining whether each process occurs stochastically or deterministically; iii) writing corresponding ordinary, stochastic, or partial differential equations and iv) solving these equations and using the solution in the form of deterministic function, stochastic process, or p.g.f. to predict a hazard function.

Mechanistic Models of Carcinogenesis. Models which use the mathematical formalism realizing one of specific scheme of the multi-stage concept of carcinogenesis are known as mechanistic carcinogenesis models.

Historically, the first carcinogenesis models were multi-stage models of Nordling [1] and Armitage-Doll [2]. Their assumption that cancer is the result of the accumulation of a critical number of mutations (Figure 1), in which the order of mutation can be important (Armitage-Doll) or not (Nordling), allowed them to explain the observation that age-specific rates of many common carcinomas increased roughly with a power of age. In both approaches, the incidence function is approximately equal

\[ I(t) = c t^k \]

where \( c = N \mu_0 \mu_1 \ldots \mu_k \) for the Armitage-Doll model and \( c = (k+1) N \mu_0 \mu_1 \ldots \mu_k \) for the Nordling model; \( \mu_i \)’s are time independent mutation rates, \( k \) is the number of stages, and \( N \) is the total number of susceptible cells.

The Armitage-Doll model predicts 4-7 stages, however there is little evidence that there are that many stages [7]. Therefore, further development to reduce the biologically implausible number of stages was required. A series of generalizations [8-11 result in a two-stage model known as the Moolgavkar-Venzon-Knudson (MVK) model. The scheme of this model is presented in Figure 2. In this Figure, \( N \), \( I \), and \( T \) denote normal, intermediate, and malignant cells, respectively. “Stage I” describes initiation.
“Stage II” describes promotion and conversion. This model became a milestone in the history of carcinogenesis modeling and served as the basis for further successful data analyses and specific generalizations [12,13].

The most popular version of the two-stage model is the Two Stage Clonal Expansion (TSCE) model which additionally assumes that i) the number of susceptible normal cells is either constant or described by a deterministic function and ii) all rates are time independent (Figure 3). An attractive property of the model is that the spontaneous hazard rate can be expressed analytically in terms of only three parameters [14]: The main disadvantage is that not all biological parameters (i.e., the number of stem cells \( N \), first \( \mu_1 \) and second \( \mu_2 \) mutation rates, and proliferation \( \alpha \) and death/differentiation \( \beta \) rates) can be identified using the data on age-specific incidence rates.

Important generalization of the Armitage-Doll and two-stage models to allow for an arbitrary number of mutation stages was performed by Little [15,16]. Corresponding multistage models (Figure 4) were successfully applied in different data analyses. An analytical solution for the hazard function is no longer possible. Instead the hazard function is expressed as an integral on function dependent on p.g.f. which satisfies the Kolmogorov backward equation. This equation is solved numerically. Although the computation becomes much more complicated, such approach allowed for many generalization and specific applications. For example, Little and Wright [17] developed the stochastic carcinogenesis model incorporating genomic instability by assuming that cells can acquire two sorts of mutation, those associated with progression to a malignant phenotype, and those associated with successive destabilization of the genome.

To represent accumulation of the two types of mutations, the diagram of the model becomes two-dimensional.

Klebanov, Rachev, and Yakovlev [18] developed a model for radiation carcinogenesis based on the following assumptions. First, the number of lesions formed by IR (accumulated by time \( T \)) is the Poisson variable with expectation \( \gamma T \), where \( \gamma \) is related to IR dose. Second, lesions are subject to the repair process. The repair system is modeled by the methods of the queueing theory. The probability for a lesion not to be served is modeled as the well-known probability of "losing a customer." Third, each promoted lesion ultimately gives rise to an overt tumor after a certain period of time. Yakovlev and Polig [19] generalized this model by an incorporation of radiation induced cell death. In this model, the hazard is quite a complex function of dose rate and the p.d.f. of promotion time distribution. This model, when promotion time is modeled by \( \gamma \)-distribution, and TSCE were
compared and tested with several sets of epidemiological data [20]. No conclusion about the better model was made.

Tan [21] has developed a number of generalizations of two-mutation models of MVK and has documented biological evidence and mathematical formalism. Since most of these models are far more complicated than the scope of the MVK two-stage model, Tan and Chen [22] developed state space models (Kalman filter models) for carcinogenesis using formalism of stochastic differential equations. They also demonstrated how their formalism is related to classical formalism based on the p.g.f. Tan, Zhang, and Chen [23] developed advanced statistical procedures to estimate the unknown parameters of the state space model via the multi-level Gibbs sampling method (i.e., using the Markov Chain Monte Carlo method, MCMC) and applied these procedures to the British physician data on lung cancer due to smoking. In many cases, it was observed that the same cancer may arise from different carcinogenic processes. Careful examination of the literature on cancer biology would reveal that multiple pathways for cancer may be quite common in the real life [21]. More recent biological evidence was documented by Tan and Chen [22]. Figure 5 illustrates a model involving a one-stage model, three two-stage models, and two three-stage models. Roughly, the hazard rate is the sum of all possible pathways. For simplest paths only (e.g., for one stage path), the hazard rate can be calculated analytically.

![Figure 5. Multiple Pathways model](image)

A description of IR induced carcinogenesis is a specific and important subtask of carcinogenesis modeling. IR can induce specific mutations or epigenetic events in cells and therefore increase the number of intermediate cells susceptible to further stages of carcinogenesis. IR can also have promoting effect to carcinogenesis. The basic argument is that stem cells inactivated by IR may be replaced by the division of intermediate stem cells in which cells have a growth advantage [24]. A typical way to incorporate these effects into mechanistic models is to assume that rates of initiation, promotion, and conversion become dose-dependent. Recently, such effects were analyzed and discussed for radon-induced lung cancer in Colorado Plateau uranium miners [25] and French and Czech miner cohorts [26, 27].

Even though TSCE is definitely one of the most popular models of IR induced carcinogenesis, it still has limitations. The first is the problem with parameter identifiability. Only three combinations of biological parameters are identifiable from the age-specific hazard function. The second is that the biological mechanisms represented by TSCE are oversimplified. The next limitation is that the parameters used in this model (as well as in the more general ones) cannot be directly measured that restricts the capability of predicting individualized risks. A possible solution is to combine data on the age-specific hazard function with additional measurements indirectly related to the model parameters, e.g., to measure apoptosis rate.

**Population models.** Formalism capable of overcoming the limitations of the mechanistic models can be developed by appealing methodological ideas of a series of population models and, first of all, the stochastic process models. Many ideas of modern population models were developed by generalization of classical population models. Review of these models including first classical population models (such as Gompertz, Strehler and Mildvan, Sacher and Trucco approaches) and recent developments (such as frailty models, correlated frailty models, debilitation models, repair capacity models) was presented by Yashin et al. [28]. The Quadratic Hazard Model was introduced [29] to perform the joint description of regularly measured physiological risk factors and information on survival status of individuals. In this model, the health state is described by a set of risk factors or covariates. The hazard function is modeled as a quadratic function of covariates that is phenomenologically justified. Mathematical formalism is based on stochastic differential equations or the Kolmogorov-Fokker-Planck equation. Recent generalizations [30,31] allow for inclusion of the
effects of “optimal” (normal) physiological states and allostatic load. An important generalization of the quadratic hazard model is the stochastic process model (SPM) [4]. Since the time between surveys may be flexible and the model automatically generates values of risk factors to fill missing data, the SPM is the appropriate model for analysis of longitudinal data with irregular measurements. The following feature of the model [3,4] will be crucial for further development of carcinogenesis model based on ideas of breaking barrier mechanisms:

- Dynamics of covariates are described by the Gaussian stochastic process, characteristics of which (i.e., the vector of means and the variance-covariance matrix) are defined by a system of ordinary differential equations solvable analytically or numerically.
- Covariates are naturally incorporated into the model in addition to the information on individual survival; adding new covariate does not change the structure of the model.
- The distribution of covariates conditional on survival is also normal if dynamic stochastic differential equation is linear over covariates and mortality is a quadratic (or linear) function of covariates.
- There exist versions of the model for different experimental designs, e.g., i) when only time of event is measured, ii) when covariates for individuals are measured, iii) when covariate dynamics is assigned but they are not measured or partially measured, and iv) whole trajectories of covariates are observed.
- There exists an exact procedure of parameter estimation for distinct experimental designs. Parameter estimates are consistent and identifiable under mild conditions.
- There are straightforward possibilities to include characteristics of exposure (e.g., dose or dose rate) into the model [32].

Two other useful classes of models are dynamic microsimulation and latent structure models. The first allows for simulating individual age trajectories, to construct projections for various “what-if” scenarios (e.g., [33]) and has the potential to investigate medical interventions [34]. The second allows us to deal with data presented in the form of multiple categorical measurements. The most known examples of such models are the Latent Class Models, Grade of Membership, and the recently developed Linear Latent Structure Analysis [35,36]. All of the above mentioned, and several other, models were recently reviewed by Akushevich et al. [32].

**Connecting two methodologies.** Tan [5] has shown that, under some mild conditions, the number of initiated cells in an extended two-stage model of carcinogenesis can be approximated by a diffusion process with accuracy to the order of \( O(N^{-2}) \) where \( N \) is the number of susceptible cells \( (N=10^6-10^9) \). This result allows us to bridge the two above formalisms for the tasks of carcinogenesis modeling. Specifically, the exact formalism of classical carcinogenesis models can be approximated with the good accuracy by the multivariate normal stochastic process, which is an underlying assumption in population models.

3. **Breaking Barrier Mechanisms (BBM) Model**

The concepts of barrier mechanisms as a basis for the approach to the mathematical model of carcinogenesis are formulated on the basis of numerous observations collected in modern science on intracellular processes occurring in the norm and in the pathology including malignant tumors. Specifically, this concept is based on the following principles. The carcinogenesis represents a set of structural and functional changes in susceptible cells ultimately expressed at a higher level of hierarchy in tissues of a human body. Detrimental changes at the cell level are caused by an insufficient quality of operation of a complex of mutually interacting barrier mechanisms (e.g., AOD, repair systems, apoptosis). The intercellular barrier mechanisms represent the complex of cell responses to similar, negative for the cell and/or the whole organism, events. They are combined in the system of the cell defense that protects from occurrence of genetic damages and their further fixation as mutations potentially promoting the cell to carcinogenesis. Cell malignization may occur due to an inefficient operation of a part of, or all, barrier mechanisms. Hierarchy and complex interaction of barrier mechanisms can compensate the inefficiency of operation of certain mechanisms by reinforcing others, and as a result decrease the risk of pathology development. Disorders in the barrier mechanism functioning occur as a result of one or more mutations/aberrations or adverse epigenetic
events, and there exists a conceptual possibility to measure the efficiency of each barrier functioning in sensitive cells of a certain individual.

Adopting the understanding of the barrier mechanisms as dynamic processes hindering carcinogenesis at its different stages (initiation, promotion, conversion), we realize that in each cell there simultaneously exist processes promoting the cell malignization (e.g., mutations, epigenetic events) and antagonistic processes preventing the neoplastic transformation of the cell. What was considered in traditional multi-stage carcinogenesis models as mutations transferring the cell over stages of initiation, promotion, and conversion, can be considered as results of inefficiency of cell barrier mechanisms within the concept of BBM. A critical conjunction of failure in several mechanisms (e.g., combinations of inefficient repair, violations in apoptosis, activation of telomerase resulting in immortality of the cell, acquired hypersensitivity to growth signals [37]) and further appearance of two daughter cells can be considered as its conversion to cancer phenotype.

The transfer of a cell from a state to a state is understood as a failure, or breaking, of a certain barrier mechanism in a given cell. Such a transfer can occur due to a mutation as in the standard multistage model, due to several mutations, due to adverse epigenetic events or their combinations. A distinguished property of the approach based on this concept is that variables describing the cell state are measurable at the individual level.

In the considered variant of the model we include the following barriers: i) removal of superfluous free radicals which induce damages in bio-molecules, ii) repair of damages occurred through the free radical mechanism or due to other causes, and iii) apoptosis (in the case if repair is not effective and/or the damage is not compatible with vital cell functions). The scheme of a compartmental model implementing the sequence of barrier breaking when a cell undergoes changes from a normal (N or ABC) to a malignant (M) state, is presented in Figure 6. Different blocks correspond to different cell states, and each cell can be in one and only one state. It is assumed in the model that a cell becomes malignant if barrier mechanisms responsible for apoptosis (A) and repair (B) are broken. Other barriers (AOD) play a role of promoters of the process. Letters on the blocks denote which barriers are effective in a certain state. Transfers from a state to a state occur with rates marked by $\nu$ or $\mu$ at the corresponding arrow. Subscript at $\nu$ and first subscript at $\mu$ denote a barrier breaking at the corresponding transfer. Remaining subscripts at $\mu$ show barriers broken before the time of the transfer.

Dynamics of the number of cells can be described by a system of stochastic differential equations, which reproduces the mathematical structure of the population models of mortality and aging [3]. All results developed for those models, including solution for hazard function, parameter estimation procedure can be applied for the model presented in Figure 6. The following experimental designs can be considered, for which the solution can be constructed analytically or numerically: design I) only the age at onset is observed for individuals of the study cohort, design II) the age at onset is detected and auxiliary information helping to model initial conditions of barrier mechanisms and their dynamics, is used, design III) The age at onset is detected and several or all barrier states are measured at regular time intervals, design IV) The age at onset is detected and several or all barrier states are measured at irregular time intervals, and design V) Epidemiological information on the age specific incidence rate is used in addition to rare measurements of barriers in a randomly selected sub-cohort of a small size.

This model for the design I can be used for analyses of age-specific incidence rates as observed in the SEER. SEER was begun in 1973 and captured approximately 14% of the US population, the
expansion registries increased the coverage to approximately 26% [6]. The results of the fit of the chronic lymphoid (CLL) and myeloid leukemia (ML) age patterns using the five parameter version of BBM and TSCE models are shown in Figure 7. Statistical criteria (such as $\chi^2$/d.o.f and AIC) show that the BBM model describes data for CLL and ML significantly better than TSCE. The age pattern of the lymphoid leukemia incidence rate (the ICD-9-CM code is 204 excluding 204.1) has a sharp peak at ages of 2-5 years. Description of the whole age region of the incidence pattern including this peak within the same model is possible by introducing a mixture of the two sub-cohorts with different distributions of initial (e.g., genetic) damage (lower plots of Figure 7). Generalization for IR-induced carcinogenesis is performed in a way similar to those used in standard models. Corrections to initiation and transformation rates can be introduced as in Heidenreich et al. [27], i.e., $\nu(d) = \nu(0)(1+\nu_1 d)$ and $\mu(d) = \mu(0)(1+\mu_1 d)$. Following refs [14,27], the promotion effect is modeled as correction parameters $\Delta$ describing proliferation/apoptosis balance, i.e., $\Delta(d)=\Delta(0)+\Delta_{level}(1-\exp(\Delta_1 d))$ or simply $\Delta(d)=\Delta(0)(1+\Delta_1 d)$. Recently it has been proven that the apoptosis rate is significantly higher for individuals who were at chronic exposure to IR in the past [38]. This can be interpreted as an indication to the existence of the promotion effect of chronic IR.

**Measurement of Barrier States.** Modern state of science opens broad possibility for individual measurements of states of barrier mechanisms, e.g., TUNEL for measurement of apoptosis, etc. For estimating the probability of BBM, the standard loading tests (e.g., experiments with additional exposure and measurements of barrier mechanism response to newly created damages) are useful. Other promising approaches of measurements and modeling barrier states are based on actively developing technologies of biochips providing estimation of gene expression and SNP analysis. Application of these technologies will allow to extend the number of measured model parameters with simultaneous broad coverage of possible mechanisms underlying carcinogenesis.

**Simulation Studies.** To prove the predictive power of the model, we performed a series of simulation studies for different designs discussed above. For each such study, we assumed true values for substantial parameters, simulated tens of datasets, estimated parameters for each simulated dataset, averaged them over all simulated datasets, and compared with true values. An example of the results of the simulation study for Design III is presented in Table 2. We simulated 50 datasets with 100,000 total person years. The quantity RAT estimated as $RAT=\frac{\text{mean}-\text{true}}{\text{SE}}$ is the characteristics of the quality of reconstruction of corresponding variables and related with p-value of the acceptance of the
hypothesis that the mean obtained by averaging over parameter estimates from all simulated databases coincides with true values used for simulations.

Conclusions from simulation studies are that i) dynamic parameters (e.g., all $v$ and all $\Delta$) are defined quite well; ii) identification of hazard parameters, i.e., $\mu_{BA}$, $\mu_{AB}$, $\mu_{BAC}$, $\mu_{ABC}$, requires large statistics, especially for rare events; iii) for Design III all parameters are identified while identifiability of all parameters in the general case (i.e., for Design IV) requires additional investigation. Design V, in which data are combined from epidemiological measurements (as in Design I) and physiological (i.e., barrier state) measurements would be the most promising and beneficial for efficient data collection and parameter estimation. To illustrate this fact, we performed an additional simulation study in which information on dynamics of barrier state variables was only used for 1% of randomly selected individuals. Statistical errors of the estimates responsible for dynamics declined, but those responsible for mortality were almost unchanged. The ratio of statistical errors is given in the last row of Table 1.

Table 1. Results of simulation studies

<table>
<thead>
<tr>
<th>$V_i$</th>
<th>$V_s$</th>
<th>$V_c$</th>
<th>$\Delta_{AC}$</th>
<th>$\Delta_{AB}$</th>
<th>$\Delta_{A}$</th>
<th>$\mu_{CA}$</th>
<th>$\mu_{CA}$</th>
<th>$\mu_{CA}$</th>
<th>$\mu_{AC}$</th>
<th>$\mu_{AC}$</th>
<th>$\mu_{ABC}$</th>
<th>$\mu_{ABC}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>true</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>5.0</td>
<td>5.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>mean</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.8</td>
<td>2.1</td>
<td>2.1</td>
<td>1.6</td>
<td>4.5</td>
<td>9.3</td>
<td>2.6</td>
</tr>
<tr>
<td>SE</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>3.4</td>
<td>0.1</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
<td>1.2</td>
<td>0.9</td>
</tr>
<tr>
<td>RAT</td>
<td>-3.1</td>
<td>-0.1</td>
<td>-0.8</td>
<td>3.4</td>
<td>1.6</td>
<td>0.5</td>
<td>0.8</td>
<td>-0.0</td>
<td>0.4</td>
<td>0.6</td>
<td>-1.3</td>
<td>-0.9</td>
</tr>
<tr>
<td>SE'/SE</td>
<td>9.6</td>
<td>8.7</td>
<td>9.9</td>
<td>10.2</td>
<td>8.8</td>
<td>11.5</td>
<td>9.8</td>
<td>12.6</td>
<td>8.6</td>
<td>1.1</td>
<td>9.9</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Initial true values of the parameters are taken for 5 parameter model; all rates are in 1/year

4. Discussion

Summary. Biologically-motivated mathematical models of carcinogenesis constitute a complementary approach to empirical analyses and standard statistical models, and, therefore, are important for understanding carcinogenesis despite simplification of the biological reality. Widely used mechanistic multistage models have certain disadvantages, e.g., oversimplification of tumor development process in a two-stage model, difficulty of creating mathematical formalism and only partial identifiability of biological parameters in many of these models. Our studies show that the BBM model of carcinogenesis suggested in this paper has the potential to overcome at least part of the limitations of the traditional mechanistic models of carcinogenesis. It also provides perspectives for constructing individualized prognoses of cancer related health effects of IR. The approach suggested in this paper has combined recent methodological developments in modeling the cancer risk and provided new techniques for description of multistage and multiple pathways models of carcinogenesis based on diffusion type stochastic differential equations. The BMM model relies on the concept of barrier mechanisms in a cell as playing a key role in preventing the cell malignant transformation. Methodologically, this new model is a generalization of ideas of multistage and multiple pathways models to the concept of barrier mechanisms.

Modeling vs. Empirical Analysis in Radiation Epidemiology. Many seminal results in radiation epidemiology were obtained using the methods of empirical analyses and, specifically, the EPICURE program [39]. Models relying on biologically motivated principles are capable of enriching empirical findings and obtaining new modeling-associated results. There are strong advantages of the modeling approaches based on biological concepts. Such models incorporate knowledge on mechanisms of carcinogenesis and allow for estimation of biologically motivated parameters. They can circumvent some implicit assumptions of empirical analysis (e.g., on proportional hazards) and allow for estimation of baseline carcinogenesis models. Often such models are useful for elaborating clear strategies for sensitivity analysis and they also have the potential to be combined with models describing survival and effect of treatment in cancer patients. Furthermore, using the biologically motivated models is important because of their ability to deal with data of non-standard structure. Types of datasets in radiation epidemiology vary and include: i) case-control studies, ii) grouped data, i.e., measured person-years and number of cases, iii) epidemiological registers, iv) follow-ups with covariates measured at regular time intervals, v) follow-ups with covariates measured at irregular time intervals, vi) tracking individual medical histories, vii) surveys, i.e., categorical measurements, and
viii) follow-ups of categorical measurements. In many cases, data are collected under irregular study design. Data often include a combination of several sources and, in the majority of cases, they have missing data. In such situations, standard methods often fail and a comprehensive analysis becomes possible only by using mathematically appropriate models.

**Incorporating Genomic Instability.** Since the IR-induced genomic instability is one possible mechanism of carcinogenesis, a lot of attention has recently been paid to modeling this process and incorporating it into existing carcinogenesis models [17,40,41]. Such incorporation, however, should be performed with care to avoid overcomplicating the model with additional parameters, which can easily become non-identifiable. In the case of the carcinogenesis model based on measurements of the barrier mechanisms, the effects of genomic instability are naturally incorporated into the model because states of the barrier mechanisms are simultaneously the biomarkers of genomic instability. Figure 7 suggests that our model also has a potential for forecasting risks of children’s cancers. Further development of the model for offspring of people chronically exposed to low-dose IR will allow one to naturally include consideration of the effects of trans-generational transfer of genomic instability.

**Acknowledgments.** The paper was improved thanks to discussions of the main results at presentation at IRCC-2007 [42] with M. Dewhirst, D. Ilyasova, P. Jacob, W. Heidenreich, M. Little, H.G. Paretzke, D. Preston, E. Ron, M. Schubauer-Berigan, and Z. Vujaskovic. This work was supported by NIA/NIH grants 5R01-AG-028259, 5R01-AG-027019 and 5PO1-AG-008761.

**References**

[40] Little MP, Li G. Stochastic modelling of colon cancer: is there a role for genomic instability? Carcinogenesis 28, 479 (2007).