Parameter Identifiability and Redundancy in a General Class of Stochastic Carcinogenesis Models

Mark P. Little¹*, Wolfgang F. Heidenreich², Guangquan Li¹

¹ Department of Epidemiology and Public Health, Imperial College, London, United Kingdom; ² Institut für Strahlenschutz, Helmholtz Zentrum München, Neuherberg, Germany

Abstract

**Background:** Heidenreich et al. (Risk Anal 1997 17 391–399) considered parameter identifiability in the context of the two-mutation cancer model and demonstrated that combinations of all but two of the model parameters are identifiable. We consider the problem of identifiability in the recently developed carcinogenesis models of Little and Wright (Math Biosci 2003 183 111–134) and Little et al. (J Theoret Biol 2008 254 229–238). These models, which incorporate genomic instability, generalize a large number of other quasi-biological cancer models, in particular those of Armitage and Doll (Br J Cancer 1954 8 1–12), the two-mutation model (Moolgavkar et al. Math Biosci 1979 47 55–77), the generalized multistage model of Little (Biometrics 1995 51 1278–1291), and a recently developed cancer model of Nowak et al. (PNAS 2002 99 16226–16231).

**Methodology/Principal Findings:** We show that in the simpler model proposed by Little and Wright (Math Biosci 2003 183 111–134) the number of identifiable combinations of parameters is at most two less than the number of biological parameters, thereby generalizing previous results of Heidenreich et al. (Risk Anal 1997 17 391–399) for the two-mutation model. For the more general model of Little et al. (J Theoret Biol 2008 254 229–238) the number of identifiable combinations of parameters is at most \( r + 1 \) less than the number of biological parameters, where \( r \) is the number of destabilization types, thereby also generalizing all these results. Numerical evaluations suggest that these bounds are sharp. We also identify particular combinations of identifiable parameters.

**Conclusions/Significance:** We have shown that the previous results on parameter identifiability can be generalized to much larger classes of quasi-biological carcinogenesis model, and also identify particular combinations of identifiable parameters. These results are of theoretical interest, but also of practical significance to anyone attempting to estimate parameters for this large class of cancer models.

Citation: Little MP, Heidenreich WF, Li G (2009) Parameter Identifiability and Redundancy in a General Class of Stochastic Carcinogenesis Models. PLoS ONE 4(12): e8520. doi:10.1371/journal.pone.0008520

Editor: Dow J. Stekel, University of Nottingham, United Kingdom

Received September 6, 2009; Accepted November 30, 2009; Published December 31, 2009

Copyright: © 2009 Little et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was funded partially by the European Commission under contract FI6R-CT-2003-508842 (RISC-RAD) and FP6-036465 (NOTE). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: mark.little@imperial.ac.uk

Introduction

Models for complex biological systems may involve a large number of parameters. In principle it may well be that some of these parameters may not be observed, or be possible to be derived from observed data via regression techniques. Such parameters are said to be unidentifiable or non-identifiable, the remaining parameters being identifiable.

There is a substantial literature on identifiability in stochastic models in various contexts [1,2,3]. Catchpole and Morgan [3] considered identifiability and parameter redundancy and the relations between them in a general class of (exponential family) models. Catchpole and Morgan [3] defined a set of model parameters in an exponential family model to be redundant if the likelihood can be written using a strictly smaller parameter vector; otherwise they are irredundant. Rothenberg [1], Jacquez and Perry [4] and Catchpole and Morgan [3] also defined a notion of local identifiability, to mean that within a neighbourhood of each set of parameter values the likelihood differs for at least some data points. This notion has been extended by Little et al. [5] to gradient weak local identifiability and weak local identifiability. Little et al. [5] defined a set of parameters to be weakly locally identifiable if the maxima of the likelihood are isolated; they defined parameters to be gradient weakly locally identifiable if the turning points (those for which the likelihood derivative with respect to the parameters is zero) are isolated. The results obtained by Little et al. [5] (Corollary 2 (ii) and the subsequent Remark (ii)), show that, subject to some regulatory conditions, the number of locally identifiable or (gradient) weakly locally identifiable parameter combinations is equal to the rank of the Hessian matrix, or equivalently the rank of the Fisher information matrix. The notions of identifiability in stochastic models [1,2,3,5], within which framework this paper is set, should be contrasted with the consideration of identifiability in non-stochastic settings considered by some [4,6,7].

Heidenreich [8] and Heidenreich et al. [9] considered parameter identifiability in the context of the two-mutation cancer model [10] and demonstrated that of the five biological parameters in the model, on the basis of the cancer hazard function only three could
be identified. [It should be noted that given extra information, for example on numbers and sizes of intermediate cell compartment clones, there is information on an additional parameter.]

In this paper we consider the problem of identifiability in recently developed carcinogenesis models of Little and Wright [11] and Little et al. [12]. These models generalize a large number of other quasi-biological cancer models, in particular those of Armitage and Doll [13], the two-mutation model [10], the generalized multistage model of Little [14], and a recently developed cancer model of Nowak et al. [15] that incorporates genomic instability. We shall show that via a specific reparameterization, in the simpler model proposed by Little and Wright [11] in principle combinations of all but two of the model parameters are identifiable, thereby generalizing previous results of Heidenreich [8] and Heidenreich et al. [9] for the two-mutation cancer model. For the more general model of Little et al. [12] combinations of all but \( r_z \) of the model parameters are identifiable, where \( r \) is the number of destabilization types, thereby also generalizing all these results. We also identify particular forms of identifiable parameters.

**Methods**

**Parameter Identifiability in the Context of a Stochastic Cancer Model with Genomic Instability**

We consider the problem of parameter identifiability in a particular class of stochastic cancer models, those of Little and Wright [11] and Little et al. [12]. The ideas used are similar to those employed by Heidenreich et al. [9], in particular the use of Cauchy’s method of characteristics. We shall assume throughout this section that this model is embedded in a member of the exponential family so that the log-likelihood is given by

\[
L(x|\theta) = \sum_{i=1}^n \left( \frac{z_i y_i}{c} + u(y_i, \phi) \right)
\]

where the natural parameters \( z_i \) are functions of the model parameters \( \theta \) and some auxiliary data \( y_i \), but that the scaling parameter \( c \) is not. We shall assume that the parameters \( \theta \) of the model are identifiable, thereby generalizing previous results of Heidenreich [8] and Heidenreich et al. [9] for the two-mutation cancer model. For the more general model of Little et al. [12] combinations of all but \( r \) of the model parameters are identifiable, where \( r \) is the number of destabilization types, thereby also generalizing all these results. We also identify particular forms of identifiable parameters.

**Figure 1. Diagram of cancer model with \( k \) cancer-stage mutations and \( m \) destabilizing mutations, as in [12].**

doi:10.1371/journal.pone.0008520.g001
destabilizing mutation of type \( d \) \((1 \leq d \leq r)\), it and its daughter cell can acquire up to \( m_d - 1 \) further destabilizing mutations of the same type. We define \( r \) to be the multiplicity of destabilization mutation types. It is to be expected that the more destabilizing mutations cells acquire of each type, the higher the cancer stage mutation rate is, but this is not intrinsic to the model. We write \((m_1 - m_2 - \ldots - m_r)\) as the signature of the destabilizing mutation types. We habitually describe this model as of type \( k - r - (m_1 - m_2 - \ldots - m_r) \) for short. The model is illustrated schematically in Figures 1 and 2. Table 1 lists the biological parameters that are used in the model, and their multiplicity.

Cells at different stages of the process are labelled by \( I(a,b,d) \), where the first subscript, \( a \), represents the number of cancer stage mutations that the cell has accumulated, the second subscript, \( b \), represents the number of destabilizing mutations acquired, their type being given by the third subscript, \( d \). At all stages other than \( I(0,0,0) \), cells are allowed to divide symmetrically or differentiate (or undergo apoptosis) at rates \( G(a,b,d) \) and \( D(a,b,d) \), respectively.

Each cell can divide into an equivalent daughter cell and another cell with an extra cancer stage mutation at rate \( M(a,b,d) \). Likewise, cells can also divide into an equivalent daughter cell and another cell with an additional destabilizing mutation of type \( d \) at rate \( A(a,b,d) \). The model assumes that there are \( X(t) \) susceptible stem cells at age \( t \). Further details on derivation of the hazard function are given in the paper of Little et al. [12].

Results

In Text S1 Section B we derive the hazard function and show that it can be written in terms of certain combinations of the biological parameters given in Table 1. From equations (B12)–(B16) in Text S1 Section B it is seen that the characteristics and \( \psi \) are governed by certain parameter combinations. Table 2 summarizes the maximum number of identifiable parameter combinations and their forms associated with each cell compartment. The maximum number of identifiable parameters associated with each destabilization zone, \( I(a,b,d) \), are \( 4 \) when \( \alpha < k - 1 \) and \( \beta < m_d \); \( 3 \) when \( \alpha < k - 1 \) and \( \beta = m_d \) and \( 2 \) when \( \alpha = k - 1 \) and \( \beta = m_d \). The function \( \psi \) is governed by at most \( r + 1 \) parameter combinations. Therefore, we have shown that the hazard function \( h(\theta) \) can be written as \( h(G_1(\theta),G_2(\theta),\ldots,G_N(\theta)) \) for some scalar functions \( G_1(\cdot),G_2(\cdot),\ldots,G_N(\cdot) \), where \( N = (k - 2)(3 + r) + (3 + r) + 1 + (1 + r) + 4(k - 1) - 2(k - 1)r \) when \( r \geq 2 \) and \( 3(k - 2) + 4k \cdot \sum_{d=1}^{r} m_d \) when \( r = 1 \).


Table 1. The number of biological parameters in a model with \( k \) cancer stages, \( r \) types of Gl and \( m_d \) \((d = 1, \ldots, r)\) levels of destabilizations.

| Model parameter descriptions | Model parameters | Number of such parameters in the model |
|----------------------------|------------------|--------------------------------------|
| Stem cell population number | \( X(t) \)        | 1                                    |
| Growth rate                | \( A(x,\beta,d) \) | \( k - 1 + k \sum_{d=1}^{r} m_d \) |
| Death/differentiation rate  | \( D(x,\beta,d) \) | \( k - 1 + k \sum_{d=1}^{r} m_d \) |
| Cancer-stage mutation rate  | \( M(x,\beta,d) \) | \( k + k \sum_{d=1}^{r} m_d \)       |
| Destabilizing mutation rate | \( A(x,\beta,d) \) | \( k \sum_{d=1}^{r} m_d \)           |
| Total                      |                   | \( 3k - 1 + 4k \sum_{d=1}^{r} m_d \) |

which therefore has rank at most \( N \). A similar argument that if one were to reparameterise (via some invertible \( C \) mapping \( \theta = f(\omega) \)) then the embedded log-likelihood \( L(x|\omega) = L(x|\theta) \) associated with \( h^{-1}(\theta) = h(\omega) \) must also have Fisher information matrix of rank at most \( N \). By Theorems 1 and 3 of Catchpole and Morgan [3], for this embedded exponential family model there can be at most \( N \) irredundant parameters. Therefore, of the theoretically available \( 3k - 1 + 4k \sum_{d=1}^{r} m_d \) biological parameters (Table 1), at most

\[
N = 3k - 2 + 4k \sum_{d=1}^{r} m_d
\]
information matrix, so $\leq N$. For example, in the case of the familiar two-mutation model [10], with $k = 2$, $r = 1$, $d = 0$ and $m_d = 0$, there are $k(m+1) - 1 = 2 \times 1 - 1 = 1$ $G$'s (namely $G(1,0,0)$), $k(m+1) - 1 = 2 \times 1 - 1 = 1$ $D$'s (namely $D(1,0,0)$), $k(m+1) - 1 = 2 \times 2 = 4$ $M$'s (namely $M(0,0,0), M(1,0,0)$), and a single $X$, giving a total of five biological parameters. It is known from the results of Heidenreich et al. [3,9] that for the two-mutation model only three combinations of these are estimable, i.e., that there are two redundancies, precisely in agreement with the result given here for $r = 1$. This result therefore precisely generalizes the results and approach of Heidenreich et al. [3,9]. Unfortunately, analytical methods for proving that precisely this number of parameters are estimable, including some recently outlined [17], cannot be used for the model considered here. Nevertheless, we conjecture that in fact precisely this number of parameters are estimable, so that the upper bound on the number of estimable parameter combinations that we have proved above is in fact sharp. This is supported by numerical evaluation of the Hessian in a couple of example cases, which we now outline.

**Numerical Evaluation of Hessian and Determination of Its Rank**

That there are likely to be exactly this number of estimable parameters is supported by numerical evaluation of the Hessian matrix of the hazard function. We make use of the solution of the system of ordinary differential equations defining the Hessian, outlined in Text S1 Section D. We will show in two cases that the Hessian has rank two less than the number of biological parameters, $w$. By the above-mentioned results of Catchpole and Morgan [3] and Little et al. [5] this suggests that precisely $w - 2$ parameters are (gradient) weakly locally identifiable. In order to show that the Hessians are of rank two less than the number of biological parameters, we evaluate the eigenvalues of the Hessian matrix, and establish that the smallest eigenvalue among the $w - 2$ largest eigenvalues in absolute value exceeds the likely magnitude of the error by at least an order of magnitude. We know the likely size of the error in numerical evaluations of each element, $h_{ij}$, of the Hessian from the Bochisch-Stoer integrator that is employed, namely max$\{10^{-10}, 10^{-10} |h_{ij}| : 1 \leq i, j \leq w\}$ (bstep routine, Press et al. [18], p.722). It is known that if two symmetric matrices $H$ and $H'$ have eigenvalues $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_w$ and $\hat{\lambda}_1 \leq \hat{\lambda}_2 \leq \ldots \leq \hat{\lambda}_w$ then $|\lambda_i - \hat{\lambda}_i| \leq |H - H'|_2$, $1 \leq i \leq w$, where $|H|_2 = \sup \|Hx\|/\|x\|_2 : x \neq 0$ [19](p.396). Since the approximate Hessian that we calculate, $H'$, differs from the true Hessian, $H$, by an amount $|H - H'|_2 \leq \sqrt{w} \max \{|h_{ij} - \hat{h}_{ij}| : 1 \leq i, j \leq w\}$, we know that:

$$|\lambda_i - \hat{\lambda}_i| \leq \sqrt{w} \max \{|h_{ij} - \hat{h}_{ij}| : 1 \leq i, j \leq w\} \leq \sqrt{w} \max 10^{-10}, 10^{-10} |\hat{h}_{ij}| : 1 \leq i, j \leq w$$

(2)

There is also the issue of numerical roundoff error in the QR algorithm (Numerical Algorithms Group (NAG) routine F02FAF [20]) used to compute eigenvalues. If we write now $\hat{\lambda}_i, \tilde{\lambda}_i$ for the true and approximate eigenvalues associated with the approximate Hessian, $H'$, this is known to be bounded by:

$$|\hat{\lambda}_i - \tilde{\lambda}_i| \leq \epsilon(w) \cdot c \cdot \max |H|_2 \leq \epsilon(w) \cdot c \cdot \sqrt{w} \max |h_{ij} - \hat{h}_{ij}| : 1 \leq i, j \leq w,$$

(3)

where $c(w)$ is a modestly increasing function of the dimension, $w$, of the approximate Hessian $H$ and $\epsilon$ is the machine precision [19](Chapter 8). Since the machine precision (in double precision) is of the order $10^{-15}$ this expression (3) will be dominated by the error associated with the approximation to the Hessian, given by expression (2).
Table 3. Example coefficients of model with three cancer stage mutations and one destabilizing mutation.  

| Coefficient | Value          |
|------------|----------------|
| G(1,0,0)   | 8.64714335947694 × 10^{-2} |
| G(2,0,0)   | 1.06188950764276 × 10^{-3} |
| D(1,0,0)   | 4.25556779376062 × 10^{-2} |
| D(2,0,0)   | 2.68975960218019 × 10^{-1} |
| M(0,0,0)   | 1.33167380528588 × 10^{-2} |
| M(1,0,0)   | 1.08414503240502 × 10^{0}  |
| M(2,0,0)   | 9.79093689335407 × 10^{-2} |
| A(1,0,1)   | 1.3357580655960 × 10^{-1} |
| A(1,1,1)   | 7.65789029061483 × 10^{-2} |
| A(2,0,1)   | 3.73742902971737 × 10^{-2} |
| G(0,1,1)   | 5.31044255773088 × 10^{-1} |
| G(1,1,1)   | 3.12418227871070 × 10^{1}  |
| G(2,1,1)   | 6.88636709884594 × 10^{0}  |
| D(1,1,1)   | 1.14118194967370 × 10^{-2} |
| D(2,1,1)   | 2.99644033522771 × 10^{-1} |
| M(0,1,1)   | 7.55711890719015 × 10^{0}  |
| M(1,1,1)   | 6.58304565565478 × 10^{0}  |
| M(2,1,1)   | 4.33626256393215 × 10^{-3} |
| X          | 4.0699305654860 × 10^{0}  |

doi:10.1371/journal.pone.0008520.t003

We evaluated the Hessians matrix for a model with three cancer-stage mutations and one destabilizing mutation, and a model with two cancer-stage mutations and one destabilizing mutation; log-normal perturbations of all parameters were performed, assuming a geometric standard deviation (GSD) of 4, centred on models with cancer-stage mutation rates of 4.0 × 10^{-2} year^{-1}, destabilization mutation rates of 3.0 × 10^{-3} year^{-1}, intermediate cell proliferation rates of 1.0 × 10^{-1} year^{-1}, and intermediate cell death rates of 5.0 × 10^{-1} year^{-1}. For each of 1000 random sets of parameters we evaluated the Hessian by numerical integration, as outlined in Text S1 Section D. We calculated the eigenvalues of the Hessian using the QR algorithm, specifically the NAG FORTRAN subroutine F02FAF [20]. For each model we selected the set of random parameters for which the ratio of minimum to maximum among the w−2 largest eigenvalues (w being the number of biological parameters) in absolute value was greatest. These are given in Tables 3 and 4, for the three-stage and two-stage models, respectively. The associated eigenvalues are given in Table 5. The absolute value of the w−2th smallest eigenvalue associated with each set exceeds the error bound (2) by at least an order of magnitude in each case. This strongly suggests that the Hessians calculated for these two examples really are of rank w−2 for each model.

Table 4. Example coefficients of model with two cancer stage mutations and one destabilizing mutation.  

| Coefficient | Value          |
|------------|----------------|
| G(1,0,0)   | 2.22095885699822 × 10^{-3} |
| G(1,0,0)   | 3.13787396131941 × 10^{-6} |
| M(0,0,0)   | 8.12220928775447 × 10^{-4} |
| M(1,0,0)   | 1.406740103565097 × 10^{-5} |
| A(0,0,1)   | 2.066681086660923 × 10^{-1} |
| A(1,0,1)   | 4.5724970326658 × 10^{-3} |
| G(0,1,1)   | 1.5664435664010 × 10^{-2} |
| G(1,1,1)   | 3.16379145991048 × 10^{-4} |
| D(0,1,1)   | 1.29017705678554 × 10^{0}  |
| D(1,1,1)   | 1.929697373536413 × 10^{-1} |
| M(0,1,1)   | 9.35817313727697 × 10^{-9} |
| M(1,1,1)   | 2.26339224702545 × 10^{-1} |
| X          | 2.7841105650539 × 10^{-1} |

doi:10.1371/journal.pone.0008520.t004

Table 5. Eigenvalues in ascending order of Hessian matrix associated with a model with three cancer stage mutations and one destabilizing mutation (as in Table 3), and with a model with two cancer stage mutations and one destabilizing mutation (as in Table 4).  

| Number | Eigenvalues (Table 3) | Eigenvalues (Table 4) |
|--------|------------------------|------------------------|
| 1      | −1.20726415206490 × 10^{1} | −1.45810346778189 × 10^{0} |
| 2      | −4.93248755175100 × 10^{0} | −7.774141881355 × 10^{1} |
| 3      | −1.16489800886060 × 10^{0} | −2.7712189259301 × 10^{1} |
| 4      | −2.44711976272777 × 10^{1} | −6.66243518532325 × 10^{3} |
| 5      | −9.84288250086772 × 10^{2} | −3.532097768268 × 10^{4} |
| 6      | −1.238114589706358 × 10^{2} | −2.8467103286276 × 10^{4} |
| 7      | −2.95523295984747 × 10^{4} | −9.25930409562877 × 10^{4} |
| 8      | −1.53669786311497 × 10^{4} | −1.78637642487767 × 10^{11} |
| 9      | −9.80139032107413 × 10^{4} | 2.7434292675736 × 10^{4} |
| 10     | −3.36238193418727 × 10^{4} | 4.9897524563660 × 10^{4} |
| 11     | −2.14057731816777 × 10^{4} | 1.112573049368 × 10^{2} |
| 12     | −1.86967299054058 × 10^{7} | 8.1842650733286 × 10^{1} |
| 13     | 5.01559183858810 × 10^{-12} | 1.4595703291853 × 10^{0} |
| 14     | 9.49044820094881 × 10^{-7} | - |
| 15     | 4.05661819862605 × 10^{-4} | - |
| 16     | 1.92220196143433 × 10^{-3} | - |
| 17     | 1.1042617352459 × 10^{-2} | - |
| 18     | 1.03227190243219 × 10^{-1} | - |
| 19     | 1.1266702944003 × 10^{0} | - |
| 20     | 1.08248991510735 × 10^{-1} | - |

doi:10.1371/journal.pone.0008520.t005

Non-significant eigenvalues are underlined in bold.

Discussion

We have shown that in the class of stochastic cancer models incorporating genomic instability developed by Little and Wright [11] the number of identifiable combinations of parameters is at most two less than the number of biological parameters, thereby generalizing previous results of Heidenreich et al. [8,9] and Hanin et al. [21,22] for the two-mutation model, a special case of this model. For the more general genomic-instability cancer model of Little et al. [12] the number of identifiable combinations of parameters is at most r+1 less than the number of biological parameters, where r is the number of destabilization types, thereby
also generalizing all these results. Numerical evaluations in two special cases (with \( r=1 \)) suggest that this bound is tight: a combination of parameters with cardinality two less than the number of biological parameters is of full rank, and so is not redundant.

A weakness of the paper is that one cannot be absolutely sure (because of the uncertainty implicit in any numerical evaluation) that the bound demonstrated by the mathematics of section 3 and Text S1 Section B is sharp. Nevertheless, we have clearly established a maximum number of identifiable parameter combinations. We have also specified particular combinations of identifiable parameters, and these should be used in model fitting to avoid obvious numerical problems, of lack of convergence and absence of a unique set of parameters maximizing the likelihood.

These results have obvious implications for the large number of other quasi-biological cancer models that are special cases of these models, in particular those of Armitage and Doll [13], the two-mutation model [10], the generalized multistage model of Little [14], and a recently developed cancer model of Nowak et al. [15] that incorporates genomic instability. It should be noted that the results given here are for the fully stochastic solution of the model, and would not be applicable, for example, to the deterministic approximation of the multistage model of Armitage and Doll [13] that is often employed in applications.

Our results imply that for the general class of cancer models considered here, only certain specific parameter combinations should be estimated in principle, and this is the case whatever the size of the dataset being considered. Whether for complex models for even this theoretically available number of parameters there is useful information is of course uncertain, and may well depend on the particular dataset and on the likely size of the parameters to be estimated. However, fits to a large population-based registry of colon cancer, as recently analysed by Little and Li [16], suggests that, for example, the model with two cancer-stage and one destabilizing mutations can be fitted to the dataset and yields stable parameter estimates for certain combinations of 11 parameters, in accordance with the results of this paper.

### Supporting Information

**Text S1** Text S1
Found at: doi:10.1371/journal.pone.0008520.s001 (0.42 MB DOC)

### Acknowledgments

The authors are very grateful for the comments of Professor Byron Morgan on an advanced draft of the paper. The authors are also grateful for the detailed and helpful comments of the two referees.

### Author Contributions

Conceived and designed the experiments: MPL WFH GL. Performed the experiments: MPL. Analyzed the data: MPL GL. Wrote the paper: MPL WFH GL.

### References

1. Rothenberg TJ (1971) Identification in parametric models. Econometrica 39: 577–591.
2. Silvey SD (1975) Statistical inference. London: Chapman and Hall. 192 p.
3. Catchpole EA, Morgan BJT (1997) Detecting parameter redundancy. Biometrics 53: 187–196.
4. Jacquez JA, Perry T (1990) Parameter estimation: local identifiability of parameters. Am J Physiol 258: E727–E736.
5. Little MP, Heidenreich WF, Li G (2009) Parameter identifiability and redundancy: theoretical considerations. PLoS ONE, submitted (also available on arXiv:0902.1701).
6. Chappell MJ, Gunn RN (1998) A procedure for generating locally identifiable reparameterizations of unidentifiable non-linear systems by the similarity transformation approach. Math Biosci 148: 21–41.
7. Evans ND, Chappell JM (2000) Extensions to a procedure for generating locally identifiable reparameterizations of unidentifiable systems. Math Biosci 168: 137–159.
8. Heidenreich WF (1996) On the parameters of the clonal expansion model. Radiat Environ Biophys 35: 127–129.
9. Heidenreich WF, Liebeck EG, Moolgavkar SH (1997) Some properties of the hazard function of the two-mutation clonal expansion model. Risk Anal 17: 391–399.
10. Moolgavkar SH, Venzon DJ (1979) Two-event models for carcinogenesis: incidence curves for childhood and adult tumors. Math Biosci 47: 55–77.
11. Little MP, Wright EG (2003) A stochastic carcinogenesis model incorporating genomic instability fitted to colon cancer data. Math Biosci 183: 111–134.
12. Little MP, Vines P, Li G (2008) A stochastic carcinogenesis model incorporating multiple types of genomic instability fitted to colon cancer data. J Theoret Biol 254: 229–238. 255: 265–268.
13. Armitage P, Doll R (1954) The age distribution of cancer and a multi-stage theory of carcinogenesis. Br J Cancer 8: 1–12.
14. Little MP (1995) Are two mutations sufficient to cause cancer? Some generalizations of the two-mutation model of carcinogenesis of Moolgavkar, Venzon, and Knudson, and of the multistage model of Armitage and Doll. Biometrics 51: 1278–1291.
15. Nowak MA, Komarova NL, Sengupta A, Jalili-Palli PV, Shih I-M, et al. (2002) The role of chromosomal instability in tumor initiation. Proc Natl Acad Sci U S A 99: 16226–16231.
16. Little MP, Li G (2007) Stochastic modelling of colon cancer: is there a role for genomic instability? Carcinogenesis 28: 479–487.
17. Cole D, Morgan BJT (2009) Determining the parametric structure of Non-linear models, University of Kent School of Mathematics, Statistics and Actuarial Science preprint, downloadable from http://www.kent.ac.uk/ims/personal/djc24/para.pdf.
18. Press WH, Teukolsky SA, Flannery BT, Vetterling WP (1992) Numerical recipes in FORTRAN, the art of scientific computing (2nd ed.). Cambridge: Cambridge University Press. 500 p.
19. Grölb GH, van Loan CF (1996) Matrix computations (3rd ed.). Baltimore: The Johns Hopkins University Press. 728 p.
20. Numerical Algorithms Group (2006) NAG Library, Mark 21. Oxford: Numerical Algorithms Group.
21. Hanin LG, Yakovlev AY (1996) A nonidentifiability aspect of the two-stage model of carcinogenesis. Risk Anal 16: 711–715.
22. Hanin LG (2002) Identification problem for stochastic models with application to carcinogenesis, cancer detection and radiation biology. Discrete Dyn Nature Soc 7: 177–189.
23. McCullagh P, Nelder JA (1989) Generalized linear models (2nd ed.). London: Chapman and Hall 511 p.