7-18-2012

Stochastic Models of Human Kidney Cancer

Yanan Wu

STOCHASTIC MODELS OF HUMAN KIDNEY CANCER

by

Yanan Wu

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

Major: Mathematical Sciences

The University of Memphis

August 2012

ACKNOWLEDGMENTS

I would never have been able to finish my dissertation without the guidance of my committee members, help from my friends, and support from my family.

My dissertation was finished under the supervision of Dr. Wai-Yuan Tan. I would like to express my deepest gratitude to Dr. Tan, for his constant support, guidance, patience throughout my research work and for devoting his time to help me complete my dissertation. I would like to express my sincere thanks to Dr. Lih-Yuan Deng, who was willing to be my advisor and gave me a great help during my study. I would also like to thank Dr. Seok P. Wong and Dr. Xiaoping Xiong for serving on my dissertation committee and for providing valuable guidance and suggestions during the course of this study.

I would like to express my appreciation to the Department of Mathematical Sciences at the University of Memphis and to the Department of Statistics at St. Jude Children's Research Hospital for their support, education and training.

Many thanks also to my great friends, Dr. Lijing Xu, Dr. Xiaowei Yan, Sheng Ding, Yan Yan, Quan Tang, Weiwei Chen, Xueming Mei and all friends in Memphis, for encouraging me and providing help whenever I needed it.

I would also like to thank my extended family for providing a loving environment for me. My parents, my in-laws and my cousins have provided constant support and encouragement during the whole period of my study.

Finally, and most importantly, I would like to thank my husband Bin Zhang for his constant love, support and understanding.

ABSTRACT

Wu, Yanan, Ph.D. The University of Memphis. August 2012. Stochastic Models of Human Kidney Cancer. Major Professors: Wai-Yuan Tan, Ph.D. and Lih-Yuan Deng, Ph.D.

This dissertation is focused on the development of stochastic models for carcinogenesis of human kidney cancer. Based on recent biological studies, we have developed a multiple-pathway stochastic model for the human pediatric kidney cancer - Wilms' tumor. To account for hereditary cancer cases and the development of non-hereditary cancers through two different pathways in the stochastic model, we have also developed a generalized mixture model. In this mixture model, two mixing probability distributions were applied, which are a multinomial distribution to explain the genetic segregation of the stage-limiting tumor suppressor genes and a binomial distribution to account for the development of non-hereditary cancers through two pathways. We have applied this model to fit and analyze the SEER data of Wilms' tumor from NCI/NIH. Our results indicate that the proposed model involving hereditary and non-hereditary cancer cases fitted the data better than the single-pathway model with hereditary cancer cases.

We have also derived a biologically supported stochastic model for human adult kidney cancer - renal cell carcinoma (RCC) involving three pathways. These pathways are: 3-stage pathway for pRCC , 4-stage pathway for ccRCC and 5-stage pathway for chRCC. To account for different individuals in the population at risk of developing renal cell carcinoma through different pathways, we have also presented a mixture model of three pathways. We have used this model to fit and analyze the SEER data of renal cell carcinoma from NCI/NIH. Our results indicate that the model not only provides a logical avenue to incorporate biological information but also fits the data well.

These models not only would provide more insights into human kidney cancer but also would provide useful guidance for its prevention and control and for prediction of future cancer cases.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION AND SOME CANCER BIOLOGY WITH SPECIAL REVIEW OF HUMAN KIDNEY CANCER

A. Introduction

According to the American Cancer Society, Kidney cancer is the sixth most common cancer for men and the eighth most common cause of cancer for women. In the United States, an estimated 59,600 new cases are expected to be diagnosed in 2012. Kidney cancer incidence rates have been increasing steadily each year. Recent biological studies have shown that kidney cancer is not a single disease. It consists of different cancers of the kidney. The most two common types of kidney cancer are renal cell carcinoma, the kidney cancer in adult, and Wilms' tumor, the kidney cancer in children. Renal cell carcinoma originates in the lining of the proximal convoluted tubule, the very small tubes in the kidney that filter the blood and remove waste products. Wilms' tumor arises from immature kidney cells. These two cancers are developed through distinctly different genetic and biological mechanisms. The aim of this dissertation is to develop some general stochastic models for carcinogenesis of these two different cancers basing on recent results from kidney cancer biology.

The structure of the dissertation is as follows. In Chapter I, we summarize some cancer biology and review kidney cancer biology. In Chapter II, we develop a stochastic model with multiple-pathway for Wilms' tumor involving both hereditary and non-hereditary cancer cases. We also derive a generalized mixture model to account for the inherited cancer cases and different pathways for non-hereditary cancer development. Then we illustrate the application of the models and methods by analyzing the SEER data of Wilms' tumor from NIC/NIH. In Chapter III, we derive a multi-stage stochastic model for renal cell carcinoma involving multiple pathways. We also discuss the application of the models and methods by analyzing the SEER data of renal carcinoma from NIC/NIH. In Chapter IV, we present the discussions and conclusions.

B. Some Cancer Biology

The human body consists of two types of cells: stem cells and differentiated cells. Normally, stem cell grow and divide in a controlled way to produce new stem cells and new differentiated cells to replace old or damage cells; differentiated cells do not divide and are end cells to serve as components of the tissue in human body. Cancer is a disease in which stem cells grow out of control to form new abnormal cells and have the ability to invade other tissues and spread to distant body parts. Carcinogenesis or tumorigenesis is a stochastic proliferation and differentiation process by which normal stem cells become cancer cells due to a series of irreversible genetic alterations [1], [2], [3].

There are two types of cancer gene that are affected by these genetic alterations. One type of cancer gene is an oncogene, which includes instructions for controlling when cells grow, divide, and die and it can speed up cell division and stop cells from dying. The other type of cancer gene is a tumor suppressor gene, which slows down cell division, or causes cells to die at the right time.

The two basic classes of genetic alterations are germline mutations and somatic mutations. A germline mutation is an genetic alteration in the body's reproductive cells (sperm or egg) that becomes incorporated in the DNA of every cell in the body. Somatic mutations occur in any of the cells of the body except the germ cells (sperm and egg) and therefore are not passed on to offspring. Cancer caused by germline mutations is called inherited cancer (hereditary cancer). Cancer that is due to somatic mutations is called sporadic cancer (non-hereditary cancer), which accounts for 90% - 95% of all cancer cases.

In 2000, Hanahan and Weinberg [4] proposed six hallmarks that a normal stem cell must accumulate in order to develop into a malignant cancer cell. In a follow-up article in 2011 [5], they incorporated four new hallmarks. Those hallmark characteristics are: (1) Self-sufficient growth signals. Cancer cells stimulate their own growth via genetic changes and/or epigenetic changes. They can either make their own growth hormones or

have changed so that they behave as if a growth stimulus were present even in the absence of growth hormone. (2) Insensitivity to Anti-growth Signals. Cancer cells resist inhibitory signals that might otherwise stop their growth via inactivation or silencing of tumor suppressor genes. (3) Evading Apoptosis. Normal stem cells that accumulate excessive DNA damage undergo apoptosis. However, cancer cells are resistant to apoptosis, and thus they continue to grow and divide even as they accumulate mutations via genetic changes and/or epigenetic changes. (4) Limitless Replicative Potential. Cancer cells can escape the normal limits on how many times a cell can divide. These limits are set in large part by the ends of chromosomes, which are known as telomeres. In normal cells, telomeres shrink with each round of cell division, and when these telomeres become too short the cell can no longer divide. In contrast, cancer cells can lengthen their telomeres, thus allowing them to divide an indefinite number of times. (5) Sustained Angiogenesis. Cancer cells stimulate the growth of blood vessels to supply nutrients to tumors. The process of recruiting new blood vessels is called angiogenesis. (6) Tissue invasion and metastasis. Cancer cells invade local tissue and spread to distant sites (metastasis). In order for cancer to spread, cells must acquire mutations that turn on genes which allow them to break free from the primary tumor, travel through the blood stream, and establish a new colony of cells at another site in the body. (7) Genome Instability and Mutation. Cancer cells generally have severe chromosomal abnormalities, which worsen as the disease progresses. (8) Tumor-Promoting Inflammation. With inflammation, cancer cells can become tumors. (9) Reprogramming Energy Metabolism. Cancer cells come up with their own energy systems to sustain their uncontrolled growth and proliferation. (10) Evading Immune Destruction. It is assumed that many potential cancer cells are destroyed by the immune system, but many cancer cells know how to hide from such immune surveillance.

The above results indicates a multi-stage model of carcinogenesis, that involves the sequential accumulation of many genetic changes or gene mutations. Although a large number of cancer genes involving in human cancers, only a few of genes are stage and

rate-limiting. Hence the number of stage in the multi-stage model of carcinogenesis is finite. In most of human cancers, three or more rate-limiting stages are required for development of cancer [6], [7], [8], [1].

Cancer tumor can be developed by a single pathway in some type of human cancer such as retinoblastoma [1], [9]. However, in many other cancers, the same cancer may derived from multiple pathways [2], [10], [11].

From the population perspective, the human cancer developing through different pathways can be presented by a mixture model with multiple pathways. For example, human colon cancers can be purposed by a mixture of five pathways for carcinogenesis [2] and retinoblastomas can be described by a generalized mixture model to account for hereditary cancer cases [9].

Cancer in children differs significantly from cancer in adults in several important ways. Unlike adult cancers, pediatric cancers are rare. Pediatric cancers are usually much more aggressive than adult cancers. Adult cancers are often related to specific risk factors, however most pediatric cancers have no clear cause. Pediatric cancers typically respond better to current therapies than adult cancers. Adult cancer are mostly developed from highly differentiated epithelial tissues. Whereas pediatric cancer are generally derived from non-ectodermal embryonal tissues [12]. However, like adult cancers, most pediatric cancers are not inherited. They result from genetic mutations obtained during the child's life. Some of cancer cases are due to hereditary cancer syndromes.

C. Review of Kidney Cancer

The most common kidney cancers in children are Wilms' tumors, which are embryonal kidney tumors derived from immature kidney cells. They are quite rare and contribute little to the incidence data in current datasets. In the United States, about 500 new cases of Wilms' tumors are diagnosed each year. This number has been fairly stable for many years. About 6% of all pediatric cancers are Wilms' tumors. The average age at diagnosis is about 3 years. It becomes less common as children grow older and is

uncommon after age 6. So far research has not found any strong links between Wilms' tumor and environmental factors, either during a mother's pregnancy or after a child's birth. The known mutations genes in Wilms' tumors are tumor suppressor genes WT1, WTX and TP53 and an oncogene CTNNB1. Loss of heterozygosity (LOH) or loss of imprinting (LOI) on chromosome 11p15 are also observed in tumor cases. Most Wilms' tumors are not inherited. Instead, they seem to be the result of gene changes that occur early in a child's life, perhaps even before birth. Since the genes discussed above are not altered in all Wilms' tumor cases, there must be changes in other genes that have not yet been found. In many cases, more than one gene change is probably involved. (see [13], [14], [15]).

Kidney cancers in adults are renal cell carcinomas (RCCs), which arise from cells in the tubules of the filtration portion of the kidney. They account for approximately 3% of adult malignancies. The incidence of renal cell cancer has been rising steadily. The vast majority of cases are diagnosed in patients over 65. RCCs exhibit unique genetic abnormalities and differ from Wilms' tumor in biology. Most RCCs (about 80%) are classified as clear cell RCC. Other two common subtypes of RCCs are papillary RCC (about 15%) and chromophobe RCC (about 5%). When seen under a microscope, the cells that make up clear cell RCC look very pale or clear. Papillary RCC form little finger-like projections (called papillae) in some of the tumor. The cells of Chromophobe RCC are also pale, like the clear cells, but are much larger and have certain other features that can be recognized [16], [17]. The cause of RCC is not known. Epidemiologic evidence indicates that age beyond 50 years, male gender, and end-stage renal disease are risk factors for developing renal cell carcinoma. Other risk factors include smoking, obesity, hypertension and exposures to certain substances. Despite these associations, no definite causal relationship has been established. While the specific causes of kidney cancer are unknown, genetic abnormalities are consistently present in each histologic subtype.

Different subtypes of RCCs undergo genetic or epigenetic changes and are developed by different pathways.

CHAPTER II

A NEW STOCHASTIC MODEL OF PEDIATRIC KIDNEY CANCER-WILMS' TUMOR

A. Introduction

Wilms' tumor, also known as nephroblastoma, is the most common kidney cancer in children. Wilms' tumor occurs of 1 in 10,000 live births, accounting for 6% of childhood cancers and rating fourth in overall incidence among childhood cancers. In the United States, about 500 children are diagnosed with Wilms' tumor each year. The tumor most often affects children between the ages of 2 and 4 years. About 95% of cases diagnosed before the age of 10 years. Although Wilms' tumor can develop in both kidneys (called bilateral), it usually occurs in only one (unilateral). (see [13], [14], [15]).

In 1972, Knudson and Strong proposed a two-stage model for carcinogenesis of human Wilms' tumor [18]. They suggested that the development of Wilms' tumor may require two independent rate-limiting genetic events. In this model, people with bilateral tumors or a family history of cancer carry a germline mutation in one allele of a tumor-suppressor gene and need only one more genetic event to develop Wilms' tumors. People with unilateral tumors and no family history require two independent somatic mutations to develop tumors. Unlike the development of retinoblastoma, which results from the inactivation of one single tumor-suppressor gene RB, the development of Wilms' tumor is complex and likely to involve several genes by multistage and multiple pathways. In this chapter, we develop a stochastic model for human Wilms' tumor with a generalized mixture model to account for hereditary cancer cases and different pathways for non-hereditary cancer cases.

For developing a biologically supported stochastic model of carcinogenesis for Wilms' tumor, in Section B, we provide a brief summary of Wilms' tumor biology. In Section C, we present a biologically supported stochastic model of Wilms' tumor incorporating hereditary cancer cases. In Section D, we develop a statistical model for

cancer incidence data. This statistical model is basically a generalized mixture model of Poisson distributions with two mixture probability distributions. One mixture probability distributions is a bivariate multinomial distribution, which is used to account for individuals with different genotypes at embryo stage in the population. The other one is a binomial distribution which is applied to represent individuals who are normal at embryo stage at risk of Wilms' tumor by different pathways. In Section E, by using results from Section B-D, we develop a generalized Bayesian inference procedure and multi-level Gibbs sampling method to estimate unknown parameters. To illustrate the applications of the model and methods, in Section F, we apply the models and methods to the SEER Wilms' tumor incidence data. In Section G, we present the computation details of fitting the model of Wilms' tumor. Finally in Section H, we discuss the usefulness of the model and methods and provide some conclusions.

B. A Summary of Wilms' Tumour Biology

Wilms' tumor, an embryonal kidney tumor that consists of undifferentiated mesenchymal cells, poorly organized epithelium, and surrounding stromal cells, is known to be genetically heterogeneous [19]. The known mutations genes in Wilms' tumors are tumor suppressor genes WT1, WTX and TP53 and an oncogene CTNNB1 [20], [21], [22], [23].

Mutations of the WT1 gene on chromosome 11p13 are observed in about 20% of Wilms' tumors and are an early event in tumorigenesis [24]. Germline mutations in WT1 have been identified in children with WAGR syndrome, Denys-Drash syndrome and Frasier syndrome. Somatic mutations in WT1 have also been observed in people with Wilms' tumor. Loss of WT1 function alters normal differentiation of the induced nephrogenic mesenchyme. WT1 mutation alone may be insufficient to develop tumors. One or more additional rate-limiting genetic alterations is required for carcinogenesis [25].

Somatic mutations of the CTNNB1 gene on chromosome 3p22 occur in about 15% of tumor cases and are at a later step in tumorigenesis [26]. A highly significant association

of CTNNB1 mutations with WT1 mutations has been observed. These mutations lead to deregulation of the Wnt-β-catenin signaling pathway, resulting in aberrant control of cellular proliferation in the mesenchymal cells.

Somatic mutations of the WTX gene on the X chromosome are observed in up to 30$\%$ of Wilms' tumor cases. In contrast to biallelic inactivation of autosomal tumor suppressor genes, WTX is altered by a monoallelic a single event targeting the single X chromosome in males and the active X chromosome in females. WTX mutations encodes a protein that forms a complex with β-catenin and other proteins, ultimately promoting ubiquitination and degradation of β-catenin, thereby attenuating TCF-mediated transcription. WTX mutations in Wilms' tumor are negatively correlated with mutations in WT1. WTX alteration alone may not results in tumors [21].

Both germline and somatic mutations of P53 gene on chromosome 17p occur in about 5$\%$ of tumor cases, which associated with anaplastic Wilms' tumor [13].

Additionally, loss of heterozygosity (LOH) or loss of imprinting (LOI) on chromosome 11p15, which harbors a cluster of imprinted genes (also referred to as WT2), is observed in approximately 70$\%$ of tumor cases, resulting in biallelic expression of IGF2 [27], [28]. A recent study on mouse data suggested a model for Wilms' tumor in which WT1 ablation and IGF2 unpregulation are critical genetic events [29]. Loss of WT1 function alters normal differentiation of the induced nephrogenic mesenchyme and upregulation of IGF2 drive the proliferation of these abnormal cells through IGF-IR signaling transduced via pIRS1 and pERK1/2.

Based on the above genetics and molecular biology, we propose a multi-stage stochastic model involving hereditary and non-hereditary cancer cases, in which a 3-stage model is used to illuminate the development of hereditary cancers and some non-hereditary cancers, and a 2-stage model is applied to explain other non-hereditary cancers. WT1 mutation, IGF2 upregulation, CTNNB1 mutation and P53 mutation may involve in carcinogenesis by the 3-stage model ($N \rightarrow I_1 \rightarrow I_2 \rightarrow I_3 \rightarrow$ Tumor), whereas

WTX mutation and other unknown genetic event may contribute to carcinogenesis by the 2-stage model ($N \rightarrow J_1 \rightarrow J_2 \rightarrow$ Tumor).

C. A Biologically Supported Stochastic Model of Wilms' Tumor Incorporating Hereditary Cancer Cases and Involving Multiple Pathways

Given the biology of Wilms' tumor, we observe that both germline cells (egg and sperm) and somatic cells may carry mutant alleles of cancer genes such as WT1 gene and P53 gene. In population at risk of developing Wilms' tumor, based on genetic makeup at the embryo stage, people can be classified into three groups: Normal people ($N = I_0$ people), I_1 people and I_2 people. If both egg and sperm generating the individual carry mutant alleles of relevant cancer genes, then this individual is a I_2 stage person at the embryo stage, in which case with high probability the individual is born with cancer. If only one of germ line cells (egg or sperm) generating the individual carries mutant alleles of cancer genes, then the individual is a I_1 stage person at the embryo stage. If no germline cells generating the individual carry mutant alleles of cancer genes, this individual is a normal person.

To account for inherited cancer cases, we let p_i $(i = 1, 2)$ be the proportion of I_i $(i = 1, 2)$ person in the population. Then $p_0 = 1 - p_1 - p_2$ is the proportion of normal person in the population. In general large human populations, under the assumption that there are no new mutation in cancer genes one may practically assume that p_i $(i = 0, 1, 2)$ is constant [30]. Let n_j denote the number of people at risk of developing Wilms' tumor during the $j-$th age period $[t_{j-1}, t_j)$. Among the n_j people in the $j-$th age period, let n_{ij} $(i = 1, 2)$ be the number of I_i $(i = 1, 2)$ people and n_{0j} be the number of normal people ($n_{0j} = n_j - n_{1j} - n_{2j}$). Based on the Hardy-Weinberg law, under assumptions that mating between individuals in the population is random and the population size is very large [31], [30], the conditional probability distribution of $\{n_{1j}, n_{2j}\}$ given n_j is multinomial with parameters $\{n_j; p_1, p_2\}$; that is,

$\{n_{1j}, n_{2j}\}|n_j \sim$ Multinomial$\{n_j; p_1, p_2\}$.

Since the proliferation rates of all stem cells are large during pregnancy, I_2 people in the population may acquire additional genetic and/or epigenetic changes to become I_3 stage people before birth and hence develop cancer at birth. Similarly, I_1 people may acquire genetic and/or epigenetic changes during pregnancy to become I_2 people at birth. Because of protection at the embryo stage and during pregnant period, normal people at the embryo stage would remain to be normal people at birth. This model is represented schematically in Figure 1, where α_1 is the probability of a $I_2 \to I_3$ transition in I_2 people during pregnancy.

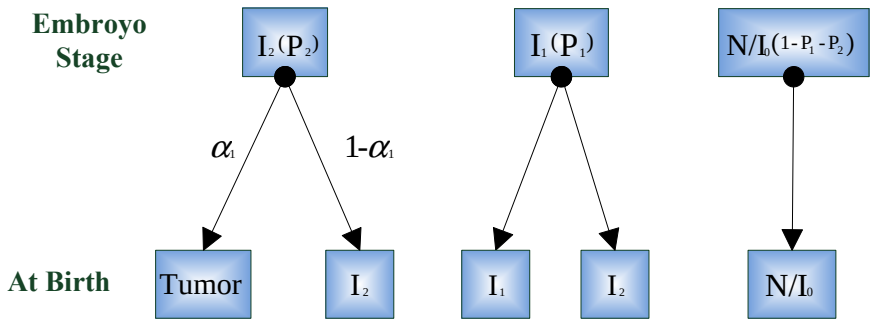

Fig. 1: Embryo Genotypes and Their Frequencies at Embryo Stage and at Birth

Based on studies of molecular biology given in Section B, we assume a stochastic model for human Wilms' tumor involving hereditary and non-hereditary cancer cases, in which a 3-stage model is used to illuminate hereditary cancers and some non-hereditary cancers, and a 2-stage model is applied to explain other non-hereditary cancers. For people who are normal at the embryo stage, Wilms' tumor is derived through two pathways: $N\ (I_0) \to I_1 \to I_2 \to I_3 \to Tumor$ and $N \to J_1 \to J_2 \to Tumor$. Let α_2 be the proportion of people in normal people at risk of developing Wilms' tumor through the 2-stage model. Thus $1 - \alpha_2$ is the proportion of people in normal people at risk of developing Wilms' tumor through the 3-stage model. Among the n_{0j} normal people at risk of cancer in the $j-$th age period, let $n_{0j}^{(I)}$ be the number of people at risk of developing tumor by 3-stage model and $n_{0j}^{(J)}$ the number of people at risk of developing

tumor by 2-stage model. Based on the Hardy-Weinberg law again, the conditional probability distribution of $n_{0j}^{(J)}$ given n_{0j} is binomial with parameters $\{n_{0j}; \alpha_2\}$; that is, $n_{0j}^{(J)}|n_{0j} \sim$ Binomial$\{n_{0j}; \alpha_2\}$. For people who have genotype I_1 at the embryo stage, Wilms' tumor is developed through $I_1 \rightarrow I_2 \rightarrow I_3 \rightarrow Tumor$. For people who have genotype I_2 at the embryo stage, Wilms' tumor is derived through $I_2 \rightarrow I_3 \rightarrow Tumor$. The proposed model for Wilms' tumor can be represented schematically by Figure 2.

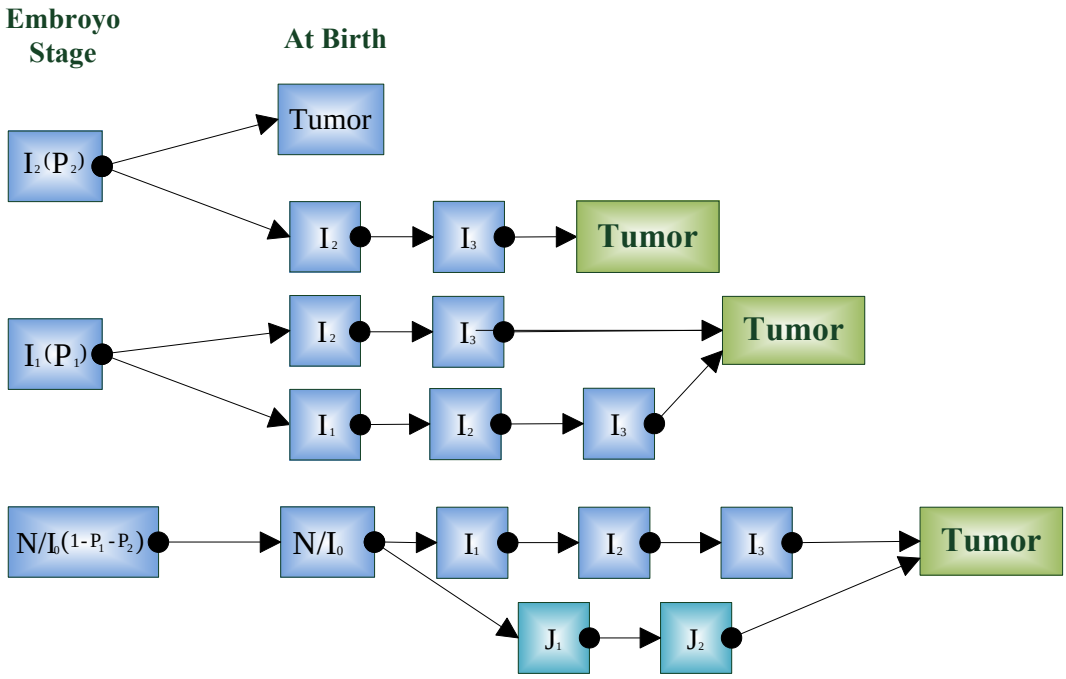

Fig. 2: Multiple-pathway for Wilms' Tumor Development

The model assumes the I_i and J_i cells are subjected to stochastic proliferation (birth) and differentiation (death). It takes into account cancer progression by following to postulate cancer tumors derive from primary I_3 (J_2) by clonal expansion [32], where primary I_3 (J_2) cells are I_3 (J_2) cells generated directly by I_2 (J_1) cells by genetic or epigenetic changes. The model also postulate that all cells proceed forward independently of other cells. The I_3 stage and J_2 stage are transient stages to cancer tumors, hence the state variables for the model are the number $I_j(t)$ ($J_k(t)$) of I_j (J_k) cells at time t for

$j = 0, 1, 2$ ($k = 0, 1$) and the number $T(t)$ of cancer tumors at time t. Therefore, the stochastic processes are $\{I_j(t), j = 0, 1, 2, J_k(t), k = 0, 1, T(t), t > 0\}$ for normal people in the population, $\{I_j(t), j = 1, 2, T(t), t > 0\}$ for I_1 people in the population, and $\{I_2(t), T(t), t > 0\}$ for I_2 people in the population. Notice that, to develop stochastic models of carcinogenesis, it is conveniently assumed that the last stage cells (i.e. I_3 cells, J_2 cells) grow instantaneously into cancer tumors as soon as they are generated as shown in [1], [8], [33]. In this case, one may assume $T(t)$ as Markov.

To develop mathematical theories for stochastic processes of carcinogenesis for Wilms' tumor, let $I_i(t; u)$ $(u = 0, 1, 2,\ i = u, \ldots, 2)$ denote the number of I_i $(i = u, \ldots, 2)$ cells at time t in people who are I_u people at the embryo stage and $J_1(t)$ the number of J_1 cells at time t in people who are normal people at the embryo stage. Let $Q_i(j)$ denote the probability of developing tumor during the $j-$th age period $[t_{j-1}, t_j)$ $(t_j > t_0)$ in people who are I_i $(i = 0, 1, 2)$ people at the embryo stage. Let $Q_0^{(I)}(j)$ ($Q_0^{(J)}(j)$) denote the probability of normal people developing tumor through 3-stage model (2-stage model) during the $j-$th age period $[t_{j-1}, t_j)$ $(t_j > t_0)$.

Let $\beta_j^{(I)}(t)$ denote the transition rate from $I_j \to I_{j+1}$ $(j = 0, 1, 2)$ at time t and $\beta_k^{(J)}(t)$ the transition rate from $J_k \to J_{k+1}$ $(k = 0, 1)$ at time t. In many practical problems, $\beta_2^{(I)}(t)$ and $\beta_1^{(J)}(t)$ are very small ($10^{-8} \sim 10^{-4}$) and one may assume that $\beta_2^{(I)}(t) = \beta_2^{(I)}$ and $\beta_1^{(J)}(t) = \beta_1^{(J)}$. Then, by using methods in Tan [34], Tan el al. [35], [2] and Tan and Yan [10], it can be shown that $\{Q_0^{(I)}(j), Q_0^{(J)}(j), Q_i(j), i = 1, 2,\ j \geq 1\}$ are given respectively by:

$$
\begin{aligned}
Q_0^{(I)}(j) &= \{e^{-\beta_2^{(I)} \int_{t_0}^{t_{j-1}} E[I_2(x;0)] P_T^{(I)}(x, t_{j-1}) dx} - e^{-\beta_2^{(I)} \int_{t_0}^{t_j} E[I_2(x;0)] P_T^{(I)}(x, t_{j-1}) dx}\} + o(\beta_2^{(I)}) \\
Q_0^{(J)}(j) &= \{e^{-\beta_1^{(J)} \int_{t_0}^{t_{j-1}} E[J_1(x)] P_T^{(J)}(x, t_{j-1}) dx} - e^{-\beta_1^{(J)} \int_{t_0}^{t_j} E[J_1(x)] P_T^{(J)}(x, t_{j-1}) dx}\} + o(\beta_1^{(J)}) \\
Q_1(j) &= \{e^{-\beta_2^{(I)} \int_{t0}^{t_{j-1}} E[I_2(x;1)] P_T^{(I)}(x, t_{j-1}) dx} - e^{-\beta_2^{(I)} \int_{t_0}^{t_j} E[I_2(x;1)] P_T^{(I)}(x, t_{j-1}) dx}\} + o(\beta_2^{(I)}) \\
Q_2(j) &= (1 - \alpha_1)\{e^{-\beta_2^{(I)} \int_{t_0}^{t_{j-1}} E[I_2(x;2)] P_T^{(I)}(x, t_{j-1}) dx} - e^{-\beta_2^{(I)} \int_{t_0}^{t_j} E[I_2(x;2)] P_T^{(I)}(x, t_{j-1}) dx}\} \\
&+ o(\beta_2^{(I)})
\end{aligned}
$$

Where $E[I_2(x;i)]$ $(i=0,1,2)$ $(E[J_1(x)])$ is the expected number of $I_2(t;i)$ $(J_1(t))$ and where $P_T^{(I)}(s,t)$ $(P_T^{(J)}(s,t))$ is the probability that a primary I_3 (J_2) cell generated from an I_2 (J_1) cell at time s develops into a detectable tumor by time t. We will derive $E[I_2(t;i)]$ and $E[J_1(t)]$ in the following subsection.

The Stochastic Model and Mathematical Analysis

For deriving mathematical analysis of the above model involving hereditary and non-hereditary cancer cases, let $b_j^{(I)}(t)$ and $d_j^{(I)}(t)$ denote the birth rate and the death rate at time t of the I_j $(j=0,1,2)$ cells respectively. Let $\{B_j^{(I)}(t;i), D_j^{(I)}(t;i), M_j^{(I)}(t;i)\}$ be the number of birth and the number of death of I_j cells, and the number of transition from $I_j \to I_{j+1}$ cells during $[t, t+\Delta t)$ respectively in people who are I_i people at the embryo stage. Similarly, let $b_k^{(J)}(t)$ and $d_k^{(J)}(t)$ denote the birth rate and the death rate at time t of the J_k $(k=0,1)$ cells. Let $\{B_k^{(J)}(t), D_k^{(J)}(t), M_k^{(J)}(t)\}$ be the number of birth and the number of death of J_k cells, and the number of transition from $J_k \to J_{k+1}$ cells during $[t, t+\Delta t)$ respectively in people who are normal people at the embryo stage. Then, to order of $o(\beta_j^{(I)}(t)\Delta t)$ and $o(\beta_k^{(J)}(t)\Delta t)$,

$$
\begin{aligned}
\{B_j^{(I)}(t;i), D_j^{(I)}(t;i)\}|I_j(t;i) &\sim \text{Multinomial}\{I_j(t;i); b_j^{(I)}(t)\Delta t, d_j^{(I)}(t)\Delta t\} \qquad (1)\\
&\quad j=0,1,2\\
\{B_k^{(J)}(t), D_k^{(J)}(t)\}|J_k(t) &\sim \text{Multinomial}\{J_k(t); b_k^{(J)}(t)\Delta t, d_k^{(J)}(t)\Delta t\} \qquad (2)\\
&\quad k=0,1\\
M_j^{(I)}(t;i)|I_j(t;i) &\sim \text{Binomial}\{I_j(t;i); \beta_j^{(I)}(t)\Delta t\} \qquad (3)\\
&\sim \text{Poisson}\{I_j(t;i)\beta_j^{(I)}(t)\Delta t\} + o(\beta_j^{(I)}(t)\Delta t)\\
&\quad \text{independently of } \{B_j^{(I)}(t;i), D_j^{(I)}(t;i)\}, j=0,1,2\\
M_k^{(J)}(t)|J_k(t) &\sim \text{Binomial}\{J_k(t); \beta_k^{(J)}(t)\Delta t\} \qquad (4)\\
&\sim \text{Poisson}\{J_k(t)\beta_k^{(J)}(t)\Delta t\} + o(\beta_k^{(J)}(t)\Delta t)\\
&\quad \text{independently of } \{B_k^{(J)}(t), D_k^{(J)}(t)\}, k=0,1
\end{aligned}
$$

For an individual who is an N (I_0) people at the embryo stage, the state variables in this individual are $\{I_j(t), j = 0, 1, 2, J_k(t), k = 0, 1, T(t)\}$; for an individual who is an I_1 people at the embryo stage, the state variables in this individual are $\{I_j(t), j = 1, 2, T(t)\}$; for an individual who is an I_2 people at the embryo stage, the staging variables in this individual are $\{I_2(t), T(t)\}$. To derive $E[I_2(t; i)]$ $(i = 0, 1, 2)$, $E[J_1(t)]$ and the probability distributions of these state variables, we have the following stochastic equations for $\{I_j(t; i), j = i, \ldots, 2, i = 0, 1, 2, J_k(t), k = 0, 1\}$ (see [34], [2], [35], [36], [37]):

$$
\begin{aligned}
I_i(t + \Delta t; i) &= I_i(t; i) + B_i^{(I)}(t; i) - D_i^{(I)}(t; i),\ i = 0, 1, 2 && (5)\\
I_j(t + \Delta t; u) &= I_j(t; u) + B_j^{(I)}(t; u) - D_j^{(I)}(t; u) + M_{j-1}^{(I)}(t; u) && (6)\\
&\quad u = 0, 1, j = u + 1, \ldots, 2 \\
J_1(t + \Delta t) &= J_1(t) + B_1^{(J)}(t) - D_1^{(J)}(t) + M_0^{(J)}(t) && (7)
\end{aligned}
$$

Given the probability distributions of the random transition variables in equations (1)-(4) and the stochastic equations in equation (5)-(7), we derive the following stochastic differential equations for the state variables
$\{I_j(t; i), i = 0, 1, 2, j = i, \ldots, 2, J_k(t), k = 0, 1\}$

$$
\begin{aligned}
\frac{d}{dt} I_i(t; i) &= I_i(t; i)\gamma_i^{(I)}(t) + e_i^{(I)}(t; i),\ i = 0, 1, 2 && (8)\\
\frac{d}{dt} I_j(t; u) &= I_j(t; u)\gamma_j^{(I)}(t) + I_{j-1}(t; u)\beta_{j-1}^{(I)}(t) + e_j^{(I)}(t; u) && (9)\\
&\quad u = 0, 1, j = u + 1, \ldots, 2 \\
\frac{d}{dt} J_1(t) &= J_1(t)\gamma_1^{(J)}(t) + J_0(t)\beta_0^{(J)}(t) + e_1(t) && (10)
\end{aligned}
$$

Where $\gamma_u^{(I)}(t) = b_u^{(I)}(t) - d_u^{(I)}(t)$ for $u = 0, 1, 2$, $\gamma_1^{(J)}(t) = b_1^{(J)}(t) - d_1^{(J)}(t)$ and the random noises are:

$$
\begin{aligned}
e_i^{(I)}(t;i)\Delta t &= [B_i^{(I)}(t;i) - I_i(t;i)b_i^{(I)}(t)\Delta t] - [D_i^{(I)}(t;i) - I_i(t;i)d_i^{(I)}(t)\Delta t],\ i=0,1,2,\\
e_j^{(I)}(t;u)\Delta t &= [B_j^{(I)}(t;u) - I_j(t;u)b_j^{(I)}(t)\Delta t] - [D_j^{(I)}(t;u) - I_j(t;u)d_j^{(I)}(t)\Delta t]\\
&+ [M_{j-1}^{(I)}(t;u) - I_{j-1}(t;u)\beta_{j-1}^{(I)}(t)\Delta t],\ u=0,1, j=u+1,\ldots,2\\
e_1^{(J)}(t)\Delta t &= [B_1^{(J)}(t) - J_1(t)b_1^{(J)}(t)\Delta t] - [D_1^{(J)}(t) - J_1(t)d_1^{(J)}(t)\Delta t]\\
&+ [M_0^{(J)}(t) - J_0(t)\beta_0^{(J)}(t)\Delta t]
\end{aligned}
$$

From the above equations, the random noises have expectation zero and are un-correlated with the state variables. Given the initial conditions $I_j(t_0;i) > 0,\ j = i, i+1$ and $I_j(t_0;i) = 0,\ j > i+1$, the solution of the above equation (8)-(10) are given respectively by:

$$
\begin{aligned}
I_i(t;i) &= I_i(t_0;i)e^{\int_{t_0}^t \gamma_i^{(I)}(x)dx} + \eta_i^{(I)}(t;i),\ i=0,1,2 \qquad (11)\\
I_j(t;u) &= I_j(t_0;u)e^{\int_{t_0}^t \gamma_j^{(I)}(x)dx} + \int_{t_0}^t I_{j-1}(x;u)\beta_{j-1}^{(I)}(x)e^{\int_x^t \gamma_j^{(I)}(y)dy}dx + \eta_j^{(I)}(t;u),\\
&\quad u=0,1, j=u+1,\ldots,2 \qquad (12)\\
J_1(t) &= J_1(t_0)e^{\int_{t_0}^t \gamma_1^{(J)}(x)dx} + \int_{t_0}^t J_0(x)\beta_0^{(J)}(x)e^{\int_x^t \gamma_1^{(J)}(y)dy}dx + \eta_1^{(J)}(t) \qquad (13)
\end{aligned}
$$

where $\eta_j^{(I)}(t;i) = \int_{t_0}^t e^{\int_x^t \gamma_j^{(I)}(y)dy} e_j^{(I)}(x;i)dx$ and $\eta_1^{(J)}(t) = \int_{t_0}^t e^{\int_x^t \gamma_1^{(J)}(y)dy} e_1^{(J)}(x)dx$.

Obviously, $E[\eta_j(t;i)] = 0$ $(j = 1,2, i = 0,1,2)$. Given $\{I_0(t_0;0) = N(t_0)$, $I_1(t_0;0) = I_2(t_0;0) = 0, J_0(t_0) = N(t_0), J_1(t_0) = 0, \gamma_0^{(I)}(x) = 0, \gamma_0^{(J)}(x) = 0\}$, the expected numbers $E[I_j(t;i)]$ $(j = 1,2, i = 0,1,2)$ of I_j $(j = 1,2)$ and $E[J(t)]$ of J_1 are:

$$
\begin{aligned}
E[I_2(t;2)] &= E[I_2(t_0;2)]e^{\int_{t_0}^{t}\gamma_2^{(I)}(x)dx} \\
E[I_2(t;1)] &= E[I_2(t_0;1)]e^{\int_{t_0}^{t}\gamma_2^{(I)}(x)dx} + E[I_1(t_0;1)]\int_{t_0}^{t}\beta_1^{(I)}(x)e^{\int_{t_0}^{x}\gamma_1^{(I)}(z)dz+\int_{x}^{t}\gamma_2^{(I)}(z)dz}dx \\
E[I_2(t;0)] &= \int_{t_0}^{t}\beta_1^{(I)}(x)E[I_1(x;0)]e^{\int_{x}^{t}\gamma_2^{(I)}(y)dy}dx \\
E[J_1(t)] &= E[N(t_0)]\int_{t_0}^{t}\beta_0^{(J)}(x)e^{\int_{x}^{t}\gamma_1^{(J)}(y)dy}dx
\end{aligned}
$$

If the model is time homogeneous so that $\{b_j^{(I)}(t) = b_j^{(I)}, d_j^{(I)}(t) = d_j^{(I)}, \gamma_j^{(I)}(t) = \gamma_j^{(I)}, \beta_j^{(I)}(t) = \beta_j^{(I)}, j = 0, 1, 2, b_j^{(I)}(t) = b_j^{(I)}, d_1^{(J)}(t) = d_1^{(J)}, \gamma_1^{(J)}(t) = \gamma_1^{(J)}, \beta_0^{(J)}(t) = \beta_0^{(J)}, \}$ and if $\gamma_i^{(I)} \neq \gamma_j^{(I)}$ for $i \neq j$, then the above expected numbers reduce to:

$$
\begin{aligned}
E[I_2(t;2)] &= E[I_2(t_0;2)]e^{\gamma_2^{(I)}(t-t_0)} && (14) \\
E[I_2(t;1)] &= E[I_2(t_0;1)]e^{\gamma_2^{(I)}(t-t_0)} + E[I_1(t_0;1)]\beta_1^{(I)}\sum_{u=1}^{2}A_{12}(u)e^{\gamma_u^{(I)}(t-t_0)} && (15) \\
E[I_2(t;0)] &= E[N(t_0)]\beta_0^{(I)}\beta_1^{(I)}\sum_{u=1}^{2}A_{12}(u)\frac{1}{\gamma_u^{(I)}}[e^{\gamma_u^{(I)}(t-t_0)} - 1], && (16) \\
E[J_1(t;0)] &= E[N(t_0)]\beta_0^{(J)}\frac{1}{\gamma_1^{(J)}}[e^{\gamma_u^{(I)}(t-t_0)} - 1], && (17)
\end{aligned}
$$

where $A_{ij}(u) = \prod_{v=i,v\neq u}^{j}(\gamma_u - \gamma_v)^{-1}$ for $i \leq u \leq j$.

According to the above results, we can derive $\{Q_0^{(I)}(j), Q_0^{(J)}(j), Q_i(j), i = 1, 2, j \geq 1\}$ for homogeneous models under the condition that $\gamma_1^{(I)} \neq \gamma_2^{(I)}$:

$$
\begin{aligned}
Q_0^{(I)}(j) &= \{e^{-\lambda_3\psi_{02}(t_{j-1})} - e^{-\lambda_3\psi_{02}(t_j)}\} + o(\beta_2^{(I)}), \\
Q_0^{(J)}(j) &= \{e^{-\lambda_4\psi_{01}(t_{j-1})} - e^{-\lambda_4\psi_{01}(t_j)}\} + o(\beta_1^{(J)}), \\
Q_1(j) &= \{e^{-\theta\psi_{22}(t_{j-1})-\lambda_2\psi_{12}(t_{j-1})} - e^{-\theta\psi_{22}(t_j)-\lambda_2\psi_{12}(t_j)}\} + o(\beta_2^{(I)}), \\
Q_2(j) &= (1-\alpha_1)\{e^{-\lambda_1\psi_{22}(t_{j-1})} - e^{-\lambda_1\psi_{22}(t_j)}\} + o(\beta_2^{(I)}).
\end{aligned}
$$

Where $\lambda_1 = \frac{1}{\gamma_2^{(I)}} E[I_2(t_0;2)]\beta_2^{(I)}$, $\lambda_2 = \{\prod_{i=1}^{2}\gamma_i^{(I)}\}^{-1} E[I_1(t_0;1)]\beta_1^{(I)}\beta_2^{(I)}$,
$\lambda_3 = \{\prod_{i=1}^{2}\gamma_i^{(I)}\}^{-1}\, E[N(t_0)]\prod_{i=0}^{2}\beta_i^{(I)}$, $\theta = \frac{1}{\gamma_2^{(I)}} E[I_2(t_0;1)]\beta_2^{(I)}$,
$\lambda_4 = \frac{1}{\gamma_1^{(J)^2}} E[N(t_0)]\beta_0^{(J)}\beta_1^{(J)}$ and

$$
\begin{aligned}
\psi_{22}(t) &= \gamma_2^{(I)} \int_{t_0}^{t} e^{\gamma_2^{(I)}(x-t_0)} P_T^{(I)}(x,t)dx \\
&= \{e^{\gamma_2^{(I)}(t-t_0)} - 1\} \text{ if } P_T^{(I)}(x,t) = 1 \text{ for } \ t > x \\
\psi_{12}(t) &= \{\prod_{i=1}^{2}\gamma_i^{(I)}\} \sum_{u=1}^{2} A_{12}(u) \int_{t_0}^{t} e^{\gamma_u^{(I)}(x-t_0)} P_T^{(I)}(x,t)dx \\
&= \{\prod_{i=1}^{2}\gamma_i^{(I)}\} \sum_{u=1}^{2} A_{12}(u) \frac{1}{\gamma_u} \{e^{\gamma_u^{(I)}(t-t_0)} - 1\} \text{ if } P_T^{(I)}(x,t) = 1 \text{ for } t > x \\
\psi_{02}(t) &= \{\prod_{i=1}^{2}\gamma_i^{(I)}\} \sum_{u=1}^{2} A_{12}(u) \frac{1}{\gamma_u^{(I)}} \int_{t_0}^{t} \{e^{\gamma_u^{(I)}(x-t_0)} - 1\} P_T^{(I)}(x,t)dx \\
&= \{\prod_{i=1}^{2}\gamma_i^{(I)}\} \sum_{u=1}^{2} A_{12}(u) \frac{1}{{\gamma_u^{(I)}}^2} \{e^{\gamma_u^{(I)}(t-t_0)} - 1 - \gamma_u^{(I)}(t-t_0)\}, \\
&\quad \text{if } P_T^{(I)}(x,t) = 1 \text{ for } t > x \\
\psi_{01}(t) &= {\gamma_1^{(J)}}^2 \int_{t_0}^{t} \{e^{\gamma_1(x-t_0)} - 1\} P_T^{(J)}(x,t)dx \\
&= \{e^{\gamma_1^{(J)}(t-t_0)} - 1 - \gamma_1^{(J)}(t-t_0)\} \text{ if } P_T^{(J)}(x,t) = 1 \text{ for } t > x
\end{aligned}
$$

The Transition Probability of State Variables

Let $g(x,y;N,p_1,p_2)$ denote the density at (x,y) of a multinomial distribution with parameters (N,p_1,p_2) and $f(x;N,p)$ denote the density at x of a binomial distribution with parameters (N,p). From equation (1)-(4), we obtain the transition probability of the markov process of state variables as, to order of $o(\Delta t)$ and for $t > t_0$:

$$
P\{I_2(t+\Delta t;2) = v | I_2(t;2) = u\} = \sum_{r=0}^{u} f(r;u,b_2^{(I)}(t)\Delta t) f(u-v+r;u-r,\frac{d_2(t)\Delta t}{1-b_2^{(I)}(t)\Delta t}).
$$

$$
P\{I_1(t+\Delta t;1) = v_1, I_2(t+\Delta t;1) = v_2 | I_1(t;1) = u_1, I_2(t;1) = u_2\} =
$$

$$
P\{I_1(t+\Delta t;1) = v_1 | I_1(t;1) = u_1\} P\{I_2(t+\Delta t;1) = v_2 | I_1(t;1) = u_1, I_2(t;1) = u_2\};
$$

$P\{I_0(t+\Delta t;0)=v_0, I_1(t+\Delta t;0)=v_1, I_2(t+\Delta t;0)=v_2|I_0(t;0)=u_0, I_1(t;0)=u_1, I_2(t;0)=u_2\} = P\{I_0(t+\Delta t;0)=v_0|I_0(t;0)=u_0\}\prod_{i=1}^{2} P\{I_i(t+\Delta t;0)=v_i|I_{i-1}(t,0)=u_{i-1}, I_i(t;0)=u_i\}$;

$P\{J_0(t+\Delta t)=v_0, J_1(t+\Delta t)=v_1, |J_0(t)=u_0, J_1(t)=u_1\} = P\{J_0(t+\Delta t)=v_0|J_0(t)=u_0\}P\{J_1(t+\Delta t)=v_1|J_0(t)=u_0, J_1(t)=u_1\}$, where

$$
\begin{aligned}
& P\{I_1(t+\Delta t;1)=v_1|I_1(t;1)=u_1\} \\
= & \sum_{i=0}^{u_1} f(i;u_1,b_1^{(I)}(t)\Delta t)\times f(u_1-v_1+i;u_1-i,\frac{d_1^{(I)}(t)\Delta t}{1-b_1^{(I)}(t)\Delta t}), \\
& P\{I_2(t+\Delta t;1)=v_2|I_1(t;1)=u_1, I_2(t;1)=u_2\} \\
= & \sum_{i=0}^{u_2}\sum_{j=0}^{u_2-i} g(i,j;u_2,b_2^{(I)}(t)\Delta t,d_2^{(I)}(t)\Delta t)\times f(v_2-u_2-i+j;u_1,\beta_1^{(I)}(t)\Delta t). \\
& P\{I_0(t+\Delta t;0)=v_0|I_0(t;0)=u_0\} \\
= & \sum_{i=0}^{u_0} f(i;u_0,b_0^{(I)}(t)\Delta t)\times f(u_0-v_0+i;u_0-i,\frac{d_0^{(I)}(t)\Delta t}{1-b_0^{(I)}(t)\Delta t}), \\
& P\{I_i(t+\Delta t;0)=v_i|I_{i-1}(t;0)=u_{i-1}, I_i(t;0)=u_i\} \\
= & \sum_{j=0}^{u_i}\sum_{m=0}^{u_i-j} g(j,m;u_i,b_i^{(I)}(t)\Delta t,d_i^{(I)}(t)\Delta t)\times f(v_i-u_i-j+m;u_{i-1},\beta_{i-1}^{(I)}(t)\Delta t), \\
& i=1,2,
\end{aligned}
$$

$$
\begin{aligned}
& P\{J_1(t+\Delta t)=v_1|J_0(t)=u_0, J_1(t)=u_1\} \\
= & \sum_{i=0}^{u_1}\sum_{j=0}^{u_1-i} g(i,j;u_1,b_1^{(J)}(t)\Delta t,d_1^{(J)}(t)\Delta t)\times f(v_1-u_1-j+m;u_0,\beta_0^{(J)}(t)\Delta t).
\end{aligned}
$$

D. A Statistical Model and The Probability Distribution of the Number of Detectable Tumors

The data available for modelling carcinogenesis are usually cancer incidence over different time periods. For example, the SEER data of NCI/NIH for human cancers are given by $\{(y_0,n_0),(y_j,n_j), j=1,\ldots,k\}$, where y_0 is the number of cancer cases at birth and n_0 the total number of birth and where for $j\geq 1$, y_j is the number of cancer cases during the $j-$th age group of 1 year period (or 5 years period) and n_j is the number of people who are at risk of cancer and from whom y_j of them have developed cancer during

the $j-$th age group. Given in Table 1 are the SEER data for Wilms' tumor cases. From this data set, notice that there are cancer cases at birth implying some number of inherited cancer cases. In this section, we will develop a statistical model for this data set.

To incorporate hereditary cancer cases, we have noted in the previous section that $n_j = \sum_{i=0}^{2} n_{ij}$, where n_{ij} is the number of individuals who have genotype I_i $(i = 0, 1, 2)$ at the embryo stage. Then, as showed in the previous section, the conditional probability distribution of (n_{1j}, n_{2j}) given n_j is multinomial with parameters $\{n_j; p_1, p_2\}$. It follows that $n_{ij}|n_j \sim$ Binomial$\{n_j, p_i\}, i = 1, 2$. Among the n_{0j} normal people at risk of cancer in the $j-$th age period, $n_{0j} = n_{0j}^{(I)} + n_{0j}^{(J)}$, where $n_{0j}^{(I)}$ is the number of people at risk of developing tumor by 3-stage model and $n_{0j}^{(J)}$ the number of people at risk of developing tumor by 2-stage model. As discussed in the previous section, $n_{0j}^{(J)}|n_{0j} \sim$ Binomial$\{n_{0j}; \alpha_2\}$. We let Y_j denote the random variable for y_j and y_j be the observed number of Y_j.

The Probability Distribution of Y_0

Y_0 is the number of cancer cases at birth and y_0 derives only from individuals who have genotype I_2 at the embryo stage. Thus, given n_{20} individuals with genotype I_2 at the embryo stage, $y_0|n_{20} \sim Poisson\{n_{20}\alpha_1\}$. Since $n_{20}|n_0 \sim Binomial\{n_0, p_2\}$, obviously we have, $Y_0 \sim Poisson\{\chi_0\}$, where $\chi_0 = n_0 p_2 \alpha_1$.

The expected number of Y_0 given n_0 is $E(Y_0|n_0) = n_0 p_2 \alpha_1$. The deviance D_0 from the conditional probability distribution of y_0 given n_0 is:

$$\begin{aligned} D_0 &= -2\{\log\{h(y_0; \chi_0)\} - \log\{h(y_0; \hat{\chi}_0)\}\} \\ &= 2\{\{\chi_0 - y_0\} - y_0 \log\{\frac{\chi_0}{y_0}\}\} \end{aligned}$$

where $\hat{\chi}_0 = y_0$.

The Probability Distribution of Y_j $(j \geq 1)$

To derive the probability distribution of Y_j $(j \geq 1)$ in the $j-$th age group, let Y_{ij} $(i = 0, 1, 2)$ be the number of cancer cases generated by people with genotype I_i at the embryo stage among these Y_j cancer cases. Let $Y_{0j}^{(I)}$ be the number of cancer cases generated in normal people by 3-stage model and $Y_{0j}^{(J)}$ be the number of cancer cases generated in normal people by 2-stage model. The conditional probability distribution of Y_{ij} given n_{ij} is, for $t_j > t_0$:

$$\begin{aligned} Y_{0j}^{(I)}|n_{0j} &\sim \text{Poisson}\{n_{0j}^{(I)} Q_0^{(I)}(j)\}, \\ Y_{0j}^{(J)}|n_{0j} &\sim \text{Poisson}\{n_{0j}^{(J)} Q_0^{(J)}(j)\}, \\ Y_{ij}|n_{ij} &\sim \text{Poisson}\{n_{ij} Q_i(j)\},\ i = 1, 2. \end{aligned}$$

Then the conditional probability distribution of $\{Y_{0j}^{(I)}, Y_{0j}^{(J)}, Y_{1j}, Y_{2j}\}$ given $\{n_{0j}^{(I)}, n_{0j}^{(J)}, n_{1j}, n_{2j}\}$ is

$$\begin{aligned} & P\{y_{1j}, y_{2j}, y_{0j}^{(I)}, y_j | n_{1j}, n_{2j}, n_{0j}^{(I)}, n_j\} \\ = \ & h\{y_{0j}^{(I)}; n_{0j}^{(I)} Q_0^{(I)}(j)\} h\{y_{0j}^{(J)}; n_{0j}^{(I)} Q_0^{(I)}(j)\} \prod_{i=1}^{2} h\{y_{ij}; n_{ij} Q_i(j)\}. \end{aligned} \tag{18}$$

Put $Q_T(j) = n_{0j}^{(I)} Q_0^{(I)}(j) + n_{0j}^{(J)} Q_0^{(J)}(j) + \sum_{i=1}^{2} n_{ij} Q_i(j)$. The conditional distribution of $Y_j|(n_{ij}, i = 0, 1, 2) \sim \text{Poisson}\{Q_T(j)\}$. It follows that the probability distribution of Y_j given n_j is

$$P(y_j|n_j) = \sum_{n_{1j}=0}^{n_j} \sum_{n_{2j}=0}^{n_j - n_{1j}} g(n_{1j}, n_{2j}; n_j, p_1, p_2) \sum_{n_{0j}^{(J)}=0}^{n_{0j}} f\{n_{0j}^{(J)}; n_{0j}, \alpha_2\} h\{y_j; Q_T(j)\}, \tag{19}$$

where $g(n_{1j}, n_{2j}; n_j, p_1, p_2)$ is the probability density of $(n_{1j}, n_{2j})|n_j \sim$ Multinomial $(n_j; p_1, p_2)$, $f(n_{0j}; n_{0j}^{(J)}, \alpha_2)$ is the probability density of $n_{0j}^{(J)}|n_{0j} \sim$ Binomial$(n_{0j}; \alpha_2)$ and $h\{y_j; Q_T(j)\}$ is the Poisson density of $Y_j|(n_{ij}, i = 0, 1, 2) \sim$ Poisson $\{Q_T(j)\}$.

The probability distribution $P(y_j|n_j)$ given by equation (19) is a mixture of Poisson distributions with two mixing probability distributions given by the multinomial distribution of $\{n_{1j}, n_{2j}\}$ given n_j and the binomial distribution of $n_{0j}^{(J)}$ given n_{0j}. The multinomial mixing probability distribution represents individuals with different genotypes at the embryo stage in the population. The binomial mixing probability distribution represents individuals who are normal at embryo stage at risk of developing Wilms' tumor through different pathway.

Let Θ be the set of all unknown parameters (i.e. the parameters $(p_1, p_2, \alpha_i, i = 1, 2)$ and the birth rates, the death rates and the mutation rates of I_j cells and J_j cells). Based on data $(y_j, j = 0, 1, \ldots, k)$, the likelihood function of Θ is

$$L\{\Theta|y_j, j = 0, 1, \ldots, k\} = h(y_0; \chi_0) \prod_{j=1}^{k} P(y_j|n_j). \tag{20}$$

The Probability Distribution of the Expanded Model

For applying the mixture distribution given by equation (20) to make inference about the unknown parameters, we expand the model to include the un-observable variables $\{n_{1j}, n_{2j}, n_{0j}^{(I)}, y_{1j}, y_{2j}, y_{0j}^{(I)}\}$. To derive the joint probability distribution of these variables, observe that for $j \geq 1$, the conditional probability distribution of $\{y_{1j}, y_{2j}\}$ given $\{n_{ij}, i = 1, 2, n_j, y_j\}$ is multinomial with parameters $\{y_j; \frac{n_{1j}Q_1(j)}{Q_T(j)}, \frac{n_{2j}Q_2(j)}{Q_T(j)}\}$. That is,

$$P\{y_{1j}, y_{2j}|n_{ij}, i = 1, 2, n_j, y_j\} \sim \text{Multinomial}\{y_j; \frac{n_{1j}Q_1(j)}{Q_T(j)}, \frac{n_{2j}Q_2(j)}{Q_T(j)}\}. \tag{21}$$

The conditional probability distribution of $y_{0j}^{(I)}$ given $\{n_{0j}, y_{0j}\}$ is binomial with parameters $\{y_{0j}; \frac{n_{0j}^{(I)}Q_0^{(I)}(j)}{n_{0j}^{(I)}Q_0^{(I)}(j)+n_{0j}^{(J)}Q_0^{(J)}(j)}\}$. That is,

$$P\{y_{0j}^{(I)}|n_{0j}, y_{0j}\} \sim \text{Binomial}\{y_{0j}; \frac{n_{0j}^{(I)}Q_0^{(I)}(j)}{n_{0j}^{(I)}Q_0^{(I)}(j) + n_{0j}^{(J)}Q_0^{(J)}(j)}\}, \text{for } j \geq 1. \tag{22}$$

Hence for $j \geq 1$, the joint density of $\{n_{ij}, y_{ij}, i = 1, 2, n_{0j}^{(I)}, y_{0j}^{(I)}, y_j\}$ given n_j is

$$\begin{aligned} & P\{n_{ij}, y_{ij}, i = 1, 2, n_{0j}^{(I)}, y_{0j}^{(I)}, y_j, j = 1, \ldots, k | n_j, \Theta\} \\ = \; & g(n_{1j}, n_{2j}; n_j, p_1, p_2) f\{n_{0j}^{(J)}; n_{0j}, \alpha_2\} \\ \times \; & h\{y_{0j}^{(I)}; n_{0j}^{(I)} Q_0^{(I)}(j)\} h\{y_{0j}^{(J)}; n_{0j}^{(I)} Q_0^{(I)}(j)\} \prod_{i=1}^{2} h\{y_{ij}; n_{ij} Q_i(j)\}. \end{aligned}$$

Put $\boldsymbol{Y} = (y_{1j}, y_{2j}, y_{0j}^{(I)}, j = 1, \ldots, k)$, $\boldsymbol{N} = (n_{1j}, n_{2j}, n_{0j}^{(I)}, j = 1, \ldots, k)$, $\underset{\sim}{y} = (y_j, j = 0, 1, \ldots, k)$ and $\underset{\sim}{n} = (n_j, j = 0, 1, \ldots k\}$. For the SEER data, the joint density $P\{\boldsymbol{Y}, \underset{\sim}{y}, \boldsymbol{N} | \underset{\sim}{n}, \Theta\}$ of $\{\boldsymbol{Y}, \underset{\sim}{y}, \boldsymbol{N}\}$ given $\{\underset{\sim}{n}, \Theta\}$ is

$$\begin{aligned} P\{\boldsymbol{Y}, \underset{\sim}{y}, \boldsymbol{N} | \underset{\sim}{n}, \Theta\} \; = \; & h(y_0; n_0 p_2 \alpha_1) \\ & \times \; \prod_{j=1}^{k} \{g(n_{1j}, n_{2j}; n_j, p_1, p_2) f\{n_{0j}^{(J)}; n_{0j}, \alpha_2\} h\{y_{0j}^{(I)}; n_{0j}^{(I)} Q_0^{(I)}(j)\} \\ & \times \; h\{y_{0j}^{(J)}; n_{0j}^{(J)} Q_0^{(J)}(j)\} \prod_{i=1}^{2} h\{y_{ij}; n_{ij} Q_i(j)\}\}. \end{aligned} \tag{23}$$

The joint density $P\{\boldsymbol{Y}, \underset{\sim}{y}, \boldsymbol{N} | \underset{\sim}{n}, \Theta\}$ of $(\boldsymbol{Y}, \underset{\sim}{y}, \boldsymbol{N})$ given by equation (23) will be used as the kernel for the Bayesian method to estimate the unknown parameters and to predict the state variables.

Notice that the above distribution is a product of multinomial distributions, binomial distributions and Poisson distributions. For this joint distribution, the deviance is $Dev = -2\{\log P[\boldsymbol{Y}, \underset{\sim}{y}, \boldsymbol{N} | \underset{\sim}{n}, \Theta] - \log P[\boldsymbol{Y}, \underset{\sim}{y}, \boldsymbol{N} | \underset{\sim}{n}, \hat{\Theta}]\}$. That is,

$$Dev = D_0 + Dev(p_1, p_2) + Dev(\alpha_2) + \sum_{j=1}^{k} D_j, \tag{24}$$

where

$$
\begin{aligned}
D_0 &= 2\{\chi_0 - y_0 - y_0 \log\{\frac{\chi_0}{y_0}\}\}, \\
Dev(p_1,p_2) &= 2\sum_{j=1}^{k}\{n_{1j}\log\{\frac{n_{1j}}{n_j p_1}\} + n_{2j}\log\{\frac{n_{2j}}{n_j p_2}\} + n_{0j}\log\{\frac{n_{0j}}{n_j(1-p_1-p_2)}\}\}, \\
Dev(\alpha_2) &= 2\sum_{j=1}^{k}\{n_{0j}^{(I)}\log\{\frac{n_{0j}^{(I)}}{n_{0j}(1-\alpha_2)}\} + n_{0j}^{(J)}\log\{\frac{n_{0j}^{(J)}}{n_{0j}\alpha_2}\}\}, \\
D_j &= 2\sum_{i=1}^{2}\{n_{ij}Q_i(j) - y_{ij} - y_{ij}\log\{\frac{n_{ij}Q_i(j)}{y_{ij}}\}\} \\
&+ 2\{n_{0j}^{(I)}Q_0^{(I)}(j) - y_{0j}^{(I)} - y_{0j}^{(I)}\log\{\frac{n_{0j}^{(I)}Q_0^{(I)}(j)}{y_{0j}^{(I)}}\}\} \\
&+ 2\{n_{0j}^{(J)}Q_0^{(J)}(j) - y_{0j}^{(J)} - y_{0j}^{(J)}\log\{\frac{n_{0j}^{(J)}Q_0^{(J)}(j)}{y_{0j}^{(J)}}\}\}.
\end{aligned}
$$

The Unknown Parameters and Fitting of the Model by Cancer Incidence Data

In the above model, the unknown parameters are $\{p_1, p_2, \alpha_1, \alpha_2, \beta_0^{(I)}(t), \beta_i^{(I)}(t), b_i^{(I)}(t), d_i^{(I)}(t), i = 1, 2, \beta_0^{(J)}(t), \beta_1^{(J)}(t), b_1^{(J)}(t), d_1^{(J)}(t)\}$. Since the mutation rates are very small, we assume $\beta_i^{(I)}(t) = \beta_i^{(I)}$ for $i = 0, 1, 2$ and $\beta_i^{(J)}(t) = \beta_i^{(I)}$ for $i = 0, 1$. The proliferation rates of I_l cells for $l = 1, 2$ and J_1 cells are expected to be small [3]. It is also reasonable to assume $b_1^{(I)}(t) = b_1^{(I)}$, $d_1^{(I)}(t) = d_1^{(I)}$, $b_2^{(I)}(t) = b_2^{(I)}$, $d_2^{(I)}(t) = d_2^{(I)}$, $b_1^{(J)}(t) = b_1^{(J)}$ and $d_1^{(J)}(t) = d_1^{(J)}$ so that $\gamma_1^{(I)}(t) = b_1^{(I)} - d_1^{(I)} = \gamma_1^{(I)}$, $\gamma_2^{(I)}(t) = b_2^{(I)} - d_2^{(I)} = \gamma_2^{(I)}$ and $\gamma_1^{(J)}(t) = b_1^{(J)} - d_1^{(J)} = \gamma_1^{(J)}$. (see [2], [37]).

To fit the SEER Wilms' tumor data, we let one time unit (i.e. $\Delta t = 1$) correspond to three months after birth and take $t_0 = 1$. Because the last stage cells (i.e. I_3 cells, J_2 cells) grow so fast, during a three months period one may practically assume $P_T(s,t) \sim 1$ if $t - s \geq 1$. Using this discrete approximation, as shown in the Appendix A, we obtain:

$$
\begin{aligned}
E[I_2(t;2)] &\approx E[I_2(t_0;2)](1+\gamma_2^{(I)})^{t-t_0}, i=1,2,\\
E[I_2(t;1)] &\approx E[I_2(t_0;1)](1+\gamma_2^{(I)})^{(t-t_0)} + E[I_1(t_0;1)]\beta_1^{(I)}\sum_{u=1}^{2} A_{12}(u)(1+\gamma_u^{(I)})^{t-t_0},\\
E[I_2(t;0)] &\approx E[N(t_0)]\beta_0^{(I)}\beta_1^{(I)}\sum_{u=1}^{2} A_{12}(u)\frac{1}{\gamma_u^{(I)}}[(1+\gamma_u^{(I)})^{t-t_0}-1],\\
E[J_1(t)] &\approx E[N(t_0)]\beta_0^{(J)}\frac{1}{\gamma_1^{(J)}}[(1+\gamma_1^{(J)})^{t-t_0}-1].
\end{aligned}
$$

Comparing these expected numbers with those given by equations (14)-(17) respectively, observe that if one replaces $e^{\gamma_1^{(I)}(t-t_0)}$ by $(1+\gamma_1^{(I)})^{(t-t_0)} = e^{(t-t_0)\log\{1+\gamma_1^{(I)}\}}$ $\approx e^{(t-t_0)\gamma_1^{(I)}}$, $e^{\gamma_2^{(I)}(t-t_0)}$ by $(1+\gamma_2^{(I)})^{(t-t_0)} = e^{(t-t_0)\log\{1+\gamma_2^{(I)}\}} \approx e^{(t-t_0)\gamma_2^{(I)}}$ and $e^{\gamma_1^{(J)}(t-t_0)}$ by $(1+\gamma_1^{(J)})^{(t-t_0)} = e^{(t-t_0)\log\{1+\gamma_1^{(j)}\}} \approx e^{(t-t_0)\gamma_1^{(J)}}$, then the above approximations are equal to those by equations (14)-(17) respectively. Using the above results, as shown in the Appendix A, we obtain, to order of $o(\beta_2^{(I)})$ and $o(\beta_1^{(J)})$:

$$
\begin{aligned}
Q_0^{(I)}(j) &= \{e^{-\lambda_3\phi_{02}(t_{j-1})} - e^{-\lambda_3\phi_{02}(t_j)}\} + o(\beta_2^{(I)}),\\
Q_0^{(J)}(j) &= \{e^{-\lambda_4\phi_{01}(t_{j-1})} - e^{-\lambda_4\phi_{01}(t_j)}\} + o(\beta_1^{(J)}),\\
Q_1(j) &= \{e^{-\theta\phi_{22}(t_{j-1})-\lambda_2\phi_{12}(t_{j-1})} - e^{-\theta\phi_{22}(t_j)-\lambda_2\phi_{12}(t_j)}\} + o(\beta_2^{(I)}),\\
Q_2(j) &= (1-\alpha_1)\{e^{-\lambda_1\phi_{22}(t_{j-1})} - e^{-\lambda_1\phi_{22}(t_j)}\} + o(\beta_2^{(I)}),
\end{aligned}
$$

where

$$
\begin{aligned}
\phi_{22}(t) &= \{(1+\gamma_2^{(I)})^{(t-t_0)}-1\},\ \text{if } \gamma_2^{(I)} \neq 0;\\
\phi_{12}(t) &= \{\prod_{i=1}^{2}\gamma_i^{(I)}\}\sum_{u=1}^{2}A_{12}(u)\frac{1}{\gamma_u^{(I)}}\{(1+\gamma_u^{(I)})^{(t-t_0)}-1\},\\
&\quad \text{if } \gamma_2^{(I)} \neq \gamma_1^{(I)} \neq 0;\\
\phi_{02}(t) &= \{\prod_{i=1}^{2}\gamma_i^{(I)}\}\sum_{u=1}^{2}A_{12}(u)\frac{1}{{\gamma_u^{(I)}}^2}\{(1+\gamma_u^{(I)})^{(t-t_0)}-1-\gamma_u^{(I)}(t-t_0)\},\\
&\quad \text{if } \gamma_2^{(I)} \neq \gamma_1^{(I)} \neq 0;\\
\phi_{01}(t) &= \{(1+\gamma_1^{(J)})^{(t-t_0)}-1-\gamma_1^{(J)}(t-t_0)\},\ \text{if } \gamma_1^{(J)} \neq 0.
\end{aligned}
$$

From the above analysis, it follows that to order of $o(\beta_2^{(I)})$, $o(\beta_1^{(J)})$, $\{Q_1(j), Q_2(j), Q_0^{(I)}(j), Q_0^{(J)}(j)\}$ depend on the parameters only through the parametric functions $\{\lambda_i, i=1,2,3,4,\ \theta, \gamma_1^{(I)}, \gamma_2^{(I)}, \gamma_1^{(J)}\}$. If $E[N(t_0)]$, $E[I_2(t_0;0)]$, $E[I_2(t_0;1)]$, $E[I_2(t_0;2)]$ and $E[J_1(t_0)]$ are unknown, it is not possible to estimate the mutation rates $\{\beta_i^{(I)}, i=0,1,2,\ \beta_i^{(J)}, i=0,1\}$ but only the functions $\{\lambda_i, i=1,2,3,4,\ \theta\}$ of these parameters. Similarly, one can not estimate $\{b_i^{(I)}, d_i^{(I)}, i=1,2,\ b_1^{(J)}, d_1^{(J)}\}$ but only the proliferation rates $\{\gamma_i^{(I)}=b_i^{(I)}-d_i^{(I)}, i=1,2,\ \gamma_1^{(J)}=b_1^{(J)}-d_i^{(J)}\}$. Thus, the estimable parameters are $\Theta=\{p_i, \alpha_i, \gamma_i^{(I)}, i=1,2,\ \theta, \gamma_1^{(J)}, \lambda_j, j=1,2,3,4\}$. Notice that there are 12 unknown estimable parameters. We will refer the model as Model-F.

Single Pathway Model with Hereditary Cancer Cases

For comparison purposes, we will also consider fitting the SEER data of Wilms' tumor by a single pathway 3-stage model with hereditary cancer cases. The model is referred to as Model-S. For people who are normal at the embryo stage, Wilms' tumor is derived by one single pathway: $N\ (I_0) \to I_1 \to I_2 \to I_3 \to Tumor$. For people who have genotype I_1 at the embryo stage, the tumor is developed by: $I_1 \to I_2 \to I_3 \to Tumor$. For people who have genotype I_2 at the embryo stage, the tumor is derived by: $I_2 \to I_3 \to Tumor$.

This model has 9 unknown parameters, which are

$\Theta = \{p_i, \gamma_i^{(I)}, i = 1, 2,\ \alpha_1, \theta, \lambda_j, j = 1, 2, 3\}$.

E. The Generalized Bayesian Method and the Gibbs Sampling Procedure

To fit the models to cancer incidence data and to validate the models, one would need to estimate the unknown parameters and to predict the state variables. We propose a generalized Bayesian inference procedure to achieve these purposes.

The generalized Bayesian inference is based on the posterior distribution $P\{\Theta|\boldsymbol{N}, \boldsymbol{Y}, \underset{\sim}{y}, \underset{\sim}{n}\}$ of Θ given $\{\boldsymbol{N}, \boldsymbol{Y}, \underset{\sim}{y}, \underset{\sim}{n}\}$. This posterior distribution is derived by combining the prior distribution $P\{\Theta\}$ of Θ with the joint probability distribution $P\{\boldsymbol{N}, \boldsymbol{Y}, \underset{\sim}{y}|\underset{\sim}{n}, \Theta\}$ given by equation (23). It follows that this inference procedure would combine information from three sources: (1) Previous information and experiences about the parameters in terms of the prior distribution $P\{\Theta\}$ of the parameters. (2) Biological information of inherited cancer cases via genetic segregation of cancer genes in the population ($P[\boldsymbol{N}|\underset{\sim}{n}, p_i, i = 1, 2]$). (3) Information from the expanded data $\boldsymbol{Y}$ and the observed data $\underset{\sim}{y}$ via the statistical model from the system ($P[\boldsymbol{Y}, \underset{\sim}{y}|\boldsymbol{N}, \Theta]$). Because of additional information from the genetic segregation of the cancer genes, this inference procedure provides an efficient procedure to extract information of effects of genotypes of individuals at the embryo stage.

The Prior Distribution of the Parameters

For the prior distributions of Θ, because biological information have suggested some lower bounds and upper bounds for the mutation rates and for the proliferation rates, we assume

$$P(\Theta) \propto c\ (c > 0)$$

where c is a positive constant if these parameters satisfy some biologically specified constraints; and equal to zero for otherwise. These biological constraints are:

1) $0 < p_1 < 10^{-2}$, $0 < p_2 < 10^{-3}$, $0 < \alpha_i < 1$ $(i = 1, 2)$, $-0.01 < \gamma_i^{(I)} < 1$ $(i = 0, 1)$, $-0.01 < \gamma_1^{(J)} < 1$, $10^{-8} < \beta_i^{(I)} < 10^{-3}$ $(i = 0, 1, 2)$, $10^{-8} < \beta_k^{(J)} < 10^{-3}$ $(k = 0, 1)$ and $10^3 < N(t_0) < 10^8$.

2) For the λ_j $(j = 1, 2, 3, 4)$ and θ, we let $0 < \lambda_1 = \frac{1}{\gamma_2^{(I)}} E[I_2(t_0; 2)]\beta_2^{(I)} < 10^4$, $0 < \lambda_2 = \frac{1}{\gamma_1^{(I)}\gamma_2^{(I)}} E[I_1(t_0; 1)]\beta_1^{(I)}\beta_2^{(I)} < 10^4$, $0 < \lambda_3 = \frac{1}{\gamma_1^{(I)}\gamma_2^{(I)}} N(t_0)\beta_0^{(I)}\beta_1^{(I)}\beta_2^{(I)} < 10^{-1}$, $0 < \lambda_4 = \frac{1}{\gamma_1^{(J)^2}} N(t_0)\beta_0^{(J)}\beta_1^{(J)} < 10^3$ and $0 < \theta = \frac{1}{\gamma_2^{(I)}} E[I_2(t_0; 1)]\beta_2^{(I)} < 10^{-1}$.

We will refer the above prior as a partially informative prior which may be considered as an extension of the traditional non-informative prior given in [38].

The Posterior Distribution of the Parameters Given $\{\boldsymbol{Y}, \boldsymbol{N}, \underset{\sim}{y}, \underset{\sim}{n}\}$

Denote by $\Theta = \{p_i, \alpha_i, \gamma_i^{(I)}, i = 1, 2, \lambda_j, j = 1, 2, 3, 4, \gamma_1^{(J)}, \theta\}$. From the posterior distribution $P\{\Theta|\boldsymbol{N}, \boldsymbol{Y}, \underset{\sim}{y}, \underset{\sim}{n}\}$, we obtain:

$$
\begin{aligned}
P\{\Theta|\boldsymbol{N}, \boldsymbol{Y}, \underset{\sim}{y}, \underset{\sim}{n}\} \;\propto\; & (\alpha_1)^{y_0} e^{-n_0\alpha_1 p_2} p_1^{\sum_{j=1}^k n_{1j}} p_2^{y_0+\sum_{j=1}^k n_{2j}} (1 - p_1 - p_2)^{\sum_{j=1}^k n_{0j}} \\
& \times \; \alpha_2^{\sum_{j=1}^k n_{0j}^{(J)}} (1 - \alpha_2)^{\sum_{j=1}^k n_{0j}^{(I)}} \\
& \times \; \prod_{j=1}^{k} \{e^{-n_{0j}^{(I)} Q_0^{(I)}(j)} [n_{0j}^{(I)} Q_0^{(I)}(j)]^{y_{0j}^{(I)}} e^{-n_{0j}^{(J)} Q_0^{(J)}(j)} [n_{0j}^{(J)} Q_0^{(J)}(j)]^{y_{0j}^{(J)}} \\
& \times \; \prod_{i=1}^{2} e^{-n_{ij} Q_i(j)} [n_{ij} Q_i(j)]^{y_{ij}} \}, \Theta \in \Omega,
\end{aligned}
$$

where Ω is the parameter space of Θ provided by the biological constraints in the previous subsection.

For computational convenience, we notice that the log of $P\{\Theta|\boldsymbol{N}, \boldsymbol{Y}, \underset{\sim}{y}, \underset{\sim}{n}\}$ is proportional to the negative of the deviance given by equation (24).

The Multi-level Gibbs Sampling Procedure For Estimating Parameters

Given the above posterior probability distributions, we will use the following multi-level Gibbs sampling procedure to derive estimates of the parameters. We notice that numerically, the Gibbs sampling procedure given below is equivalent to the EM-algorithm

from the sampling theory viewpoint with Steps (1)-(2) as the E-Step and with Step (3) as the M-Step respectively [39]. These multi-level Gibbs sampling procedures are given by:

1) Generating $\boldsymbol{N}$ given $(\boldsymbol{Y}, \underset{\sim}{y}, \underset{\sim}{n}, \Theta)$:

Given Θ and given $\underset{\sim}{n}$, use the multinomial distribution of $\{n_{1j}, n_{2j}\}$ given n_j and the binomial distribution of $n_{0j}^{(I)}$ given n_{0j} to generate a large sample of $\boldsymbol{N}$. Then, by combining this sample with $P\{\boldsymbol{Y}, \underset{\sim}{y} | \boldsymbol{N}, \underset{\sim}{n}, \Theta\}$ in equation (18) to select $\boldsymbol{N}$ through the weighted bootstrap method due to Smith and Gelfant [40]. This selected $\boldsymbol{N}$ is then a sample from $P\{\boldsymbol{N}|\boldsymbol{Y}, \underset{\sim}{y}, \underset{\sim}{n}, \Theta\}$ even though the latter is unknown. (For proof, see [31], Chapter 3.) Call the generated sample $\hat{\boldsymbol{N}}$.

2) Generating $\boldsymbol{Y}$ given $(\boldsymbol{N}, \underset{\sim}{y}, \underset{\sim}{n}, \Theta)$:

Given $\{\underset{\sim}{y}, \underset{\sim}{n}, \Theta\}$ and given $\boldsymbol{N} = \hat{\boldsymbol{N}}$ generated from Step (1), generate $\boldsymbol{Y}$ from the multinomial distribution and the binomial distribution given by equation (21) and (22). Call the generated sample $\hat{\boldsymbol{Y}}$.

3) Estimation of Θ given $\{\boldsymbol{N}, \boldsymbol{Y}, \underset{\sim}{y}, \underset{\sim}{n}\}$:

Given $\{\underset{\sim}{y}, \underset{\sim}{n}\}$ and given $(\boldsymbol{N}, \boldsymbol{Y}) = (\hat{\boldsymbol{N}}, \hat{\boldsymbol{Y}})$ from Step (1) and Step (2), derive the posterior mode of Θ by maximizing the conditional posterior distribution $P\{\Theta|\hat{\boldsymbol{N}}, \hat{\boldsymbol{Y}}, \underset{\sim}{y}, \underset{\sim}{n}\}$. Under the partially informative prior, this is equivalent to maximize the negative of the deviance given by equation (24) under the constraints given in this section. Denote this generated mode by $\hat{\Theta}$.

4) Recycling Step:

With $\{(\boldsymbol{N}, \boldsymbol{Y}, \Theta) = (\hat{\boldsymbol{N}}, \hat{\boldsymbol{Y}}, \hat{\Theta})\}$ given above, go back to Step (1) and continue until convergence.

The proof of convergence of the above steps can be derived by using procedure given in [31], Chapter 3. At convergence, the $\hat{\Theta}$ are the generated values from the posterior distribution of Θ given $\{\underset{\sim}{y}, \underset{\sim}{n}\}$ independently of $(\boldsymbol{N}, \boldsymbol{Y})$ (for proof, see [31], Chapter 3). Repeat the above procedures one then generates a random sample of Θ from the posterior distribution of Θ given $\{\underset{\sim}{y}, \underset{\sim}{n}\}$; then one uses the sample mean as the estimates of Θ and

use the sample variances and covariances as estimates of the variances and covariances of these estimates.

F. Application to Fit the SEER Data

In this section, we will apply the above model to the Wilms' tumor incidence data from NCI/NIH's SEER program over the years 1973 - 2007. Given in Table 1 are the numbers of people at risk and the Wilms' tumor observed cases in the age groups together with the predicted cases from the models. These data give cancer incidence at birth and incidence for 84 age groups ($k = 84$) with each group spanning over a 1 year period.

To fit the data, we let one time unit be three months after birth and let $t_0 = 1$. To compare different models, we fit the following two mixture models: (1) The mixture model with a 3-stage model for hereditary cancer cases and some non-hereditary cancer cases and a 2-stage model for other non-hereditary cancer cases (Model-F). (2) The single pathway 3-stage mixture model (Model-S). We fit these models to the SEER data given in Table 1.

Given in Table 2 are the natural log of the likelihood functions, the AIC (Akaike Information Criterion) and the BIC (Bayesian Information Criterion) for these models. Table 3 shows estimates of parameters in Model-F. Figure 3 gives the plots of predicted cancer cases from the Model-F and Model-S. For comparison purposes, in Table 1, we also provide numbers of predicted cancer cases from Model-F and Model-S together with the observed cancer cases over time from SEER. From these results, we have made the following observations:

1) As shown by results in Table 1 and Figure 3, it appeared that Model-F fitted the SEER data better than Model-S. The AIC (Akaike Information Criteria) and BIC (Bayesian Information Criteria) values of Model-F are given by (AIC=352.332, BIC=381.644) which are smaller than those of Model-S respectively. These results suggest that Wilms' tumor may best be described by a model with the first pathway given

as 3-stage model for both hereditary and non-hereditary cancer cases and the second pathway given as a 2-stage model for non-hereditary cancer cancers.

2) From Table 3, the estimates of p_1 and p_2 from the SEER data are of orders 10^{-3} and 10^{-4} respectively. This indicates that in the US population, the frequency of stage limiting cancer genes for Wilms' tumor is approximately around 4×10^{-3}.

3) Table 3 shows that the estimate of α_1 is 0.1459. This indicates that about 15% individuals with genotype I_2 would develop cancer at birth. The estimate of α_2 is of order of 10^{-5}. This indicates that the proportion of normal people in the population at risk of Wilms' tumor by 2-stage pathway is very small.

4) From Table 3, estimate of proliferation rate of I_1 is order of 10^{-7}, quite small, indicating that the phenotype I_1 is almost identical to N/I_0; further confirming staging-limiting genes are basically tumor suppressor genes. Estimate of proliferation rate of I_2 is order of 10^{-5}, which is 10^2 times greater than those of I_1. Estimate of proliferation rate of J_1 is order of 10^{-3}, which is 10^4 times greater than those of I_1.

5) Results in Table 3 show that the estimates $\{\hat{\lambda}_j, j = 1, 2, 3, 4,\ \hat{\theta}\}$ of $\{\lambda_j, \theta\}$ are of order $\{10^2, 10^3, 10^{-2}, 10^2, 10^{-2}\}$ respectively. Because $\{\lambda_1 = \frac{1}{\gamma_2^{(I)}} E[I_2(t_0; 2)]\beta_2^{(I)}$, $\lambda_2 = \frac{1}{\gamma_1^{(I)}\gamma_2^{(I)}} E[I_1\ (t_0; 1)]\beta_1^{(I)}\beta_2^{(I)}, \lambda_3 = \frac{1}{\gamma_1^{(I)}\gamma_2^{(I)}} N(t_0)\beta_0^{(I)}\beta_1^{(I)}\beta_2^{(I)}, \lambda_4 = \frac{1}{\gamma_1^{(J)^2}} N(t_0)\beta_0^{(J)}\beta_1^{(J)}$, $\theta = \frac{1}{\gamma_2^{(I)}} E[I_2(t_0; 1)]\beta_2^{(I)}\}$, assuming some values of $\{E[N(t_0)], E[I_1(t_0; 1)], E[I_2(t_0; 2)]\}$ from some biological observations, one can have some rough ideas about the magnitude of $\beta_j^{(I)}$ $(j = 0, 1, 2)$ $(\beta_k^{(J)}$ $(k = 0, 1))$. If we follow [41] to assume $E[N(t_0)] = E[I_1(t_0; 1)] = E[I_2(t_0; 2)] \sim 10^8$, then $\{\beta_j^{(I)}, j = 0, 1, 2,\ \beta_k^{(J)}, k = 0, 1\} \approx 10^{-7} \sim 10^{-5}$.

Table 1: Wilms' Tumor Incidence Data from SEER (Overall Population)

Age Groups	Number of People at Risk	Observed Incidence	Model-F Predicated	Model-S Predicated
0	12495777	212	209	183
1	12221582	239	233	254
2	12120990	221	229	189
3	12112995	217	210	142
4	12146174	167	170	107
5	12161336	137	123	81
6	12111854	73	81	60
7	12160452	61	54	46
8	11942586	38	36	34
9	12381299	22	27	26
10	12512703	20	21	20
11	12410338	11	16	15
12	12449244	8	13	12
13	12527781	7	10	9
14	12602883	7	8	7
15	12719598	7	6	6
16	12766107	6	5	4
17	12831400	6	4	4
18	12382047	2	3	3
19	12581638	0	3	2
20	12636509	3	3	2
21	12682601	0	2	2

Continued on next page

Table 1 – continued from previous page

Age Groups	Number of People at Risk	Observed Incidence	Model-F Predicated	Model-S Predicated
22	12840510	6	2	2
23	13075528	4	2	2
24	13358635	6	2	2
25	13473849	4	2	2
26	13426340	4	2	1
27	13525264	2	2	1
28	13149674	3	2	1
29	13812811	0	2	1
30	13886874	0	2	2
31	13488332	1	2	1
32	13460286	3	2	2
33	13256067	3	2	2
34	13428827	1	2	2
35	13220037	1	2	2
36	12870265	2	2	2
37	12689592	0	2	2
38	12157014	0	2	2
39	12494081	1	2	2
40	12272125	3	2	2
41	11826573	2	2	2
42	11663153	3	2	2
43	11407082	3	2	2
44	11296848	2	2	2

Continued on next page

Table 1 – continued from previous page

Age Groups	Number of People at Risk	Observed Incidence	Model-F Predicated	Model-S Predicated
45	11016369	2	2	2
46	10651593	0	2	2
47	10475708	0	2	2
48	9994684	0	2	1
49	10138908	1	2	2
50	9836359	1	2	2
51	9475641	1	2	1
52	9250985	3	2	1
53	9027382	0	2	1
54	8883737	2	2	1
55	8547883	0	2	1
56	8279648	1	2	1
57	8062368	2	2	1
58	7654610	0	2	1
59	7563706	2	2	1
60	7232719	2	2	1
61	6927332	0	2	1
62	6708273	2	2	1
63	6543931	0	2	1
64	6404652	1	2	1
65	6168486	2	2	1
66	5913479	0	2	1
67	5746766	0	2	1

Continued on next page

Table 1 – continued from previous page

Age Groups	Number of People at Risk	Observed Incidence	Model-F Predicated	Model-S Predicated
68	5480517	2	2	1
69	5363912	0	2	1
70	5110728	0	1	1
71	4925076	1	1	1
72	4696825	1	1	1
73	4512136	1	1	1
74	4345300	1	1	1
75	4148801	2	1	1
76	3900900	1	1	1
77	3681587	1	1	1
78	3481918	1	1	1
79	3243631	1	1	1
80	2961234	0	1	1
81	2724984	0	1	1
82	2495219	0	1	1
83	2271595	0	1	1
84	2041351	0	1	1

Table 2: The Log-Likelihood, AIC and BIC of the Fitted Models of Wilms' Tumor

Models	Log-Likelihood	AIC	BIC
Model-F	-164.166	352.332	381.644
Model-S	-211.840	441.680	463.664

Table 3: Estimates of Parameters for the Stochastic Model of Wilms' Tumor

Parameters	p_1	p_2	α_1	α_2
Estimates	3.999E-03	1.147E-04	1.459E-01	3.501E-05
St.D	1.057E-05	8.252E-07	8.051E-03	5.386E-07
95%CL-Lower	3.996E-03	1.145E-04	1.436E-01	3.485E-05
95%CL-Upper	4.002E-03	1.150E-04	1.482E-01	3.516E-05
Parameters	$\gamma_1^{(I)}$	$\gamma_2^{(I)}$	$\gamma_1^{(J)}$	λ_1
Estimates	5.134E-07	6.745E-05	5.471E-03	9.955E+02
St.D	8.636E-09	1.592E-06	2.627E-04	5.744E+01
95%CL-Lower	5.110E-07	6.700E-05	5.396E-03	9.792E+02
95%CL-Upper	5.159E-07	6.790E-05	5.545E-03	1.012E+03
Parameters	λ_2	λ_3	λ_4	θ
Estimates	1.757E+03	1.000E-02	3.796E+02	5.453E-02
St.D	1.465E+02	5.895E-04	1.296E+01	2.562E-03
95%CL-Lower	1.715E+03	9.835E-03	3.760E+02	5.380E-02
95%CL-Upper	1.799E+03	1.017E-02	3.833E+02	5.526E-02

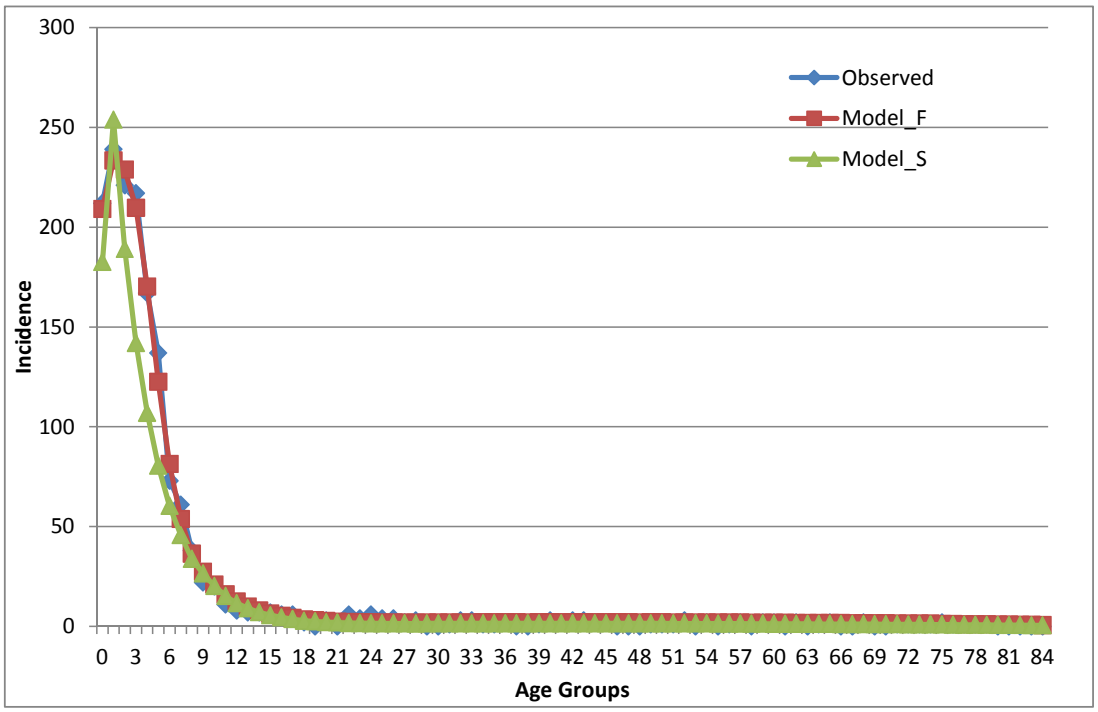

Fig. 3: Curve Fitting of Wilms' Tumor SEER Data by Proposed Models

G. Computation Details

The multi-level Gibbs sampling procedure for estimating unknown parameters is implemented in Fortran 90. The Fortran code is shown in the Appendix B. Figure 4 shows the program flow chart. The subroutine NGENERNOR01 and the subroutine NGENERNOR are used to generate $\boldsymbol{N}$ from multinomial distribution of $\{n_{1j}, n_{2j}\}$ given n_j and binomial distribution of $n_{0j}^{(I)}$ given n_{0j}. Since $\underset{\sim}{n}$ is very large and p_1 is very small, the normal approximation is applied. The subroutine PICK is applied to select the k-th $\boldsymbol{N}$ from a large sample of $\boldsymbol{N}$ through the Weighted Bootstrap Method. The selected $\boldsymbol{N}$ is a sample from $P\{\boldsymbol{N}|\boldsymbol{Y}, \underset{\sim}{y}, \underset{\sim}{n}, \Theta\}$. $\boldsymbol{Y}$ is generated by the subroutine YGENER from the multinomial distribution and the subroutine Y3GENER from the binomial distribution. The publicly available Genetic Algorithm PIKAIA is applied to derive the posterior mode of Θ by maximizing the conditional posterior distribution $P\{\Theta|\hat{\boldsymbol{N}}, \hat{\boldsymbol{Y}}, \underset{\sim}{y}, \underset{\sim}{n}\}$. The genetic algorithms are a class of search techniques inspired from the biological process of

evolution by means of natural selection. The basic principle is that those with the largest fitness will be selected as the generation progresses. Given the fitness, the genetic algorithm would choose the parameter values to maximize the fitness according to evolutionary principle as described above. The function Fit is called in PIKAIA as fitness function that is the negative of the deviance given in (2.26). The subroutine CalculateQ is used to calculate the probability of developing tumor during each age period in people through different pathways.

H. Discussion and Conclusion

Based on studies of molecular biology on human Wilms' tumor as discussed in Section B, in this chapter we have developed a multi-pathway stochastic model of carcinogenesis for human Wilms' tumor. To account for hereditary cancer cases and the development of non-hereditary cancers through two different pathways in stochastic model, we have also developed a generalized mixture model. In this mixture model, the mixing probability distributions are a multinomial distribution to explain genetic segregation of the stage-limiting tumor suppressor genes for Wilms' tumor and a binomial distribution is to account for the development of non-hereditary Wilms' tumor through two pathways. This mixture model allows us to estimate for the first time the frequency p_1 of the stage-limiting tumor suppressor genes for Wilms' tumor in the US population.

For using the proposed models to fit the cancer incidence data, in this Chapter we have developed a generalized Bayesian inference procedure to estimate the unknown parameters and to predict cancer cases. This inference procedure is advantageous over the classical sampling theory inference (i.e. maximum likelihood method) because the procedure combines information from three sources: (1) Previous information and experiences about the parameters in terms of the prior distribution $P\{\Theta\}$ of the parameters. (2) Biological information of hereditary cancer cases via the genetic segregation of stage-limiting tumor suppressor genes in the population. (3) Information

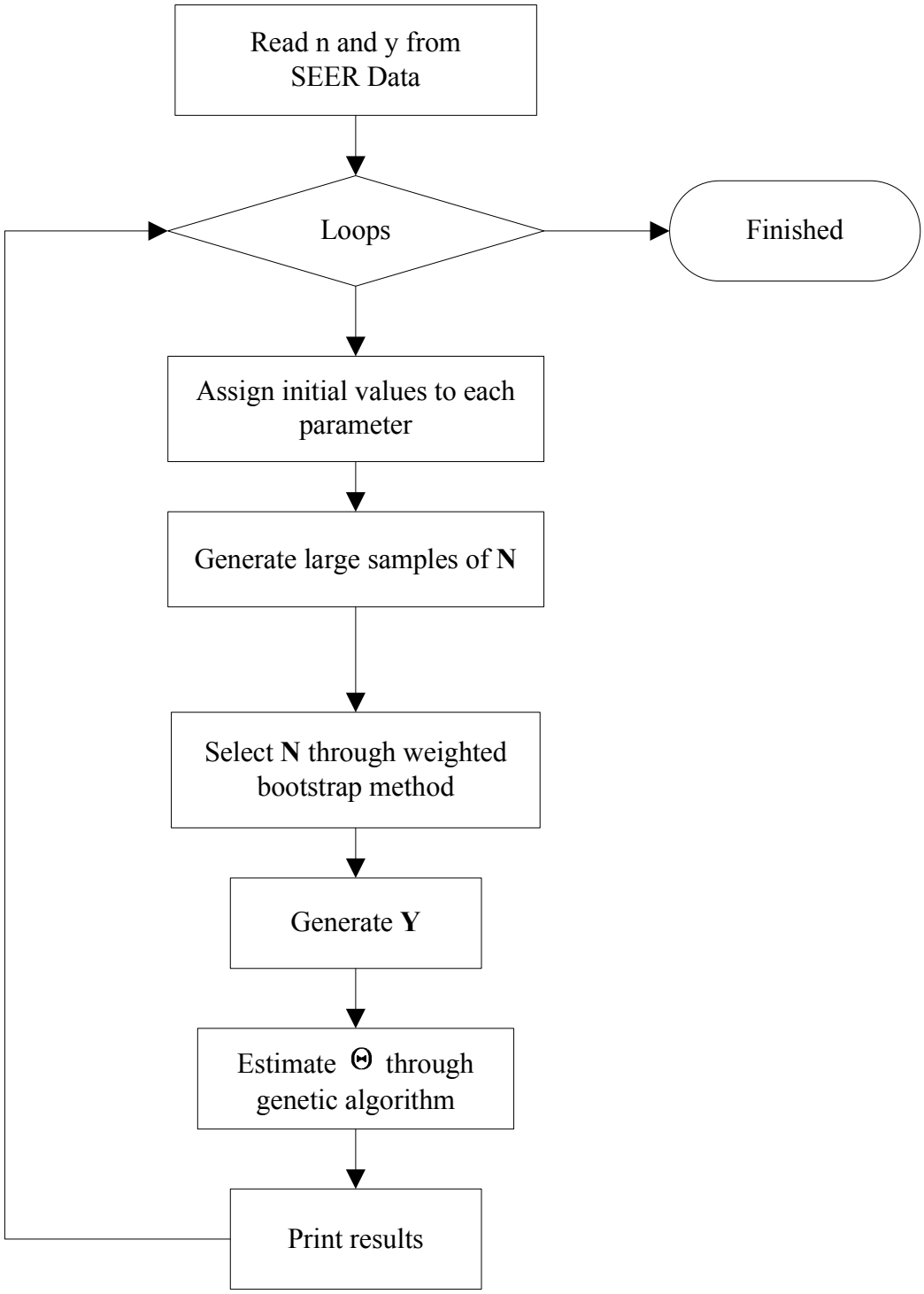

Fig. 4: Program Flow Chart

from the expanded data ($\boldsymbol{Y}$) and the observed data ($\underset{\sim}{y}$) via the statistical model from $P\{\boldsymbol{Y}, \underset{\sim}{y} | \boldsymbol{N}, \Theta\}$.

To illustrate the usefulness and applications of our models and methods, we have applied our models and methods to the SEER Wilms' tumor data of NCI/NIH. Our analysis clearly showed that the proposed multiple pathway model with hereditary cancer cases fitted the data better than the single pathway 3-stage model with hereditary cancer cases (see Table 2 and Figure 3).

Applying our models and methods to the SEER data of human Wilms' tumor, we have derived for the first time some useful information. Specifically, we mention: (1) For the first time, we have estimated the frequency of the stage-limiting tumor suppressor genes in the US population ($\hat{p_1} \sim 4.003 \times 10^{-3}$). (2) With the estimate of α_1 as $\hat{\alpha}_1 = 0.146$, the predicted number of Wilms' tumor at birth is $\hat{y}_0 = n_0 \hat{\alpha}_1 \hat{p}_2 = 209$ by multiple pathway model with inherited cancer component (Model-F). (The observed number of Wilms' tumor at birth is 212.)

Using models and methods of this paper, one can easily predict future cancer cases for human Wilms' tumor. Thus, by comparing results from different populations, our models and methods can be used to assess cancer prevention and control procedures. This will be our future research topics; we will not go any further here.

CHAPTER III

A NEW STOCHASTIC MODEL OF ADULT KIDNEY CANCER-RENAL CELL CARCINOMAS

A. Introduction

Renal cell carcinona (RCC), the most common form of kidney cancer in adults, originates in the lining of the proximal convoluted tubule, the very small tubes in the kidney that filter the blood and remove waste products. The incidence of renal cell carcinomas worldwide has been increasing at an annual rate of approximately 2%. It most commonly occurs in adults between the ages of 50 and 70 years. Based on histopathology, three common subtypes of carcinomas are distinguished: clear cell carcinomas (ccRCC, about 80% of RCC), papillary carcinomas (pRCC, about 15%) and chromophobe carcinomas(chRCC, about 5%) [16], [17]. The purpose of this chapter is to develop a multiple-pathway model for renal cell carcinomas basing on studies of genetics and biology.

In Section B, we present a brief summary of renal cell carcinomas biology. In Section C, we develop a biologically supported a multiple-pathway stochastic model of renal cell carcinomas. In Section D, we derive a statistical model for cancer incidence data of human renal cell carcinomas. This statistical model is basically a generalized mixture model of Poisson distributions with a bivariate multinomial distribution as the mixing probability distribution. By combining results from Section B-D, in Section E, we develop a generalized Bayesian inference procedure to estimate unknown parameters. To illustrate the applications of the model and methods, in Section F, we apply the models and methods to the SEER renal carcinomas incidence data. In Section G, we show the computation details of fitting the model of renal cell carcinoma. Finally in Section H, we discuss the usefulness of the model and methods and provide some conclusions.

B. A Brief Summary of Renal Cell Carcinoma Biology

Clear cell renal cell carcinoma (ccRCC) is the most common renal malignancy, accounting for about 80% of renal cell carcinomas. It is characterized by loss of genetic material on the short arm of chromosome 3. The von Hippel Lindau tumor suppressor gene (VHL) on chromosome 3p25-p26 is mutated or inactivated in about 50% of ccRCC cases [42]. It is an early event in carcinogenesis of this tumor. VHL gene plays a role in the regulation of cell-cycle control, intercellular signaling, extracellular fibronectin formation and angiogenesis. The VHL protein targets hypoxia inducible factors (HIF) for ubiquitin mediated degradation. When VHL gene is mutated or inactivated, HIF accumulates in the nucleus, resulting in increased transcription of downstream pathway genes that promote tumor angiogenesis, growth and survival, such as VEGF, GLUT1 and TGFB1 [43]. Although inactivation of VHL is necessary, VHL loss alone is insufficient of ccRCC tumorigenesis and additional genetic events are needed [44]. Mutations of PBRM1 on chromosome 3p21.1 and mutations of SETD2 on chromosome 3p21.31 have been reported recently, which are observed in approximately 40% and 3% of ccRCCs. Both genes are involved in chromatin regulation at the level of histone H3 modification and recognition [45]. Based on these combined molecular genetic results, we propose a 4-stage model for ccRCC development. The proposed model is represented schematically by Figure 5.

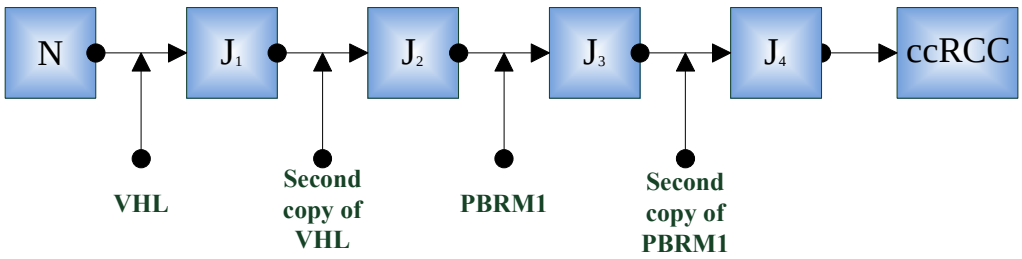

Fig. 5: Four-stage Model for ccRCC Development

Papillary renal cell carcinoma (pRCC) comprises approximately 15% of renal cell carcinomas. Mutations of MET oncogene on chromosome 7q31 are observed to play a role in about 13% of patients with pRCC. These mutations result in ligand-independent activation of intracytoplasmic tyrosine kinase domains, which activate the hepatocyte growth factor(HGF)/MET pathway [46], [16]. Gains of Chromosomal 7 (85%) and 17 (92%) and additional gains of chromosome 3q, 8q, 12q, 16q and 20q (24%-67%) are found in pRCC. It has been proposed that combined trisomy of chromosomes 7 and 17 induce renal papillary adenomas, trisomies at chromosome 3q, 8q, 12q, 16q and 20q mark papillary RCCs and gain of chromosome 1q and loss of chromosomes 6q, 9p and 14q relate to an aggressive clinical behavior and deadly progression [47]. The mutation of the oncogene MET in chromosome 7q would speed up these transitions by promoting the proliferation rates of the respective intermediate initiated cells. Based on these combined molecular genetic findings, we propose a 3-stage model for pRCC development. The proposed model is represented schematically by Figure 6.

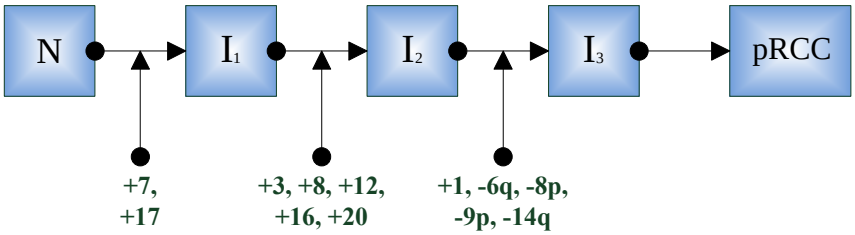

Fig. 6: Three-stage Model for pRCC Development

Chromophobe renal cell carcinoma (chRCC) accounts for approximately 5% of renal cell carcinomas. Recent studies have revealed that losses of multiple chromosomes 1, 2, 6, 10, 13, 17 and 21 occur in most of chRCC (up to 95%) [48], [49]. These studies suggest that at least 5-7 chromosomes should be lost before a clinically recognizable chRCC develops. Thus, we may assume a 5-stage model for chRCC development.

C. *A Stochastic Multi-Stage Model of Renal Cell Carcinomas Involving Multiple Pathways*

Based on results of molecular biology in the previous section, in this chapter we will propose a three-pathway stochastic model of carcinogenesis for RCC with I-pathway as a 3-stage multi-stage model ($N \to I_1 \to I_2 \to I_3 \to tumor$), with J-pathway as a 4-stage multi-stage model ($N \to J_1 \to J_2 \to J_3 \to J_4 \to tumor$) and with K-pathway as a 5-stage multi-stage model ($N \to K_1 \to K_2 \to K_3 \to K_4 \to K_5 \to tumor$). The proposed model can be represented schematically by Figure 7. We let p_i $(i = 1, 2, 3)$ be the proportion of people in the population at risk of developing RCC through I-pathway, J-pathway and K-pathway respectively $(p_1 + p_2 + p_3 = 1)$. Let n_j denote the number of people at risk of cancer during the $j-$th age period $[t_{j-1}, t_j)$. Among the n_j people at risk of cancers in the $j-$th age period, let n_{ij} $(i = 1, 2, 3)$ be the number of people at risk of developing RCC through I-pathway, J-pathway and K-pathway respectively $(n_{1j} + n_{2j} + n_{3j} = n_j)$. The conditional probability distribution of $\{n_{1j}, n_{2j}\}$ given n_j is multinomial with parameters $\{n_j; p_1, p_2\}$; that is, $(n_{1j}, n_{2j}) \sim$ Multinomial$\{n_j; p_1, p_2\}$.

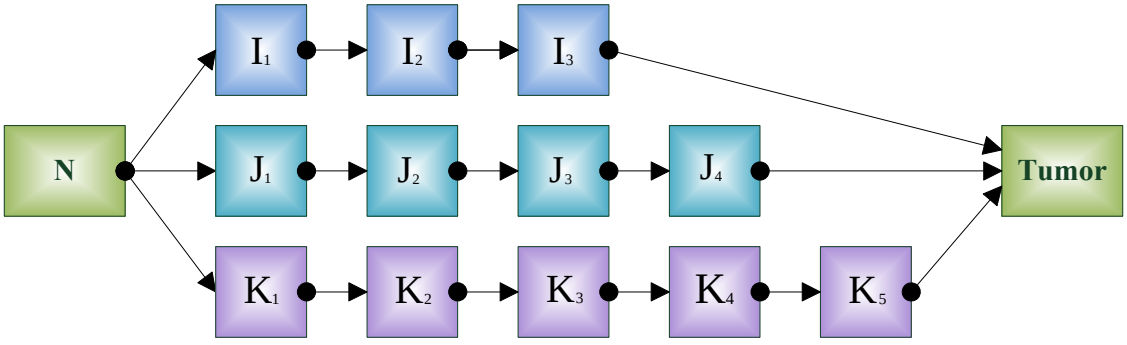

Fig. 7: Three Pathways for RCC Development

Let Y_j denote the random variable for the observed cancer cases y_i in the $j-$th time period $[t_{j-1}, t_j)$ and $f_i(Y_j)$ $(i = 1, 2, 3)$ the probability density of Y_j given that the individual is an person who is at risk to develop RCC through I-pathway, J-pathway and

K-pathway respectively. Then for any person taken randomly from the population , the probability density of Y_j is $f(Y_j) = \sum_{i=1}^{3} p_i f_i(Y_j)$.

The Stochastic Model and Mathematical Analysis

The stochastic model proposed above postulates that the I_l $(l = 1, 2, 3)$, J_u $(u = 1, 2, 3, 4)$ and K_v $(v = 1, 2, 3, 4, 5)$ cells undergo stochastic proliferation (birth) and differentiation (death); cancer tumors arise from primary I_3 (J_4, K_5) cells through clonal expansion by following stochastic birth-death process [32], where primary I_3 (J_4, K_5) cells are I_3 (J_4, K_5) cells generated directly by I_2 (J_3, K_4) cells by genetic alterations; all cells proceed forward independently of other cells. The last stage in each pathway (i.e. I_3 stage, J_4 stage, K_5 stage) is a transient stage to cancer tumors. Therefore, the state variables for this model are $\underset{\sim}{X}(t) = \{I_l(t), l = 1, 2,\ J_u(t), u = 1, 2, 3,\ K_v(t), v = 1, 2, 3, 4\}$ and $T(t)$, where $I_l(t)$ ($J_u(t)$, $K_v(t)$) denotes the number of the I_l (J_u, K_v) cells for $\{l = 1, 2,\ u = 1, 2, 3,\ v = 1, 2, 3, 4\}$ respectively at time t and $T(t)$ the number of cancer tumors at time t. Notice that $\{\underset{\sim}{X}(t), t \geq t_0\}$ can be assumed as Markov, where t_0 represents time at birth. However, $T(t)$ is in general not Markov [50]. To develop stochastic models of carcinogenesis, it is conveniently assumed that the last stage cells (i.e. I_3 cells, J_4 cells, K_5 cells) grow instantaneously into cancer tumors as soon as they are generated as shown in Tan [1], Little [8] and Zheng [33]. In this case, one may assume $T(t)$ as Markov.

Let $Q_i(j)$ $(i = 1, 2, 3)$ denote the probability of developing tumor during the $j-$th age period $[t_{j-1}, t_j)$ $(t_j > t_0)$ by the I-pathway, J-pathway and K-pathway respectively. Let $\beta_l^{(I)}(t)$ denote the transition rate from $I_l \to I_{l+1}$ $(l = 0, 1, 2)$ at time t, $\beta_u^{(J)}(t)$ the transition rate from $J_u \to J_{u+1}$ $(u = 0, 1, 2, 3)$ at time t and $\beta_v^{(K)}(t)$ the transition rate from $K_v \to K_{v+1}$ $(v = 0, 1, 2, 3, 4)$ at time t. Then by using methods in Tan [34], Tan el al. [35], [2] and Tan and Yan [10], it can be shown that $Q_i(j),\ (i = 1, 2, 3)$ are given

respectively by:

$$
\begin{aligned}
Q_1(j) &= \{e^{-\int_{t_0}^{t_{j-1}} \beta_2^{(I)} E[I_2(x)] P_T^{(I)}(x,t_{j-1})dx} - e^{-\int_{t_0}^{t_j} \beta_2^{(I)}(x) E[I_2(x)] P_T^{(I)}(x,t_{j-1})dx}\} + o(\beta_2^{(I)}), \\
Q_2(j) &= \{e^{-\int_{t_0}^{t_{j-1}} \beta_3^{(J)}(x) E[J_3(x)] P_T^{(J)}(x,t_{j-1})dx} - e^{-\int_{t_0}^{t_j} \beta_3^{(J)}(x) E[J_3(x)] P_T^{(J)}(x,t_{j-1})dx}\} + o(\beta_3^{(J)}), \\
Q_3(j) &= \{e^{-\int_{t_0}^{t_{j-1}} \beta_4^{(K)}(x) E[K_4(x)] P_T^{(K)}(x,t_{j-1})dx} - e^{-\int_{t_0}^{t_j} \beta_4^{(K)}(x) E[K_4(x)] P_T^{(K)}(x,t_{j-1})dx}\} \\
&+ o(\beta_4^{(K)}).
\end{aligned}
$$

Where $E[I_2(x)]$ ($E[J_3(x)]$, $E[K_4(x)]$) is the expected number of $I_2(t)$ ($J_3(t)$, $K_4(t)$) and where $P_T^{(I)}(s,t)$ ($P_T^{(J)}(s,t)$, $P_T^{(K)}(s,t)$) is the probability that a primary I_3 (J_4, K_5) cell generated from an I_2 (J_3, K_4) cell at time s develops into a detectable tumor by time t.

For derive mathematical analysis for the above model, let $b_l^{(I)}(t)$ ($b_u^{(J)}(t), b_v^{(K)}(t)$) and $d_l^{(I)}(t)$ ($d_u^{(J)}(t), d_v^{(K)}(t)$) denote the birth rate and the death rate at time t of the $I_l, l = 1, 2$ ($J_u, u = 1, 2, 3,\ K_v, v = 1, 2, 3, 4$) cells respectively. Let $\{B_l^{(I)}(t), D_l^{(I)}(t),\ M_l^{(I)}(t)\}$ ($\{B_u^{(J)}(t), D_u^{(J)}(t), M_u^{(J)}(t)\}$, $\{B_v^{(K)}(t), D_v^{(K)}(t), M_v^{(K)}(t)\}$) be the number of birth and the number of death of I_l (J_u, K_v) cells and the number of transition from $I_l \to I_{l+1}$ ($J_u \to J_{u+1}, K_v \to K_{v+1}$) cells during $[t, t + \Delta t)$ respectively. Also let $M_0^{(I)}$ ($M_0^{(J)}, M_0^{(K)}$) be the number of mutation of $N \to I_1$ ($N \to J_1$, $N \to K_1$) during $[t, t + \Delta t)$. Then the conditional probability distributions of random transition variables given the state variables are, to order of $o(\Delta t)$,

$$
\begin{aligned}
M_0^{(I)}(t)|N(t) &\sim \text{Poisson}\{N(t)\beta_0^{(I)}(t)\Delta t\} && (25) \\
\{B_l^{(I)}(t), D_l^{(I)}(t)\}|I_l(t) &\sim \text{Multinomial}\{I_l(t); b_l^{(I)}(t)\Delta t, d_l^{(I)}(t)\Delta t\}, && (26) \\
&\quad \text{independently of } M_0^{(I)}(t),\ l = 1, 2 \\
M_l^{(I)}(t)|I_l(t) &\sim \text{Poisson}\{I_l(t)\beta_l^{(I)}(t)\Delta t\}, && (27) \\
&\quad \text{independently of } \{M_0^{(I)}(t), B_l^{(I)}(t), D_l^{(I)}(t)\}, l = 1, 2 \\
M_0^{(J)}(t)|N(t) &\sim \text{Poisson}\{N(t)\beta_0^{(J)}(t)\Delta t\} && (28)
\end{aligned}
$$

$$\{B_l^{(J)}(t), D_u^{(J)}(t)\}|J_u(t) \sim \text{Multinomial}\{J_u(t); b_u^{(J)}(t)\Delta t, d_u^{(J)}(t)\Delta t\}, \quad (29)$$
$$\text{independently of } M_0^{(J)}(t),\ u = 1, 2, 3$$
$$M_u^{(J)}(t)|J_u(t) \sim \text{Poisson}\{J_u(t)\beta_u^{(J)}(t)\Delta t\}, \quad (30)$$
$$\text{independently of } \{M_0^{(J)}(t), B_u^{(J)}(t), D_u^{(J)}(t)\}, u = 1, 2, 3$$
$$M_0^{(K)}(t)|N(t) \sim \text{Poisson}\{N(t)\beta_0^{(K)}(t)\Delta t\} \quad (31)$$
$$\{B_v^{(K)}(t), D_v^{(K)}(t)\}|K_v(t) \sim \text{Multinomial}\{K_v(t); b_v^{(K)}(t)\Delta t, d_v^{(K)}(t)\Delta t\}, \quad (32)$$
$$\text{independently of } M_0^{(K)}(t),\ v = 1, 2, 3, 4$$
$$M_v^{(K)}(t)|K_v(t) \sim \text{Poisson}\{K_v(t)\beta_v^{(K)}(t)\Delta t\}, \quad (33)$$
$$\text{independently of } \{M_0^{(K)}(t), B_v^{(K)}(t), D_v^{(K)}(t)\},$$
$$k = 1, 2, 3, 4$$

We have the following stochastic equations of the state variables $\{I_l(t), l = 1, 2,$ $J_u(t), u = 1, 2, 3,\ K_v(t), v = 1, 2, 3, 4\}$:

$$I_l(t + \Delta t) = I_l(t) + B_l^{(I)}(t) - D_l^{(I)}(t) + M_{l-1}^{(I)}(t),\ l = 1, 2, \quad (34)$$
$$J_u(t + \Delta t) = J_u(t) + B_u^{(J)}(t) - D_u^{(J)}(t) + M_{u-1}^{(J)}(t),\ u = 1, 2, 3, \quad (35)$$
$$K_v(t + \Delta t) = K_v(t) + B_v^{(K)}(t) - D_v^{(K)}(t) + M_{v-1}^{(K)}(t),\ v = 1, 2, 3, 4. \quad (36)$$

Given the probability distributions of the random transition variables in equations (25)-(33) and the stochastic equations in equations (34)-(36), we derive the following stochastic differential equations for the state variables $\{I_l(t), l = 1, 2,\ J_u(t),$ $u = 1, 2, 3,\ K_v(t), v = 1, 2, 3, 4\}$:

$$dI_l(t) = I_l(t)\gamma_l^{(I)}(t)\Delta t + I_{l-1}(t)\beta_{l-1}^{(I)}(t)\Delta t + e_l^{(I)}(t)\Delta t,\ l = 1, 2, \quad (37)$$
$$dJ_u(t) = J_u(t)\gamma_u^{(J)}(t)\Delta t + J_{u-1}(t)\beta_{u-1}^{(J)}(t)\Delta t + e_u^{(J)}(t)\Delta t,\ u = 1, 2, 3, \quad (38)$$
$$dK_v(t) = K_v(t)\gamma_v^{(K)}(t)\Delta t + K_{v-1}(t)\beta_{v-1}^{(K)}(t)\Delta t + e_v^{(K)}(t)\Delta t,\ v = 1, 2, 3, 4. \quad (39)$$

Where $\gamma_l^{(I)}(t) = b_l^{(I)}(t) - d_l^{(I)}(t)$ for $l = 1, 2$, $\gamma_u^{(J)}(t) = b_u^{(J)}(t) - d_u^{(J)}(t)$ for $u = 1, 2, 3$ and $\gamma_v^{(K)}(t) = b_v^{(K)}(t) - d_v^{(K)}(t)$ for $v = 1, 2, 3, 4$ and the random noises $\{e_l^{(I)}(t), e_u^{(J)}(t), e_v^{(K)}(t)\}$ are:

$$\begin{aligned}
e_l^{(I)}(t)\Delta t &= [B_l^{(I)}(t) - I_l(t)b_l^{(I)}(t)\Delta t] - [D_l^{(I)}(t) - I_l(t)d_l^{(I)}(t)\Delta t] \\
&+ [M_{l-1}^{(I)}(t) - I_{l-1}(t)\beta_{l-1}^{(I)}(t)\Delta t],\ l = 1, 2, \\
e_u^{(J)}(t)\Delta t &= [B_u^{(J)}(t) - J_u(t)b_u^{(J)}(t)\Delta t] - [D_u^{(J)}(t) - J_u(t)d_u^{(J)}(t)\Delta t] \\
&+ [M_{u-1}^{(J)}(t) - J_{u-1}(t)\beta_{u-1}^{(J)}(t)\Delta t],\ u = 1, 2, 3, \\
e_v^{(K)}(t)\Delta t &= [B_v^{(K)}(t) - K_v(t)b_v^{(K)}(t)\Delta t] - [D_v^{(K)}(t) - K_v(t)d_v^{(K)}(t)\Delta t] \\
&+ [M_{v-1}^{(K)}(t) - K_{v-1}(t)\beta_{v-1}^{(K)}(t)\Delta t],\ v = 1, 2, 3, 4.
\end{aligned}$$

From the above equations, the random noises have expectation zero and are un-correlated with the state variables $\underset{\sim}{X}(t)$ and $T(t)$. The covariances between these random noises are $o(\Delta t)$ and the variances of the random noises are given by:

$$\begin{aligned}
\text{var}[e_l^{(I)}(t)\Delta t] &= E[I_{l-1}(t)]\beta_{l-1}^{(I)}\Delta t + E[I_l(t)][b_l^{(I)}(t) + d_l^{(I)}(t)]\Delta t + o(\Delta t), \\
\text{var}[e_u^{(J)}(t)\Delta t] &= E[J_{u-1}(t)]\beta_{u-1}^{(J)}\Delta t + E[J_u(t)][b_u^{(J)}(t) + d_u^{(J)}(t)]\Delta t + o(\Delta t), \\
\text{var}[e_v^{(K)}(t)\Delta t] &= E[K_{v-1}(t)]\beta_{v-1}^{(K)}\Delta t + E[K_v(t)][b_v^{(K)}(t) + d_v^{(K)}(t)]\Delta t + o(\Delta t).
\end{aligned}$$

The solution of the above equation (37)-(39) are given respectively by:

$$I_l(t) = \int_{t_0}^{t} I_{l-1}(x)\beta_{l-1}^{(I)}(x)e^{\int_x^t \gamma_l^{(I)}(y)dy}dx + \eta_l^{(I)}(t),\ l = 1, 2, \tag{40}$$

$$J_u(t) = \int_{t_0}^{t} J_{u-1}(x)\beta_{u-1}^{(J)}(x)e^{\int_x^t \gamma_u^{(J)}(y)dy}dx + \eta_u^{(J)}(t),\ u = 1, 2, 3, \tag{41}$$

$$K_v(t) = \int_{t_0}^{t} K_{v-1}(x)\beta_{v-1}^{(K)}(x)e^{\int_x^t \gamma_v^{(K)}(y)dy}dx + \eta_v^{(K)}(t),\ v = 1, 2, 3, 4, \tag{42}$$

where $\eta_l^{(I)}(t) = \int_{t_0}^{t} e^{\int_x^t \gamma_l^{(I)}(y)dy} e_l^{(I)}(x)dx,\ \eta_u^{(J)}(t) = \int_{t_0}^{t} e^{\int_x^t \gamma_u^{(J)}(y)dy} e_u^{(J)}(x)dx,$
$\eta_v^{(K)}(t) = \int_{t_0}^{t} e^{\int_x^t \gamma_v^{(K)}(y)dy} e_v^{(K)}(x)dx$ and $I_0(t_0) = J_0(t_0) = K_0(t_0) = N(t_0)$.

Since $E[\eta_l^{(I)}(t)]$ $(l = 1, 2)$, $E[\eta_u^{(J)}(t)]$ $(u = 1, 2, 3)$ and $E[\eta_v^{(K)}(t)]$ $(v = 1, 2, 3, 4)$ are all zeros, the expected numbers $E[I_l(t)]$ of I_l $(l = 1, 2)$, $E[J_u(t)]$ of J_u $(l = 1, 2, 3)$ and $E[K_v(t)]$ of K_v $(v = 1, 2, 3, 4)$ are:

$$E[I_l(t)] = \int_{t_0}^{t} E[I_{l-1}(x)]\beta_{l-1}^{(I)}(x)e^{\int_x^t \gamma_l^{(I)}(y)dy}dx,\ l = 1, 2, \tag{43}$$

$$E[J_u(t)] = \int_{t_0}^{t} E[J_{u-1}(x)]\beta_{u-1}^{(J)}(x)e^{\int_x^t \gamma_u^{(J)}(y)dy}dx,\ u = 1, 2, 3, \tag{44}$$

$$E[K_v(t)] = \int_{t_0}^{t} E[K_{v-1}(x)]\beta_{v-1}^{(K)}(x)e^{\int_x^t \gamma_v^{(K)}(y)dy}dx,\ v = 1, 2, 3, 4. \tag{45}$$

If the model is time homogeneous so that $\{b_l^{(I)}(t) = b_l^{(I)}, d_l^{(I)}(t) = d_l^{(I)}, \gamma_l^{(I)}(t) = \gamma_l^{(I)}, \beta_l^{(I)}(t) = \beta_l^{(I)},\ l = 0, 1, 2,\ b_u^{(J)}(t) = b_u^{(J)}, d_u^{(J)}(t) = d_u^{(J)}, \gamma_u^{(J)}(t) = \gamma_u^{(J)}, \beta_u^{(J)}(t) = \beta_u^{(J)},\ u = 0, 1, 2, 3,\ b_v^{(K)}(t) = b_v^{(K)}, d_v^{(K)}(t) = d_v^{(K)}, \gamma_v^{(K)}(t) = \gamma_v^{(K)}, \beta_v^{(K)}(t) = \beta_v^{(K)},\ v = 0, 1, 2, 3, 4\}$ and if $\{\gamma_i^{(I)} \neq \gamma_j^{(I)}, \gamma_i^{(J)} \neq \gamma_j^{(J)}, \gamma_i^{(K)} \neq \gamma_j^{(K)}\}$ for $i \neq j$, then the above expected numbers reduce to:

$$E[I_2(t)] = E[N(t_0)]\prod_{j=0}^{1}\beta_j^{(I)}\sum_{u=1}^{2}A_{12}(u)\frac{1}{\gamma_u^{(I)}}[e^{\gamma_u^{(I)}(t-t_0)} - 1], \tag{46}$$

$$E[J_3(t)] = E[N(t_0)]\prod_{j=0}^{2}\beta_j^{(J)}\sum_{u=1}^{3}A_{13}(u)\frac{1}{\gamma_u^{(J)}}[e^{\gamma_u^{(J)}(t-t_0)} - 1], \tag{47}$$

$$E[K_4(t)] = E[N(t_0)]\prod_{j=0}^{3}\beta_j^{(K)}\sum_{u=1}^{4}A_{14}(u)\frac{1}{\gamma_u^{(K)}}[e^{\gamma_u^{(K)}(t-t_0)} - 1], \tag{48}$$

where $A_{ij}(u) = \prod_{v=i, v\neq u}^{j}(\gamma_u - \gamma_v)^{-1}$ for $i \leq u \leq j$.

According to the above results, we can derive $Q_i(j),\ i = 1, 2, 3$ for homogeneous models under the condition that $\{\gamma_i^{(I)} \neq \gamma_j^{(I)}, \gamma_i^{(J)} \neq \gamma_j^{(J)}, \gamma_i^{(K)} \neq \gamma_j^{(K)}\}$ for $i \neq j$:

$$Q_1(j) = \{e^{-\lambda_1\psi_{02}(t_{j-1})} - e^{-\lambda_1\psi_{02}(t_j)}\} + o(\beta_2^{(I)}), \tag{49}$$

$$Q_2(j) = \{e^{-\lambda_2\psi_{03}(t_{j-1})} - e^{-\lambda_2\psi_{03}(t_j)}\} + o(\beta_3^{(J)}), \quad (50)$$

$$Q_3(j) = \{e^{-\lambda_3\psi_{04}(t_{j-1})} - e^{-\lambda_3\psi_{04}(t_j)}\} + o(\beta_4^{(K)}), \quad (51)$$

where $\lambda_1 = \{\prod_{i=1}^{2}\gamma_i^{(I)}\}^{-1}E[N(t_0)]\prod_{j=0}^{2}\beta_j^{(I)}$, $\lambda_2 = \{\prod_{i=1}^{2}\gamma_i^{(J)}\}^{-2}E[N(t_0)]\prod_{j=0}^{3}\beta_j^{(J)}$, $\lambda_3 = \{\prod_{i=1}^{3}\gamma_i^{(K)}\}^{-2}E[N(t_0)]\prod_{j=0}^{4}\beta_j^{(K)}$, and

$$\begin{aligned}
\psi_{02}(t) &= \{\prod_{i=1}^{2}\gamma_i^{(I)}\}\sum_{u=1}^{2}A_{12}(u)\frac{1}{\gamma_u^{(I)}}\int_{t_0}^{t}\{e^{\gamma_u^{(I)}(x-t_0)}-1\}P_T^{(I)}(x,t)dx \\
&= \{\prod_{i=1}^{2}\gamma_i^{(I)}\}\sum_{u=1}^{2}A_{12}(u)\frac{1}{{\gamma_u^{(I)}}^2}\{e^{\gamma_u^{(I)}(t-t_0)}-1-\gamma_u^{(I)}(t-t_0)\}, \\
&\quad \text{if } P_T^{(I)}(x,t)=1 \text{ for } t>x; \\
\psi_{03}(t) &= \{\prod_{i=1}^{2}\gamma_i^{(J)}\}^2\sum_{u=1}^{3}A_{13}(u)\frac{1}{\gamma_u^{(J)}}\int_{t_0}^{t}\{e^{\gamma_u^{(J)}(x-t_0)}-1\}P_T^{(J)}(x,t)dx \\
&= \{\prod_{i=1}^{2}\gamma_i^{(J)}\}^2\sum_{u=1}^{3}A_{13}(u)\frac{1}{{\gamma_u^{(J)}}^2}\{e^{\gamma_u^{(J)}(t-t_0)}-1-\gamma_u^{(J)}(t-t_0)\}, \\
&\quad \text{if } P_T^{(J)}(x,t)=1 \text{ for } t>x; \\
\psi_{04}(t) &= \{\prod_{i=1}^{3}\gamma_i^{(K)}\}^2\sum_{u=1}^{4}A_{14}(u)\frac{1}{\gamma_u^{(K)}}\int_{t_0}^{t}\{e^{\gamma_u^{(K)}(x-t_0)}-1\}P_T^{(K)}(x,t)dx \\
&= \{\prod_{i=1}^{3}\gamma_i^{(K)}\}^2\sum_{u=1}^{4}A_{14}(u)\frac{1}{{\gamma_u^{(K)}}^2}\{e^{\gamma_u^{(K)}(t-t_0)}-1-\gamma_u^{(K)}(t-t_0)\}, \\
&\quad \text{if } P_T^{(K)}(x,t)=1 \text{ for } t>x.
\end{aligned}$$

The Transition Probability of State Variables

Let $g(x,y;N,p_1,p_2)$ denote the density at (x,y) of a multinomial distribution with parameters (N,p_1,p_2) and $h(x;\lambda)$ the density at x of Poisson distribution with mean λ. From equation (25)-(33), we obtain the transition probability of the Markov process of state variables as, to order of $o(\Delta t)$ and for $t > t_0$:

$$\begin{aligned}
& P\{I_j(t+\Delta t)=v_j,\ j=1,2|I_j(t)=u_j,\ j=1,2\} \\
= \ & P\{I_1(t+\Delta t)=v_1|I_1(t)=u_1\}P\{I_2(t+\Delta t)=v_2|I_1(t)=u_1,I_2(t)=u_2\},
\end{aligned}$$

$$
\begin{aligned}
& P\{I_1(t+\Delta t) = v_1 | I_1(t) = u_1\} \\
= & \sum_{r_1=0}^{u_1} \sum_{m_1=0}^{u_1 - r_1} g(r_1, m_1; u_1, b_1^{(I)}(t)\Delta t, d_1^{(I)}(t)\Delta t) h(v_1 - u_1 - r_1 + m_1; N(t)\beta_0^{(I)}(t)\Delta t), \\
& P\{I_2(t+\Delta t) = v_2 | I_1(t) = u_1, I_2(t) = u_2\} \\
= & \sum_{r_2=0}^{u_2} \sum_{m_2=0}^{u_2 - r} g(r_2, m_2; u_2, b_2^{(I)}(t)\Delta t, d_2^{(I)}(t)\Delta t) h(v_2 - u_2 - r_2 + m_2; u_1\beta_1^{(I)}(t)\Delta t).
\end{aligned}
$$

$$
\begin{aligned}
& P\{J_j(t+\Delta t) = v_j,\ j = 1,2,3 | J_j(t) = u_j,\ j = 1,2,3\} \\
= & P\{J_1(t+\Delta t) = v_1 | J_1(t) = u_1\} \\
\times & \prod_{j=2}^{3} P\{J_j(t+\Delta t) = v_j | J_{j-1}(t) = u_{j-1}, J_j(t) = u_j\}, \\
& P\{J_1(t+\Delta t) = v_1 | J_1(t) = u_1\} \\
= & \sum_{r_1=0}^{u_1} \sum_{m_1=0}^{u_1 - r_1} g(r_1, m_1; u_1, b_1^{(J)}(t)\Delta t, d_1^{(J)}(t)\Delta t) h(v_1 - u_1 - r_1 + m_1; N(t)\beta_0^{(J)}(t)\Delta t), \\
& P\{J_j(t+\Delta t) = v_j | J_{j-1}(t) = u_{j-1}, J_j(t) = u_j\} \\
= & \sum_{r_j=0}^{u_j} \sum_{m_j=0}^{u_j - r} g(r_j, m_j; u_j, b_j^{(J)}(t)\Delta t, d_j^{(J)}(t)\Delta t) h(v_j - u_j - r_j + m_j; u_{j-1}\beta_{j-1}^{(J)}(t)\Delta t), \\
& j = 2,3.
\end{aligned}
$$

$$
\begin{aligned}
& P\{K_j(t+\Delta t) = v_j,\ j = 1,2,3 | K_j(t) = u_j,\ j = 1,2,3\} \\
= & P\{K_1(t+\Delta t) = v_1 | K_1(t) = u_1\} \\
\times & \prod_{j=2}^{4} P\{K_j(t+\Delta t) = v_j | K_{j-1}(t) = u_{j-1}, K_j(t) = u_j\}, \\
& P\{K_1(t+\Delta t) = v_1 | K_1(t) = u_1\} \\
= & \sum_{r_1=0}^{u_1} \sum_{m_1=0}^{u_1 - r_1} g(r_1, m_1; u_1, b_1^{(K)}(t)\Delta t, d_1^{(K)}(t)\Delta t) h(v_1 - u_1 - r_1 + m_1; N(t)\beta_0^{(K)}(t)\Delta t), \\
& P\{K_j(t+\Delta t) = v_j | K_{j-1}(t) = u_{j-1}, K_j(t) = u_j\} \\
= & \sum_{r_j=0}^{u_j} \sum_{m_j=0}^{u_j - r} g(r_j, m_j; u_j, b_j^{(K)}(t)\Delta t, d_j^{(J)}(t)\Delta t) h(v_j - u_j - r_j + m_j; u_{j-1}, \beta_{j-1}^{(K)}(t)\Delta t), \\
& j = 2,3,4.
\end{aligned}
$$

D. A Statistical Model and The Probability Distribution of the Number of Detectable Tumors

The data available for modeling carcinogenesis are usually cancer incidence over different time periods. For example, the SEER data of NCI/NIH for human renal carcinomas are given by $\{(y_j, n_j), j = 1, \ldots, k\}$, where y_j is the number of cancer cases during the $j-$th age group of 1 year period and n_j is the number of people who are at risk of renal carcinomas and from whom y_j of them have developed cancer during the $j-$th age group. Given in Table 4 are the SEER data for human renal carcinomas cases. We let Y_j denote the random variable for the observed cancer cases y_j in the $j-$th age group. In this section, we will develop a statistical model for this dataset.

The Probability Distribution of Y_j

To derive the probability distribution of Y_j in the $j-$th age group, let Y_{ij} $(i = 1, 2, 3)$ be the number of cancer cases generated by I-pathway, J-pathway and K-pathway respectively. The conditional distribution of $Y_{ij}|n_{ij} \sim$ Poisson$\{n_{ij}Q_i(j)\}$. Then The conditional probability distribution of $\{Y_{1j}, Y_{2j}, Y_j\}$ given $\{n_{1j}, n_{2j}, n_j\}$ is

$$P\{y_{1j}, y_{2j}, y_j|n_{1j}, n_{2j}, n_j\} = P\{y_{1j}, y_{2j}, y_{3j}|n_{1j}, n_{2j}, n_{3j}\} = \prod_{i=1}^{3} h\{y_{ij}; n_{ij}Q_i(j)\}. \quad (52)$$

Let $Q_T(j) = \sum_{i=1}^{3} n_{ij}Q_i(j)$. The conditional distribution of $Y_j|(n_{ij}, i = 1, 2, 3) \sim$ Poisson$\{Q_T(j)\}$. It follows that the probability distribution of Y_j given n_j is

$$P(y_j|n_j) = \sum_{n_{1j}=0}^{n_j} \sum_{n_{2j}=0}^{n_j - n_{1j}} g(n_{1j}, n_{2j}; n_j, p_1, p_2) h\{y_j; Q_T(j)\}, \quad (53)$$

where $g(n_{1j}, n_{2j}; n_j, p_1, p_2)$ is the probability density of $(n_{1j}, n_{2j})|n_j \sim$ Multinomial$(n_j; p_1, p_2)$ and $h\{y_j; Q_T(j)\}$ is the Poisson density of $Y_j|(n_{ij}, i = 1, 2, 3) \sim$ Poisson$\{Q_T(j)\}$.

The probability distribution $P(y_j|n_j)$ given by equation (53) is a mixture of Poisson distributions with a mixing probability distribution given by the multinomial distribution of $\{n_{1j}, n_{2j}\}$ given n_j. This mixing probability distribution represents individuals at risk of developing RCC through different pathway in the population. Let Θ be the set of all unknown parameters (i.e. the parameters (p_1, p_2) and the birth rates, the death rates and the mutation rates of I_l cells, J_u cells and K_v cells). Based on data $(y_j, j = 1, \ldots, k)$, the likelihood function of Θ is

$$L\{\Theta|y_j, j = 1, \ldots, k\} = \prod_{j=1}^{k} P(y_j|n_j). \tag{54}$$

The Probability Distribution of the Expanded Model

For applying the mixture distribution given by equation (54) to make inference about the unknown parameters, we expand the model to include the un-observable variables $\{n_{1j}, n_{2j}, y_{1j}, y_{2j}\}$. To derive the joint probability distribution of these variables, observe that the conditional probability distribution of $\{y_{1j}, y_{2j}\}$ given $\{n_{ij}, i = 1, 2, n_j, y_j\}$ is multinomial with parameters $\{y_j; \frac{n_{1j}Q_1(j)}{Q_T(j)}, \frac{n_{2j}Q_2(j)}{Q_T(j)}\}$. That is,

$$P\{y_{1j}, y_{2j}|n_{ij}, i = 1, 2, n_j, y_j\} \sim \text{Multinomial}\{y_j; \frac{n_{1j}Q_1(j)}{Q_T(j)}, \frac{n_{2j}Q_2(j)}{Q_T(j)}\}. \tag{55}$$

Hence the joint density of $\{n_{ij}, y_{ij}, i = 1, 2, y_j\}$ given n_j is:

$$\begin{aligned} & P\{n_{ij}, y_{ij}, i = 1, 2, y_j, j = 1, \ldots, k|n_j, \Theta\} \\ = \;& g(n_{1j}, n_{2j}; n_j, p_1, p_2) \prod_{i=1}^{3} h\{y_{ij}; n_{ij}Q_i(j)\}. \end{aligned} \tag{56}$$

Put $\boldsymbol{Y} = (y_{ij}, i = 1, 2, j = 1, \ldots, k)$, $\boldsymbol{N} = (n_{ij}, i = 1, 2, j = 1, \ldots, k)$, $\underset{\sim}{y} = (y_j, j = 1, \ldots, k)$ and $\underset{\sim}{n} = (n_j, j = 1, \ldots k\}$. For the SEER data, the joint density

$P\{\boldsymbol{Y}, \underset{\sim}{y}, \boldsymbol{N}|\underset{\sim}{n}, \Theta\}$ of $\{\boldsymbol{Y}, \underset{\sim}{y}, \boldsymbol{N}\}$ given $\{\underset{\sim}{n}, \Theta\}$ is:

$$P\{\boldsymbol{Y}, \underset{\sim}{y}, \boldsymbol{N}|\underset{\sim}{n}, \Theta\} = \prod_{j=1}^{k}\{g(n_{1j}, n_{2j}; n_j, p_1, p_2)\prod_{i=1}^{3} h[y_{ij}; n_{ij}Q_i(j)]\}. \tag{57}$$

Notice that the above distribution is a product of multinomial distributions and Poisson distributions. For this joint distribution, the deviance is:

$$\begin{aligned} Dev &= -2\{\log P[\boldsymbol{Y}, \underset{\sim}{y}, \boldsymbol{N}|\underset{\sim}{n}, \Theta] - \log P[\boldsymbol{Y}, \underset{\sim}{y}, \boldsymbol{N}|\underset{\sim}{n}, \hat{\Theta}]\} \\ &= 2\sum_{j=1}^{k}\{n_{1j}\log[\frac{n_{1j}}{n_j p_1}] + n_{2j}\log[\frac{n_{2j}}{n_j p_2}] + n_{3j}\log[\frac{n_{3j}}{n_j(1-p_1-p_2)}]\} \\ &+ 2\sum_{j=1}^{k}\sum_{i=1}^{3}\{n_{ij}Q_i(j) - y_{ij} - y_{ij}\log[\frac{n_{ij}Q_i(j)}{y_{ij}}]\}, \end{aligned} \tag{58}$$

The joint density $P\{\boldsymbol{Y}, \underset{\sim}{y}, \boldsymbol{N}|\underset{\sim}{n}, \Theta\}$ of $(\boldsymbol{Y}, \underset{\sim}{y}, \boldsymbol{N})$ given by equation (57) will be used as the kernel for the Bayesian method to estimate the unknown parameters and to predict the state variables.

The Unknown Parameters and Fitting of the Model by Cancer Incidence Data

In the above model, the unknown parameters are $\{p_1, p_2, \beta_0^{(I)}(t), \beta_l^{(I)}(t), b_l^{(I)}(t), d_l^{(I)}(t), l = 1, 2,\ \beta_0^{(J)}(t), \beta_u^{(J)}(t), b_u^{(J)}(t), d_u^{(J)}(t), u = 1, 2, 3,\ \beta_0^{(K)}(t), \beta_v^{(K)}(t), b_v^{(K)}(t), d_v^{(K)}(t), v = 1, 2, 3, 4\}$. Since the mutation rates are very small, it is reasonable to assume $\beta_l^{(I)}(t) = \beta_l^{(I)}$ for $l = 0, 1, 2$, $\beta_u^{(J)}(t) = \beta_u^{(J)}$ for $u = 0, 1, 2, 3$ and $\beta_v^{(K)}(t) = \beta_v^{(K)}$ for $v = 0, 1, 2, 3, 4$. The proliferation rates of I_l cells for $l = 1, 2$, J_u cells for $u = 1, 2$ and K_v cells for $v = 1, 2, 3, 4$ are expected to be small [3]. It is also reasonable to assume $b_l^{(I)}(t) = b_l^{(I)}$, $d_l^{(I)}(t) = d_l^{(I)}$ for $l = 1, 2$; $b_u^{(J)}(t) = b_u^{(J)}$, $d_u^{(J)}(t) = d_u^{(J)}$ for $u = 1, 2$; $b_v^{(K)}(t) = b_v^{(K)}$, $d_v^{(K)}(t) = d_v^{(K)}$ for $v = 1, 2, 3, 4$, hence $\gamma_l^{(I)}(t) = b_l^{(I)} - d_l^{(I)} = \gamma_l^{(I)}$, $\gamma_u^{(J)}(t) = b_u^{(J)} - d_u^{(J)} = \gamma_u^{(J)}$ and $\gamma_v^{(K)}(t) = b_v^{(K)} - d_v^{(K)} = \gamma_v^{(K)}$ (see Tan et al. [2], [36]). Because the protection devises such as the apoptosis and cell cycle inhibition are activated when the number of J_3 cells are very large, one would expect a Gompertz curve for the

growth of J_3 cells. Then we may assume $b_3^{(J)}(t) = b_3^{(J)} e^{-\delta(t-t_0)},\ d_3^{(J)}(t) = d_3^{(J)} e^{-\delta(t-t_0)}$, then $\gamma_3^{(J)}(t) = \gamma_3^{(J)} e^{-\delta(t-t_0)}$.

To fit the SEER data for renal cell carcinoma, we let one time unit (i.e. $\Delta t = 1$) correspond to three months after birth. Since the growth of last stage cells (i.e. I_3 cells, J_4 cells and K_5 cells) is very rapid, during a three months period one may practically assume $P_T^{(I)}(s,t) \sim 1$ ($P_T^{(J)}(s,t) \sim 1$, $P_T^{(K)}(s,t) \sim 1$) if $t - s \geq 1$. Using this discrete approximation, we obtain:

$$\begin{aligned}
E[I_2(t)] &\approx E[N(t_0)] \prod_{j=0}^{1} \beta_j^{(I)} \sum_{u=1}^{2} A_{12}(u) \frac{1}{\gamma_u^{(I)}} [(1+\gamma_u^{(I)})^{t-t_0} - 1] \\
E[J_3(t)] &\approx E[N(t_0)] \prod_{j=0}^{2} \beta_j^{(J)} \sum_{u=1}^{3} A_{13}(u) \frac{1}{\gamma_u^{(J)}} [(1+\gamma_u^{(J)})^{t-t_0} - 1] \\
E[K_4(t)] &\approx E[N(t_0)] \prod_{j=0}^{3} \beta_j^{(K)} \sum_{u=1}^{4} A_{14}(u) \frac{1}{\gamma_u^{(K)}} [(1+\gamma_u^{(K)})^{t-t_0} - 1]
\end{aligned}$$

If one replaces $e^{\gamma_l^{(I)}(t-t_0)}$ by $(1+\gamma_l^{(I)})^{(t-t_0)} = e^{(t-t_0)\log\{1+\gamma_l^{(I)}\}} \approx e^{(t-t_0)\gamma_l^{(I)}}$, $e^{\gamma_u^{(J)}(t-t_0)}$ by $(1+\gamma_u^{(J)})^{(t-t_0)} = e^{(t-t_0)\log\{1+\gamma_u^{(J)}\}} \approx e^{(t-t_0)\gamma_u^{(J)}}$ and $e^{\gamma_v^{(K)}(t-t_0)}$ by $(1+\gamma_v^{(K)})^{(t-t_0)} = e^{(t-t_0)\log\{1+\gamma_v^{(K)}\}} \approx e^{(t-t_0)\gamma_v^{(K)}}$, then the above approximations are equal to those by equations (46)-(48) respectively. For time homogeneous models, the $Q_i(j),\ i = 1,2,3$ under discrete approximation are:

$$\begin{aligned}
Q_1(j) &= \{e^{-\lambda_1 \phi_{02}(t_{j-1})} - e^{-\lambda_1 \phi_{02}(t_j)}\} + o(\beta_2^{(I)}) \\
Q_2(j) &= \{e^{-\lambda_2 \phi_{03}(t_{j-1})} - e^{-\lambda_2 \phi_{03}(t_j)}\} + o(\beta_3^{(J)}) \\
Q_3(j) &= \{e^{-\lambda_3 \phi_{04}(t_{j-1})} - e^{-\lambda_3 \phi_{04}(t_j)}\} + o(\beta_4^{(K)})
\end{aligned}$$

where $\lambda_1 = \{\prod_{i=1}^{2} \gamma_i^{(I)}\}^{-1} E[N(t_0)] \prod_{j=0}^{2} \beta_j^{(I)}$, $\lambda_2 = \{\prod_{i=1}^{2} \gamma_i^{(J)}\}^{-2} E[N(t_0)] \prod_{j=0}^{3} \beta_j^{(J)}$, $\lambda_3 = \{\prod_{i=1}^{3} \gamma_i^{(K)}\}^{-2} E[N(t_0)] \prod_{j=0}^{4} \beta_j^{(K)}$, and

$$\phi_{02}(t) = \{\prod_{i=1}^{2}\gamma_i^{(I)}\}\sum_{u=1}^{2}A_{12}(u)\frac{1}{{\gamma_u^{(I)}}^2}\{(1+\gamma_u^{(I)})^{(t-t_0)}-1-\gamma_u^{(I)}(t-t_0)\},$$
$$\text{if } \gamma_2^{(I)} \neq \gamma_1^{(I)} \neq 0$$
$$\phi_{03}(t) = \{\prod_{i=1}^{2}\gamma_i^{(J)}\}^2\sum_{u=1}^{3}A_{13}(u)\frac{1}{{\gamma_u^{(J)}}^2}\{(1+\gamma_u^{(J)})^{(t-t_0)}-1-\gamma_u^{(J)}(t-t_0)\},$$
$$\text{if } \gamma_i^{(J)} \neq \gamma_j^{(J)} \neq 0,\ i \neq j$$
$$\phi_{04}(t) = \{\prod_{i=1}^{3}\gamma_i^{(K)}\}^2\sum_{u=1}^{4}A_{14}(u)\frac{1}{{\gamma_u^{(K)}}^2}\{(1+\gamma_u^{(K)})^{(t-t_0)}-1-\gamma_u^{(K)}(t-t_0)\},$$
$$\text{if } \gamma_i^{(K)} \neq \gamma_j^{(K)} \neq 0,\ i \neq j.$$

From the above analysis, it follows that to order of $o(\beta_2^{(I)})$, $o(\beta_3^{(J)})$ and $o(\beta_4^{(K)})$, the $Q_i(j)$'s depend on the parameters only through the parametric functions $\{\lambda_i, i=1,2,3,\ \gamma_l^{(I)}, l=1,2,\ \gamma_u^{(J)}, u=1,2,3,\ \gamma_v^{(K)}, v=1,2,3,4,\ \delta\}$. Thus the estimable parameters are $\Theta = \{p_i, \gamma_i^{(I)}, i=1,2,\ \lambda_j, \gamma_j^{(J)}, j=1,2,3,\ \gamma_v^{(K)}, v=1,2,3,4,\ \delta\}$. There are 14 unknown estimable parameters in the model.

E. The Generalized Bayesian Method and the Gibbs Sampling Procedure

To fit the models to the data and to validate the models, one would need to estimate the unknown parameters and to predict the state variables. We propose a generalized Bayesian inference procedure to achieve these purposes.

The generalized Bayesian inference is based on the posterior distribution $P\{\Theta|\boldsymbol{N}, \boldsymbol{Y}, \underset{\sim}{y}, \underset{\sim}{n}\}$ of Θ given $\{\boldsymbol{N}, \boldsymbol{Y}, \underset{\sim}{y}, \underset{\sim}{n}\}$. This posterior distribution is derived by combining the prior distribution $P\{\Theta\}$ of Θ with the joint probability distribution $P\{\boldsymbol{N}, \boldsymbol{Y}, \underset{\sim}{y} | \underset{\sim}{n}, \Theta\}$ given $\{\underset{\sim}{n}, \Theta\}$ given by equation (57). It follows that this inference procedure would combine information from three sources: (1) Previous information and experiences about the parameters in terms of the prior distribution $P\{\Theta\}$ of the parameters. (2) Biological information of cancer cases via different pathways of RCC in the population

($P[\boldsymbol{N}|\underset{\sim}{n}, p_1, p_2]$). (3) Information from the expanded data ($\boldsymbol{Y}$) and the observed data ($\underset{\sim}{y}$) via the statistical model from the system ($P[\boldsymbol{Y}, \underset{\sim}{y}|\boldsymbol{N}, \underset{\sim}{n}, \Theta]$).

The Prior Distribution of the Parameters

For the prior distributions of Θ, because biological information have suggested some lower bounds and upper bounds for the mutation rates and for the proliferation rates, we assume

$$P(\Theta) \propto c \ (c > 0)$$

where c is a positive constant if these parameters satisfy some biologically specified constraints; and equal to zero for otherwise. These biological constraints are:

1) $0.05 < p_1 < 0.15$, $0.75 < p_2 < 0.85$, $-0.01 < \gamma_l^{(I)} < 1 \ \ (l = 1, 2)$, $-0.01 < \gamma_u^{(J)} < 1 \ (u = 1, 2, 3)$, $-0.01 < \gamma_v^{(K)} < 1 \ (v = 1, 2, 3, 4)$ and $0 < \delta < 10^{-2}$;

2) For the $\lambda_j \ (j = 1, 2, 3)$, we let $0 < \lambda_1 = \{\prod_{i=1}^{2} \gamma_i^{(I)}\}^{-1} E[N(t_0)] \prod_{j=0}^{2} \beta_j^{(I)} < 10$, $0 < \lambda_2 = \{\prod_{i=1}^{2} \gamma_i^{(J)}\}^{-2} E[N(t_0)] \prod_{j=0}^{3} \beta_j^{(J)} < 1, \ 0 < \lambda_3 = \{\prod_{i=1}^{3} \gamma_i^{(K)}\}^{-2} E[N(t_0)] \prod_{j=0}^{4} \beta_j^{(K)} < 10^2$, where $10^{-8} < \beta_i^{(I)} < 10^{-3} \ (i = 0, 1, 2)$, $10^{-8} < \beta_k^{(J)} < 10^{-3} \ (k = 0, 1)$, $N(t_0) \approx 10^8$.

We will refer the above prior as a partially informative prior which may be considered as an extension of the traditional non-informative prior given in Box and Tiao [38].

The Posterior Distribution of the Parameters Given $\{\boldsymbol{Y}, \boldsymbol{N}, \underset{\sim}{y}, \underset{\sim}{n}\}$

Denote by $\Theta = \{p_i, \gamma_i^{(I)}, i = 1, 2, \ \lambda_j, \gamma_j^{(J)}, j = 1, 2, 3, \ \gamma_v^{(K)}, v = 1, 2, 3, 4, \ \delta\}$. From the posterior distribution $P\{\Theta|\boldsymbol{N}, \boldsymbol{Y}, \underset{\sim}{y}, \underset{\sim}{n}\}$, we obtain:

$$P\{\Theta|\boldsymbol{N}, \boldsymbol{Y}, \underset{\sim}{y}, \underset{\sim}{n}\} \ \propto \ p_1^{\sum_{j=1}^{k} n_{1j}} p_2^{\sum_{j=1}^{k} n_{2j}} (1 - p_1 - p_2)^{\sum_{j=1}^{k} n_{oj}} \prod_{j=1}^{k} \prod_{i=1}^{3} e^{-n_{ij} Q_i(j)} \{n_{ij} Q_i(j)\}^{y_{ij}}, \ \Theta \in \Omega,$$

where Ω is the parameter space of Θ provided by the biological constraints in the previous subsection. We notice that the log of $P\{\Theta|\boldsymbol{N}, \boldsymbol{Y}, \underset{\sim}{y}, \underset{\sim}{n}\}$ is proportional to the negative of deviance given by equation (58).

The Multi-level Gibbs Sampling Procedure For Estimating Parameters

Given the above posterior probability distributions, we apply the following multi-level Gibbs sampling procedure to derive estimates of the parameters:

1) Generating $\boldsymbol{N}$ Given $(\boldsymbol{Y}, \underset{\sim}{y}, \underset{\sim}{n}, \Theta)$:

Given Θ and given $\underset{\sim}{n}$, use the multinomial distribution of $\{n_{1j}, n_{2j}\}$ given n_j to generate a large sample of $\boldsymbol{N}$. Then, by combining this sample with $P\{\boldsymbol{Y}, \underset{\sim}{y}|\boldsymbol{N}, \underset{\sim}{n}, \Theta\}$ in equation (52) to select $\boldsymbol{N}$ through the weighted bootstrap method due to Smith and Gelfant [40]. This selected $\boldsymbol{N}$ is then a sample from $P\{\boldsymbol{N}|\boldsymbol{Y}, \underset{\sim}{y}, \underset{\sim}{n}, \Theta\}$ even though the latter is unknown. (For proof, see Tan[31], Chapter 3.) Call the generated sample $\hat{\boldsymbol{N}}$.

2) Generating $\boldsymbol{Y}$ Given $(\boldsymbol{N}, \underset{\sim}{y}, \underset{\sim}{n}, \Theta)$:

Given $\{\underset{\sim}{y}, \underset{\sim}{n}, \Theta\}$ and given $\boldsymbol{N} = \hat{\boldsymbol{N}}$ generated from step (1), generate $\boldsymbol{Y}$ from the probability distribution $P\{\boldsymbol{Y}|\hat{\boldsymbol{N}}, \underset{\sim}{y}, \underset{\sim}{n}, \Theta\}$ given by equation (55). Call the generated sample $\hat{\boldsymbol{Y}}$.

3) Estimation of Θ Given $\{\boldsymbol{N}, \boldsymbol{Y}, \underset{\sim}{y}, \underset{\sim}{n}\}$:

Given $\{\underset{\sim}{y}, \underset{\sim}{n}\}$ and given $(\boldsymbol{N}, \boldsymbol{Y}) = (\hat{\boldsymbol{N}}, \hat{\boldsymbol{Y}})$ from step (1) and step (2), derive the posterior mode of Θ by maximizing the conditional posterior distribution $P\{\Theta|\hat{\boldsymbol{N}}, \hat{\boldsymbol{Y}}, \underset{\sim}{y}, \underset{\sim}{n}\}$. Under the partially informative prior, this is equivalent to maximize the negative of the deviance given by equation (58) under the constraints given in this section. Denote this generated mode by $\hat{\Theta}$. In this step, Genetic Algorithm is used to derive the posterior mode of Θ.

4) Recycling Step:

With $(\boldsymbol{N}, \boldsymbol{Y}, \Theta) = (\hat{\boldsymbol{N}}, \hat{\boldsymbol{Y}}, \hat{\Theta})$, go back to Step (1) and continue until convergence.

The proof of convergence of the above steps can be derived by using procedure given in Tan ([31], Chapter 3). At convergence, the $\hat{\Theta}$ are the generated values from the

posterior distribution of Θ given $\{\underset{\sim}{y}, \underset{\sim}{n}\}$ independently of $(\boldsymbol{N}, \boldsymbol{Y})$ (for proof, see Tan [31], Chapter 3). Repeat the above procedures one then generates a random sample of Θ from the posterior distribution of Θ given $\{\underset{\sim}{y}, \underset{\sim}{n}\}$; then one uses the sample mean as the estimates of Θ and use the sample variances and covariances as estimates of the variances and covariances of these estimates.

F. Application to Fit the SEER Data

In this section, we will apply the above model to the renal carcinoma incidence data from NCI/NIH's SEER program over the years 1973 - 2007. Given in Table 4 are the numbers of people at risk and the renal carcinoma cases in the age groups together with the predicted cases from the model. This data are incidence for 84 age groups ($k = 84$) with each group spanning over a one year period. Notice that there are few cancer cases before 10 years old implying the inclusion of some inherited cancer cases in the SEER dataset. Since our modeling in this chapter focus on the adult cancers, the quite rare cancer cases before 10 years old in the dataset are ignored. Estimates of parameters in the model are given in Table 5. The plot of the observed and predicted cancer incidence of renal cell carcinoma are shown in Figure 8. From these results, we have made the following observations:

1) As shown by results in Table 4 and Figure 8, the predicted number of cancer cases are very close to the observed cases. This indicates that the three-pathway model fits the data well and that one can safely assume that the human renal cell carcinoma can be described by a mixture model of three pathways.

2) From results in Table 5, the estimates of p_1 and p_2 from the SEER data are 0.1602 and 0.8097 respectively. This indicates that about 81% individuals in US population at risk of developing renal cell carcinoma by 4-stage pathway, 16% individuals by 3-stage pathway and 3% individuals by 5-stage pathway.

3) Results in Table 5 show that the estimates $\hat{\lambda}_j$ $(j = 1, 2, 3)$ of λ_j are of order $\{10^0, 10^{-1}, 10^1\}$ respectively. Because $\{\lambda_1 = \{\prod_{i=1}^{2} \gamma_i^{(I)}\}^{-1} E[N(t_0)] \prod_{j=0}^{2} \beta_j^{(I)}$,

$\lambda_2 = \{\prod_{i=1}^{2} \gamma_i^{(J)}\}^{-2} E[N(t_0)] \prod_{j=0}^{3} \beta_j^{(J)}$, $\lambda_3 = \{\prod_{i=1}^{3} \gamma_i^{(K)}\}^{-2} E[N(t_0)] \prod_{j=0}^{4} \beta_j^{(K)}\}$, one can have some rough ideas about the magnitude of $\beta_l^{(I)}$ $(l = 0, 1, 2)$, $\beta_u^{(J)}$ $(u = 0, 1, 2, 3)$ and $\beta_v^{(K)}$ $(v = 0, 1, 2, 3, 4)$ by assuming the value of $E[N(t_0)]$. If we follow Potten et al. [41] to assume $E[N(t_0)] \sim 10^8$, then $\beta_j^{(I)}(\beta_j^{(J)}, \beta_j^{(K)}) \approx 10^{-6} \sim 10^{-5}$.

4) From Table 5, it is observed that the estimate of $\gamma_1^{(I)}$ is of order 10^{-6} and the estimate of $\gamma_2^{(I)}$ is of order 10^{-4} which is about 100 times greater than those of I_1 cells. The estimate of $\gamma_2^{(J)}$ is of order 10^{-3}, and the estimate of $\gamma_3^{(J)}$ is of order 10^{-1} which is about 100 times greater than those of J_2 cells. The estimate of $\gamma_4^{(K)}$ is of order 10^{-2} which is about 10 times greater than the estimate of $\gamma_2^{(K)}$. These observations are due presumably to the effects of the silencing or inactivation of the cell cycle inhibition genes and the apoptosis inhibition genes.

Table 4: Renal Cell Carcinomas Incidence Data from SEER (Overall Population)

Age Groups	Number of People at Risk	Observed Incidence	Predicated Incidence
1	12221582	1	0
2	12120990	4	0
3	12112995	2	0
4	12146174	2	0
5	12161336	1	1
6	12111854	4	1
7	12160452	0	1
8	11942586	3	2
9	12381299	4	3
10	12512703	4	3
11	12410338	6	4
12	12449244	5	5

Continued on next page

Table 4 – continued from previous page

Age Groups	Number of People at Risk	Observed Incidence	Predicted Incidence
13	12527781	7	6
14	12602883	7	8
15	12719598	9	9
16	12766107	14	11
17	12831400	12	13
18	12382047	14	14
19	12581638	17	17
20	12636509	21	20
21	12682601	30	23
22	12840510	31	28
23	13075528	23	33
24	13358635	31	40
25	13473849	56	47
26	13426340	52	55
27	13525264	54	65
28	13149674	71	75
29	13812811	90	93
30	13886874	114	111
31	13488332	129	127
32	13460286	143	149
33	13256067	166	173
34	13428827	179	205
35	13220037	266	236

Continued on next page

Table 4 – continued from previous page

Age Groups	Number of People at Risk	Observed Incidence	Predicted Incidence
36	12870265	297	269
37	12689592	300	308
38	12157014	302	343
39	12494081	380	407
40	12272125	522	460
41	11826573	482	508
42	11663153	585	572
43	11407082	652	636
44	11296848	772	712
45	11016369	801	783
46	10651593	890	850
47	10475708	970	934
48	9994684	1012	991
49	10138908	1086	1114
50	9836359	1176	1192
51	9475641	1245	1262
52	9250985	1324	1348
53	9027382	1412	1433
54	8883737	1488	1531
55	8547883	1541	1592
56	8279648	1690	1660
57	8062368	1773	1734
58	7654610	1756	1759

Continued on next page

Table 4 – continued from previous page

Age Groups	Number of People at Risk	Observed Incidence	Predicted Incidence
59	7563706	1789	1850
60	7232719	1907	1876
61	6927332	1870	1898
62	6708273	1911	1936
63	6543931	1871	1981
64	6404652	1982	2028
65	6168486	1995	2035
66	5913479	1958	2027
67	5746766	1962	2039
68	5480517	2084	2007
69	5363912	1995	2021
70	5110728	2017	1975
71	4925076	1933	1947
72	4696825	1942	1893
73	4512136	1870	1849
74	4345300	1781	1805
75	4148801	1772	1742
76	3900900	1738	1652
77	3681587	1650	1567
78	3481918	1513	1487
79	3243631	1374	1385
80	2961234	1259	1262
81	2724984	1150	1155

Continued on next page

Table 4 – continued from previous page

Age Groups	Number of People at Risk	Observed Incidence	Predicted Incidence
82	2495219	1025	1050
83	2271595	885	947
84	2041351	846	841

Table 5: Estimates of Parameters for the Stochastic Model of Renal Cell Carcinoma

Parameters	p_1	p_2	$\gamma_1^{(I)}$	$\gamma_2^{(I)}$	λ_1
Estimates	1.602E-01	8.097E-01	5.457E-06	9.777E-05	1.565E+00
St.D	7.171E-05	7.823E-05	7.221E-07	1.115E-05	5.356E-02
95%CL-Lower	1.600E-01	8.095E-01	3.663E-06	7.007E-05	1.432E+00
95%CL-Upper	1.604E-01	8.100E-01	8.010E-06	1.127E-04	1.710E+00
Parameters	$\gamma_1^{(J)}$	$\gamma_2^{(J)}$	$\gamma_3^{(J)}$	δ	λ_2
Estimates	2.632E-04	3.927E-03	1.222E-01	3.554E-03	1.517E-01
St.D	1.856E-05	4.984E-04	9.652E-04	1.367E-05	1.353E-02
95%CL-Lower	2.171E-04	2.689E-03	1.198E-01	3.520E-03	1.181E-01
95%CL-Upper	2.893E-04	5.639E-03	1.250E-01	3.598E-03	1.704E-01
Parameters	$\gamma_1^{(K)}$	$\gamma_2^{(K)}$	$\gamma_3^{(K)}$	$\gamma_4^{(K)}$	λ_3
Estimates	2.879E-04	3.453E-03	8.679E-03	1.427E-02	8.853E+01
St.D	8.175E-05	2.446E-04	8.455E-04	6.165E-04	4.333E+00
95%CL-Lower	8.487E-05	2.846E-03	6.579E-03	1.274E-02	7.776E+01
95%CL-Upper	5.047E-04	4.057E-03	1.175E-02	1.586E-02	1.039E+02

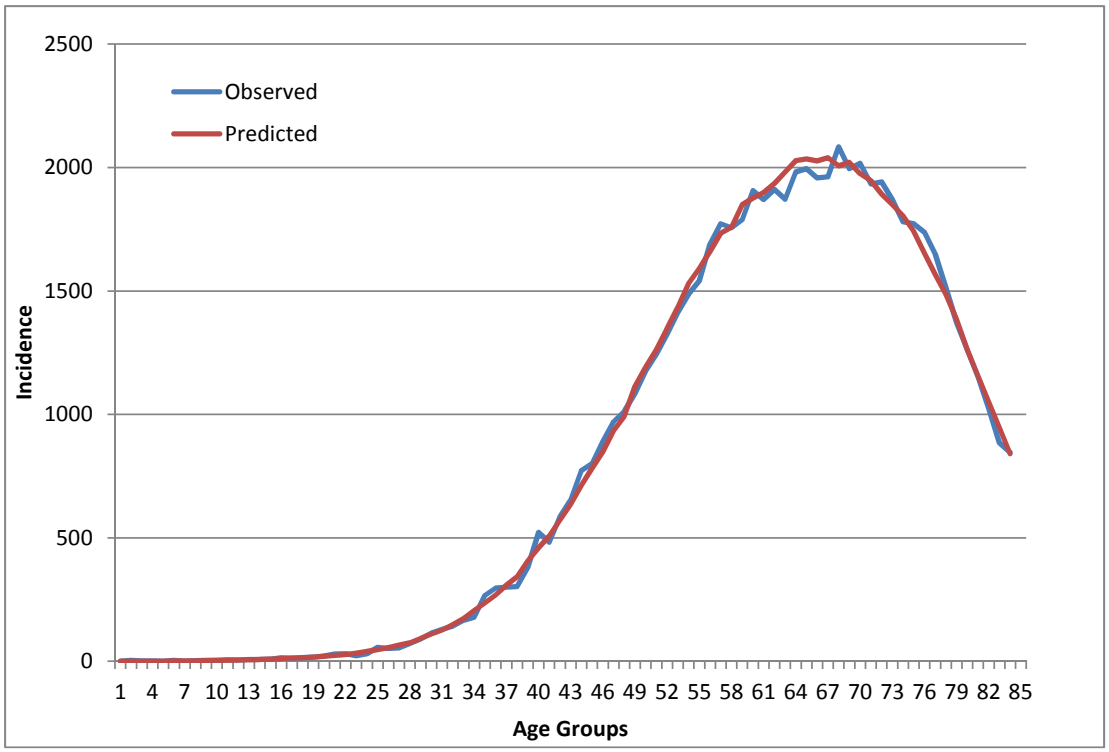

Fig. 8: Curve Fitting of Renal Cell Carcinoma SEER Data by Proposed Model

G. *Computation Details*

The multi-level Gibbs sampling procedure for estimating unknown parameters is implemented in Fortran 90. The Fortran code is shown in the Appendix C. The subroutine NGENERNOR01 are used to generate $\boldsymbol{N}$ from multinomial distribution of $\{n_{1j}, n_{2j}\}$ given n_j. Since $\underset{\sim}{n}$ is very large and $\{p_1, p_2\}$ are very small, the normal approximation is applied. The subroutine PICK is applied to select the k-th $\boldsymbol{N}$ from a large sample of $\boldsymbol{N}$ through the Weighted Bootstrap Method. The selected $\boldsymbol{N}$ is a sample from $P\{\boldsymbol{N}|\boldsymbol{Y}, \underset{\sim}{y}, \underset{\sim}{n}, \Theta\}$. $\boldsymbol{Y}$ is generated by the subroutine YGENER from the multinomial distribution. The publicly available Genetic Algorithm PIKAIA is applied to derive the posterior mode of Θ by maximizing the conditional posterior distribution $P\{\Theta|\hat{\boldsymbol{N}}, \hat{\boldsymbol{Y}}, \underset{\sim}{y}, \underset{\sim}{n}\}$. The genetic algorithms are a class of search techniques inspired from the biological process of evolution by means of natural selection. The basic principle is

that those with the largest fitness will be selected as the generation progresses. Given the fitness, the genetic algorithm would choose the parameter values to maximize the fitness according to evolutionary principle as described above. The function Fit is called in PIKAIA as fitness function that is the negative of the deviance given in (3.34). The subroutine CalculateQ is used to calculate the probability of developing tumor during each age period through different pathways.

H. Discussion and Conclusion

Based on most recent biological studies on renal cell carcinomas as discussed in Section B, in this chapter we have presented a stochastic model for carcinogenesis of renal cell carcinomas involving three different pathways, with each pathway being a multi-stage model. To account for different individuals in the population at risk of developing cancer through different pathways, we have also developed a mixture model of three pathways: (1) 3-stage model for pRCCs which account for about 15% of all RCCs, (2) 4-stage model for ccRCCs which account for about 80% of all RCCs, and (3) 5-stage model for chRCCs which account for about 5% of all RCCs.

For using the proposed model to fit the cancer incidence data, we have developed a generalized Bayesian inference procedure to estimate the unknown parameters and to predict cancer cases. This inference procedure is advantageous over the classical sampling theory inference (i.e. maximum likelihood method) because the procedure combines information from three sources: previous information and experiences about the parameters in terms of the prior distribution $P\{\Theta\}$ of the parameters; Biological information of cancer cases via different pathways for developing RCC in the population ($P[\boldsymbol{N}|\underset{\sim}{n}, p_1, p_2]$); information from the expanded data $\boldsymbol{Y}$ and the observed data $\underset{\sim}{y}$ via the statistical model from the system ($P[\boldsymbol{Y}, \underset{\sim}{y}|\boldsymbol{N}, \underset{\sim}{n}, \Theta]$).

We have applied our models and methods to the renal carcinomas incidence data from NCI/NIH's SEER program. Our analysis clearly showed that the proposed three-pathway

model fitted the data well (see Table 4 and Figure 8); The estimates from the model are consistent with biological findings.

Using models and methods of this chapter, one can easily predict future cancer cases for renal cell carcinomas. Thus, by comparing results from different populations, our models and methods can be used to assess cancer prevention and control procedures. This will be our future research topics; we will not go any further here.

CHAPTER IV

DISCUSSION AND CONCLUSION

Recent studies of cancer molecular biology have indicated that the carcinogenesis of Wilms' tumor was more complex than was first proposed by a 'two-hit' model of Knudson and Strong [18] in the early 1970's. Based on biological information, we have developed a general two-pathway stochastic model for Wilms' tumor. The fist pathway is a 3-stage model for both hereditary and non-hereditary cancer cases, in which the major genetic alterations may include WT1 mutation, IGF2 upregulation, CTNNB1 mutation and P53 mutation. The second pathway is a 2-stage model for non-hereditary cancer cases, in which WTX mutation and one more unknown genetic event may be the major genetic alterations. To account for hereditary cancer cases and the development of non-hereditary cancers through two different pathways in the stochastic model, we have also developed a generalized mixture model. In this mixture model, two mixing probability distributions were applied, which are a multinomial distribution to explain the genetic segregation of the stage-limiting tumor suppressor genes for Wilms' tumor and a binomial distribution to account for the development of non-hereditary Wilms' tumor through two pathways. We have fitted the model to the Wilms' tumor incidence data from NIH/NCI's SEER program. The fitting results have showed that the proposed two-pathway model involving hereditary and non-hereditary cancer cases fitted the data better than the single-pathway model with hereditary cancer cases. The results have confirmed the finding from molecular biology that Wilms' tumor is more genetically heterogeneous than other pediatric cancers such as retinoblastoma. In combination of generalized Bayesian approach using multi-level Gibbs sampling procedures, we have estimated the genetic segregation frequency of the stage-limiting tumor suppressor genes, the proportion of the individual at risk of developing non-hereditary cancer cases by different pathways and the proliferation rates of cells in each stage. Furthermore, we have obtained some rough ideas about the magnitude of the mutations rate of intermediate cells from the estimates of parameter functions.

Based on kidney cancer biology, we have known that the mechanism of adult kidney cancer (renal cell carcinoma) distinctly differs from that of pediatric kidney cancer. In addition, we have also found that renal cell carcinomas consist of three main histological subtypes and each subtype of renal cell carcinomas develops through different pathway. Thus, based on recent biological results, we have developed a stochastic model for human renal cell carcinoma involving three pathways, with each pathway being a multi-stage model. To account for different individuals in the population at risk of developing renal cell carcinoma through different pathways, we have also developed a mixture model of three pathways. The first pathway is a 3-stage model for pRCCs, the second pathway is a 4-stage model for ccRCCs, and the third pathway is a 5-stage model for chRCCs. We have also applied these models and procedure to the renal carcinoma incidence data from NIH/NCI's SEER program. Our results showed that the proposed multiple-pathway model fitted the data nicely. Using the stochastic model and the mixture model, we have developed a generalized Bayesian procedure to estimate the unknown parameters.

Overall, we have developed stochastic models for multiple-pathway carcinogenesis of different kidney cancers. We have developed a model to analyse tumor development for hereditary and non-hereditary Wilms' tumor, and a model for renal cell carcinoma. The stochastic models we have developed are based on biological information and hence are more realistic and applicable in practice.

In this study, we have not predicted the future cancer cases for human kidney cancer. Thus, our research in the future will be focused on developing predictive inference for cancer incidence and progression. Furthermore, by comparing results from different populations, we will apply our models and methods to assess cancer prevention and control procedures.

REFERENCES

[1] W. Tan, *Stochastic models of carcinogenesis*, vol. 116. CRC, 1991.

[2] W. Tan, C. Chen, and L. Zhang, "Cancer biology, cancer models and stochastic mathematical analysis of carcinogenesis," in *Handbook of Cancer Models and Applications.* (W. Tan and L. Hanin, eds.), ch. 3, pp. 45–90, World Scientific, River Edge, NJ, 2008.

[3] R. A. Weinberg, *The Biology of Human Cancer*. Garland Sciences, Taylor and Frances Group, New York, 2007.

[4] D. Hanahan and R. A. Weinberg, "The hallmarks of cancer," *cell*, vol. 100, no. 1, pp. 57–70, 2000.

[5] D. Hanahan and R. A. Weinberg, "Hallmarks of cancer: the next generation," *Cell*, vol. 144, no. 5, pp. 646–674, 2011.

[6] P. Armitage and R. Doll, "The age distribution of cancer and a multi-stage theory of carcinogenesis," *British journal of cancer*, vol. 8, no. 1, pp. 1–12, 1954.

[7] M. P. Little, "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*, pp. 1278–1291, 1995.

[8] M. P. Little, "Cancer models, ionizing radiation, and genomic instability: A review," in *Handbook of Cancer Models with Applications.* (W. Tan and L. Hanin, eds.), ch. 5, pp. 109–148, World Scientific, River Edge, NJ, 2008.

[9] W. Tan and H. Zhou, "A new stochastic model of retinoblastoma involvin g both hereditary and non-hereditary cancer cases," *J Carcinogene Mutagen e*, vol. 2, p. 117, 2011.

[10] W. Tan and X. Yan, "A new stochastic and state space model of human colon cancer incorporating multiple pathways," *Biology Direct*, vol. 5, p. 26, 2010.

[11] M. P. Little and E. Wright, "A stochastic carcinogenesis model incorporating genomic instability fitted to colon cancer data," *Mathematical biosciences*, vol. 183, no. 2, pp. 111–134, 2003.

[12] G. Friedman and G. Yancey Gillespie, "Cancer stem cells and pediatric solid tumors," *Cancers*, vol. 3, no. 1, pp. 298–318, 2011.

[13] V. Huff, "Wilms' tumours: about tumour suppressor genes, an oncogene and a chameleon gene," *Nat Rev Cancer*, vol. 11, no. 2, pp. 111–121, 2011.

[14] M. N. Rivera and D. A. Haber, "Wilms' tumour: connecting tumorigenesis and organ development in the kidney," *Nat Rev Cancer*, vol. 5, no. 9, pp. 699–712, 2005.

[15] K. W. Brown and K. T. Malik, "The molecular biology of wilms tumour," *Expert Rev Mol Med*, vol. 2001, pp. 1–16, 2001.

[16] J. Hagenkord, Z. Gatalica, E. Jonasch, and F. Monzon, "Clinical genomics of renal epithelial tumors," *Cancer Genetics*, vol. 204, no. 6, pp. 285–297, 2011.

[17] R. Motzer, N. Bander, and D. Nanus, "Renal-cell carcinoma," *New England Journal of Medicine*, vol. 335, no. 12, pp. 865–875, 1996.

[18] A. G. Knudson and L. C. Strong, "Mutation and cancer: a model for wilms' tumor of the kidney," *J. Natl Cancer Inst.*, vol. 48, pp. 313–324, 1972.

[19] P. Vize, A. Woolf, and J. Bard, *The kidney: from normal development to congenital disease*. Academic Pr, 2003.

[20] V. Huff, "Wilms tumor genetics," *Am. J. Med. Genet.*, vol. 79, pp. 260–267, 1998.

[21] M. N. Rivera, W. J. Kim, J. Wells, D. R. Driscoll, B. W. Brannigan, M. Han, J. C. Kim, A. P. Feinberg, W. L. Gerald, S. O. Vargas, L. Chin, A. J. Iafrate, D. W. Bell, and D. A. Haber, "An x chromosome gene, wtx, is commonly inactivated in wilms tumor," *Science*, vol. 315, no. 5812, pp. 642–645, 2007.

[22] N. Bardeesy, "Anaplastic wilms' tumour, a subtype displaying poor prognosis, harbours p53 gene mutations," *Nature Genet.*, vol. 7, pp. 91–97, 1994.

[23] R. Koesters, "Mutational activation of the -catenin proto-oncogene is a common event in the development of wilms' tumors," *Cancer Res.*, vol. 59, pp. 3880–3882, 1999.

[24] F. RAUSCHER, "The wt1 wilms tumor gene product: a developmentally regulated transcription factor in the kidney that functions as a tumor suppressor," *The FASEB journal*, vol. 7, no. 10, pp. 896–903, 1993.

[25] S. Park, A. Bernard, K. E. Bove, D. A. Sens, D. J. Hazen-Martin, A. J. Garvin, and D. A. Haber, "Inactivation of wt1 in nephrogenic rests, genetic precursors to wilms' tumour," *Nat Genet*, vol. 5, no. 4, pp. 363–367, 1993.

[26] R. Fukuzawa, R. W. Heathcott, H. E. More, and A. E. Reeve, "Sequential wt1 and ctnnb1 mutations and alterations of -catenin localization in intralobar nephrogenic rests and associated wilms tumours: two case studies," *J. Clin. Pathol.*, vol. 60, pp. 1013–1016, 2007.

[27] O. Ogawa, "Relaxation of insulin-like growth factor ii gene imprinting implicated in wilms' tumour," *Nature*, vol. 362, pp. 749–751, 1993.

[28] M. J. Steenman, "Loss of imprinting of igf2 is linked to reduced expression and abnormal methylation of h19 in wilms' tumour," *Nature Genet.*, vol. 7, pp. 433–439, 1994.

[29] Q. Hu, F. Gao, W. Tian, E. Ruteshouser, Y. Wang, A. Lazar, J. Stewart, L. Strong, R. Behringer, and V. Huff, "Wt1 ablation and igf2 upregulation in mice result in wilms tumors with elevated erk1/2 phosphorylation," *The Journal of clinical investigation*, vol. 121, no. 1, pp. 174–183, 2011.

[30] J. Crow and M. Kimura, *An introduction to population genetics theory*. New York, Evanston and London: Harper & Row, Publishers, 1970.

[31] W. Tan, *Stochastic models with applications to genetics, cancers, AIDS and other biomedical systems*, vol. 4. World Scientific Pub Co Inc, 2002.

[32] G. Yang and C. Chen, "A stochastic two-stage carcinogenesis model: a new approach to computing the probability of observing tumor in animal bioassays," *Mathematical biosciences*, vol. 104, no. 2, pp. 247–258, 1991.

[33] Q. Zheng, "Stochastic multistage cancer models: A fresh look at an old approach," in *Handbook of Cancer Models and Applications.* (W. Tan and L. Hanin, eds.), ch. 2, pp. 25–44, World Scientific, River Edge, NJ, 2008.

[34] W. Tan, "Stochastic multiti-stage models of carcinogenesis as hidden markov models: A new approach," *Int. J. Systems and Synthetic Biology*, vol. 313-337, 2010.

[35] W. Tan, L. Zhang, and C. Chen, "Stochastic modeling of carcinogenesis: state space models and estimation of parameters," *Discrete and Continuous Dynamical Systems Series B*, vol. 4, no. 1, pp. 297–322, 2004.

[36] W. Tan, C. Chen, and L. Zhang, "Cancer risk assessment of environmental agents by stochastic and state space models of carcinogenesis," in *Handbook of Cancer Models and Applications.* (W. Tan and L. Hanin, eds.), ch. 12, pp. 375–396, World Scientific, River Edge, NJ, 2008.

[37] W. Tan, C. Chen, and L. Zhang, "A stochastic model of human colon cancer involving multiple pathways," in *Handbook of Cancer Models and Applications.* (W. Tan and L. Hanin, eds.), ch. 11, pp. 345–374, World Scientific, River Edge, NJ, 2008.

[38] G. Box and G. Tiao, *Bayesian inference in statistical analysis*. Wiley Online Library, 1992.

[39] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 1–38, 1977.

[40] A. Smith and A. Gelfand, "Bayesian statistics without tears: a sampling-resampling perspective," *American statistician*, pp. 84–88, 1992.

[41] C. Potten, C. Booth, and D. Hargreaves, "The small intestine as a model for evaluating adult tissue stem cell drug targets1," *Cell proliferation*, vol. 36, no. 3, pp. 115–129, 2003.

[42] W. Kaelin Jr, "The von hippel-lindau tumor suppressor gene and kidney cancer," *Clinical cancer research*, vol. 10, no. 18, pp. 6290S–6295S, 2004.

[43] S. Clifford, A. Prowse, N. Affara, C. Buys, and E. Maher, "Inactivation of the von hippel–lindau (vhl) tumour suppressor gene and allelic losses at chromosome arm 3p in primary renal cell carcinoma: Evidence for a vhl-independent pathway in clear cell renal tumourigenesis," *Genes, Chromosomes and Cancer*, vol. 22, no. 3, pp. 200–209, 1998.

[44] S. Mandriota, K. Turner, D. Davies, P. Murray, N. Morgan, H. Sowter, C. Wykoff, E. Maher, A. Harris, P. Ratcliffe, *et al.*, "Hif activation identifies early lesions in vhl kidneys: evidence for site-specific tumor suppressor function in the nephron," *Cancer cell*, vol. 1, no. 5, pp. 459–468, 2002.

[45] I. Varela, P. Tarpey, K. Raine, D. Huang, C. Ong, P. Stephens, H. Davies, D. Jones, M. Lin, J. Teague, *et al.*, "Exome sequencing identifies frequent mutation of the swi/snf complex gene pbrm1 in renal carcinoma," *Nature*, vol. 469, no. 7331, pp. 539–542, 2011.

[46] G. Banumathy and P. Cairns, "Signaling pathways in renal cell carcinoma," *Cancer biology & therapy*, vol. 10, no. 7, pp. 658–664, 2010.

[47] A. Szponar, D. Zubakov, J. Pawlak, A. Jauch, and G. Kovacs, "Three genetic developmental stages of papillary renal cell tumors: Duplication of chromosome 1q marks fatal progression," *International Journal of Cancer*, vol. 124, no. 9, pp. 2071–2076, 2009.

[48] M. R. Speicher, B. Schoell, S. du Manoir, E. Schröck, T. Ried, T. Cremer, S. Störkel, A. Kovacs, and G. Kovacs, "Specific loss of chromosomes 1, 2, 6, 10, 13, 17, and 21 in chromophobe renal cell carcinomas revealed by comparative genomic hybridization," *The American journal of pathology*, vol. 145, no. 2, pp. 356–364, 1994.

[49] M. Brunelli, J. Eble, S. Zhang, G. Martignoni, B. Delahunt, and L. Cheng, "Eosinophilic and classic chromophobe renal cell carcinomas have similar frequent losses of multiple chromosomes from among chromosomes 1, 2, 6, 10, and 17, and this pattern of genetic abnormality is not present in renal oncocytoma," *Modern pathology*, vol. 18, no. 2, pp. 161–169, 2004.

[50] A. Yakovlev, A. Tsodikov, and B. Asselain, *Stochastic models of tumor latency and their biostatistical applications*, vol. 1. World Scientific Pub Co Inc, 1996.

APPENDIX A

DERIVATION OF $\{Q_1(J), Q_2(J), Q_0^{(I)}(J), Q_0^{(J)}(J)\}$ BY DISCRETE APPROXIMATION

Under discrete time with one time unit (i.e. $\Delta t = 1$) corresponding to 3 months or longer, one may practically assume $P_T(s,t) = 1$ for $t - s \geq 1$; further $Q_1(j)$, $Q_2(j)$, $Q_0^{(I)}(j)$, $Q_0^{(J)}(j)$ are approximated by:

$$\begin{aligned}
Q_0^{(I)}(j) &= \{e^{-\beta_2^{(I)} \sum_{s=t_0}^{t_{j-1}-1} E[I_2(x;0)]} - e^{-\beta_2^{(I)} \sum_{s=t_0}^{t_j-1} E[I_2(x;0)]}\} + o(\beta_2^{(I)}) \\
Q_0^{(J)}(j) &= \{e^{-\beta_1^{(J)} \sum_{s=t_0}^{t_{j-1}-1} E[J_1(x)]} - e^{-\beta_1^{(J)} \sum_{s=t_0}^{t_j-1} E[J_1(x)]}\} + o(\beta_1^{(J)}) \\
Q_1(j) &= \{e^{-\beta_2^{(I)} \sum_{s=t_0}^{t_{j-1}-1} E[I_2(x;1)]} - e^{-\beta_2^{(I)} \sum_{s=t_0}^{t_j-1} E[I_2(x;1)]}\} + o(\beta_2^{(I)}) \\
Q_2(j) &= (1-\alpha_1)\{e^{-\beta_2^{(I)} \sum_{s=t_0}^{t_{j-1}-1} E[I_2(x;2)]} - e^{-\beta_2^{(I)} \sum_{s=t_0}^{t_j-1} E[I_2(x;2)]}\} + o(\beta_2^{(I)})
\end{aligned}$$

Under discrete time, the stochastic differential equations for staging state variables become stochastic difference equations for these state variables respectively. Thus, for deriving $\{E[I_2(t;i)]\ (i = 0, 1, 2), E[J_1(t)]\}$ and $\{Q_1(j), Q_2(j), Q_0^{(I)}(j), Q_0^{(J)}(j)\}$ under discrete time, we have the following difference equation for $I_2(t;2)$. This stochastic difference equation is derived from equation (5) by putting $\Delta t = 1$:

$$\begin{aligned}
I_2(t+1;2) &= I_2(t;2) + B_2(t;2) - D_2(t;2) \\
&= I_2(t;2)(1+\gamma_2^{(I)}) + \epsilon_2(t;2), t > t_0, \qquad (59)
\end{aligned}$$

where $\epsilon_2(t;2) = [B_2(t;2) - I_2(t;2)b_2] - [D_2(t;2) - I_2(t;2)d_2]$.

From the above equation (59), we obtain:

$$E[I_2(t;2)] = E[I_2(t-1;2)](1+\gamma_2^{(I)}) = \cdots = E[I_2(t_0;2)](1+\gamma_2^{(I)})^{t-t_0}.$$

Put $\phi_{22}(t) = \gamma_2^{(I)} \sum_{s=t_0}^{t-1} (1+\gamma_2^{(I)})^{s-t_0}$ $(t-1 > t_0)$. Using the result $\sum_{i=0}^{t-1} a^i = \frac{a^t-1}{a-1}$, we obtain:

$$\phi_{22}(t) \;=\; \gamma_2^{(I)} \sum_{s=t_0}^{t-1} (1+\gamma_2^{(I)})^{s-t_0} = \gamma_2^{(I)} \sum_{s=0}^{t-t_0-1} (1+\gamma_2^{(I)})^{s} = (1+\gamma_2^{(I)})^{t-t_0} - 1.$$

It follows that

$$Q_2(j) \;\approx\; (1-\alpha_1)\{e^{-\lambda_1 \phi_{22}(t_{j-1})} - e^{-\lambda_1 \phi_{22}(t_j)}\},$$

where $\lambda_1 = \frac{1}{\gamma_2^{(I)}} E[I_2(t_0;2)]$.

For deriving $E[I_2(t;1)]$ and $Q_1(j)$ under discrete time, from equations (5)-(6) we obtain the following difference equations for $\{I_1(t;1), I_2(t;1)\}$.

$$\begin{aligned} I_1(t+1;1) &= I_1(t;1) + B_1(t;1) - D_1(t;1) \\ &= I_1(t;1)(1+\gamma_1^{(I)}) + \epsilon_1(t+1;1), \qquad (60) \\ I_2(t+1;1) &= I_2(t;1) + M_1(t;1) + B_2(t;1) - D_2(t;1) \\ &= I_2(t;1)(1+\gamma_2^{(I)}) + I_1(t;1)\beta_1^{(I)} + \epsilon_2(t+1;1), \qquad (61) \end{aligned}$$

where

$$\begin{aligned} \epsilon_1(t+1;1) &= [B_1(t;1) - I_1(t;1)b_1] - [D_1(t;1) - I_1(t;1)d_1], \\ \epsilon_2(t+1;1) &= [M_1(t;1) - I_1(t;1)\beta_1^{(I)}] + [B_2(t;1) - I_2(t;1)b_2] \\ &- [D_2(t;1) - I_2(t;1)d_2]. \end{aligned}$$

From equation (60), we obtain $E[I_1(t;1)] = E[I_1(t_0;1)](1+\gamma_1^{(I)})^{t-t_0}$. From equation (61), we obtain:

$$
\begin{aligned}
& E[I_2(t;1)] \\
= \;& E[I_2(t_0;1)](1+\gamma_2^{(I)})^{t-t_0} + \beta_1^{(I)} \sum_{s=t_0}^{t-1} E[I_1(s;1)](1+\gamma_2^{(I)})^{t-1-s} \\
= \;& E[I_2(t_0;1)](1+\gamma_2^{(I)})^{t-t_0} + \beta_1^{(I)} E[I_1(t_0;1)] \sum_{s=t_0}^{t-1} (1+\gamma_1^{(I)})^{s-t_0} (1+\gamma_2^{(I)})^{t-1-s} \\
= \;& E[I_2(t_0;1)](1+\gamma_2^{(I)})^{t-t_0} + \beta_1^{(I)} E[I_1(t_0;1)](1+\gamma_2^{(I)})^{t-1-t_0} \sum_{s=t_0}^{t-1} (\frac{1+\gamma_1^{(I)}}{1+\gamma_2^{(I)}})^{s-t_0} \\
= \;& E[I_2(t_0;1)](1+\gamma_2^{(I)})^{t-t_0} + E[I_1(t_0;1)]\beta_1^{(I)} \sum_{u=1}^{2} A_{12}(u)(1+\gamma_u^{(I)})^{t-t_0}.
\end{aligned}
$$

From the above expected numbers, it follows that

$$
Q_1(j) \;\approx\; \{e^{-\theta\phi_{22}(t_{j-1})-\lambda_2\phi_{12}(t_{j-1})} - e^{-\theta\phi_{22}(t_j)-\lambda_2\phi_{12}(t_j)}\},
$$

where $\{\lambda_2 = \frac{1}{\gamma_1^{(I)}\gamma_2^{(I)}} E[I_1(t_0;1)]\beta_1^{(I)}\beta_2^{(I)},\ \theta = \frac{1}{\gamma_2^{(I)}} E[I_2(t_0;1)]\beta_2^{(I)}\}$, and

$$
\begin{aligned}
\phi_{12}(t) &= \{\prod_{i=1}^{2} \gamma_i^{(I)}\} \sum_{u=1}^{2} A_{12}(u) \frac{1}{\gamma_u^{(I)}} \{(1+\gamma_u^{(I)})^{(t-t_0)} - 1\}, \\
\phi_{22}(t) &= (1+\gamma_2^{(I)})^{t-t_0} - 1.
\end{aligned}
$$

For deriving $E[I_2(t;0)]$ and $Q_0^{(I)}(j)$ $(i=0,1)$ under discrete time, from equations (5)-(6) we obtain the following difference equations for $\{I_0(t), I_1(t), I_2(t)\}$ with initial conditions $\{I_0(t_0) = N(t_0), I_i(t_0) = I_0(t_0;0) = 0,\ i = 1,2\}$ at birth (t_0):

$$
\begin{aligned}
I_0(t+1;0) &= I_0(t;0) + B_0(t;0) - D_0(t;0) = I_0(t;0)(1+\gamma_0^{(I)}) + \epsilon_0(t+1;0), \\
&\quad \text{with } I_0(t_0;0) = N(t_0), \qquad (62) \\
I_i(t+1;0) &= I_i(t;0) + M_{i-1}(t;0) + B_i(t;0) - D_i(t;0) = I_i(t;0)(1+\gamma_i^{(I)}) \\
&+ I_{i-1}(t;0)\beta_{i-1}^{(I)} + \epsilon_i(t+1;0), i = 1,2. \qquad (63)
\end{aligned}
$$

where

$$\begin{aligned}
\epsilon_0^{(I)}(t+1;0) &= [B_0(t;0) - I_0(t;0)b_0] - [D_0(t;0) - I_0(t;0)d_0], \\
\epsilon_i^{(I)}(t+1;0) &= [M_{i-1}(t;0) - I_{i-1}(t;0)\beta_{i-1}] + [B_i(t;0) - I_i(t;0)b_i] \\
&- [D_i(t;0) - I_i(t;0)d_i], i = 1, 2.
\end{aligned}$$

Practically we will assume $\gamma_0^{(I)} = 0$ after birth. Hence, from equation (62), $E[I_0(t;0)] = E[I_0(t_0;0)] = E[N(t_0)]$. From equation (63), we obtain:

$$\begin{aligned}
E[I_1(t+1;0)] &= E[I_1(t;0)](1+\gamma_1^{(I)}) + E[N(t_0)]\beta_0^{(I)}, \\
E[I_2(t+1;0)] &= E[I_2(t;0)](1+\gamma_2^{(I)}) + E[I_1(t;0)]\beta_1^{(I)}, t \geq t_0.
\end{aligned}$$

The solution of $E[I_1(t;0)]$ under the initial condition $I_1(t_0;0) = 0$ is

$$\begin{aligned}
E[I_1(t;0)] &= I_1(t_0)(1+\gamma_1^{(I)})^{t-t_0} + E[N(t_0)]\beta_0^{(I)} \sum_{s=0}^{t-t_0-1} (1+\gamma_1^{(I)})^s \\
&= \frac{E[N(t_0)]\beta_0^{(I)}}{\gamma_1^{(I)}}\{(1+\gamma_1^{(I)})^{t-t_0} - 1\} \text{ if } \gamma_1^{(I)} \neq 0;
\end{aligned}$$

The solution of $E[I_2(t;0)]$ under the initial conditions $\{I_0(t_0;0) = N(t_0), I_i(t_0;0) = 0, i = 1, 2\}$ is

$$\begin{aligned}
E[I_2(t;0)] &= I_2(t_0)(1+\gamma_2^{(I)})^{t-t_0} + \beta_1^{(I)} \sum_{s=t_0}^{t-1} E[I_1(s;0)](1+\gamma_2^{(I)})^{t-s-1} \\
&= \frac{E[N(t_0)]\beta_0^{(I)}\beta_1^{(I)}}{\gamma_1^{(I)}} \sum_{s=t_0}^{t-1} [(1+\gamma_1^{(I)})^{s-t_0} - 1](1+\gamma_2^{(I)})^{t-s-1} \\
&= E[N(t_0)]\beta_0^{(I)}\beta_1^{(I)} \sum_{u=1}^{2} A_{12}(u) \frac{1}{\gamma_u^{(I)}}\{(1+\gamma_u^{(I)})^{t-t_0} - 1\}.
\end{aligned}$$

From this result, it follows that

$$Q_0^{(I)}(j) \approx e^{-\lambda_3 \phi_{02}(t_{j-1})} - e^{-\lambda_3 \phi_{02}(t_j)},$$

where $\lambda_3 = \{\prod_{j=1}^{2} \gamma_j^{(I)}\}^{-1} E[N(t_0)](\prod_{i=0}^{2} \beta_i^{(I)})$, and

$$\phi_{02}(t) = \{\prod_{i=1}^{2} \gamma_i^{(I)}\} \sum_{u=1}^{2} A_{12}(u) \frac{1}{{\gamma_u^{(I)}}^2} \{(1+\gamma_u^{(I)})^{(t-t_0)} - 1 - \gamma_u^{(I)}(t-t_0)\}$$

For deriving $E[J_1(t)]$ and $Q_0^{(J)}(j)$ under discrete time, from equations (7) we obtain the following difference equations for $\{J_0(t), J_1(t)\}$ with initial conditions $\{J_0(t_0) = N(t_0), J_1(t_0) = 0\}$ at birth (t_0):

$$\begin{aligned} J_0(t+1) &= J_0(t) + B_0(t) - D_0(t) = J_0(t)(1+\gamma_0^{(J)}) + \epsilon_0(t+1), \\ &\quad \text{with } J_0(t_0) = N(t_0), \qquad (64) \\ J_1(t+1) &= J_1(t) + M_0(t) + B_1(t) - D_1(t) = J_1(t)(1+\gamma_1^{(J)}) \\ &+ J_0(t)\beta_0^{(J)} + \epsilon_i(t+1). \qquad (65) \end{aligned}$$

where

$$\begin{aligned} \epsilon_0^{(J)}(t+1) &= [B_0(t) - J_0(t)b_0] - [D_0(t) - J_0(t)d_0], \\ \epsilon_1^{(J)}(t+1) &= [M_0(t) - J_0(t)\beta_0] + [B_1(t) - J_1(t)b_1] \\ &- [D_1(t) - J_1(t)d_1]. \end{aligned}$$

Practically we will assume $\gamma_0^{(J)} = 0$ after birth. Hence, from equation (64), $E[J_0(t)] = E[J_0(t_0)] = E[N(t_0)]$. From equation (65), we obtain:

$$E[J_1(t+1)] = E[J_1(t)](1+\gamma_1^{(J)}) + E[N(t_0)]\beta_0^{(J)}.$$

The solution of $E[J_1(t)]$ under the initial condition $J_1(t_0) = 0$ is

$$\begin{aligned} E[J_1(t)] &= J_1(t_0)(1+\gamma_1^{(J)})^{t-t_0} + E[N(t_0)]\beta_0^{(J)} \sum_{s=0}^{t-t_0-1} (1+\gamma_1^{(J)})^s \\ &= \frac{E[N(t_0)]\beta_0^{(J)}}{\gamma_1^{(J)}}\{(1+\gamma_1^{(J)})^{t-t_0} - 1\} \text{ if } \gamma_1^{(J)} \neq 0. \end{aligned}$$

From this result, it follows that

$$Q_0^{(J)}(j) \approx \{e^{-\lambda_4 \phi_{01}(t_{j-1})} - e^{-\lambda_4 \phi_{01}(t_j)}\} + o(\beta_1^{(J)}),$$

where $\lambda_4 = \frac{1}{\gamma_1^{(J)^2}} N(t_0)\beta_0^{(J)}\beta_1^{(J)}$, and

$$\phi_{01}(t) = \{(1+\gamma_1^{(J)})^{(t-t_0)} - 1 - \gamma_1^{(J)}(t-t_0)\}, \text{ if } \gamma_1^{(J)} \neq 0.$$

APPENDIX B

PROGRAM CODE TO FIT A MODEL OF WILMS' TUMOR

```
PROGRAM WimlsTumor
USE RNNOR_INT
USE RNBIN_INT
USE UMACH_INT
USE RNSET_INT
USE FAC_INT
IMPLICIT NONE
INTEGER, PARAMETER:: NB=500, n_fit2 = 12 , MaxAge = 84, dt=4
INTEGER SEED(NB), ISEED, K, control, Y(MaxAge),Y1(1:NB),Y2
    (1:NB),Y3(1:NB),Y31(1:NB),Y32(1:NB),N(MaxAge),N1(MaxAge),
    N2(MaxAge),N3(MaxAge), N31(MaxAge), N32(MaxAge), i, j,
    pikaia_status, n_loop
INTEGER Y0,N0,N10,N20,N30,N1Temp,N2Temp,N3Temp,N31Temp,
    N32Temp,kk,Y_Orig(MaxAge), N_Orig(MaxAge),qstatus,T
DOUBLE PRECISION Q1(1:NB),Q2(1:NB), Q3(1:NB), Q31(1:NB), Q32
    (1:NB), P22(MaxAge),P030(MaxAge),P130(MaxAge),P131(MaxAge
    ),P231(MaxAge),P232(MaxAge),P020(MaxAge), Q(1:NB),N1_pred
    (MaxAge),N2_pred(MaxAge),N3_pred(MaxAge),N31_pred(MaxAge)
    ,N32_pred(MaxAge)
DOUBLE PRECISION mylambda(1:4),mygamma1,mydelta, myP, myP1,
    myP2,myalpha2,myalpha1,myalpha0,myalpha3,mygamma2,
    mygamma3,mytheta1,mytheta2,y0_pred, fit2_result,
    Q_from_pikaia_output(1:MaxAge)
DOUBLE PRECISION,DIMENSION(1:NB,1:MaxAge)::H, N1B, N2B, N3B,
    N31B,N32B
DOUBLE PRECISION,DIMENSION(1:MaxAge)::N1_PICKED,N2_PICKED,
    N3_PICKED, N31_PICKED,N32_PICKED,K_PICKED,Y1_PICKED,
    Y2_PICKED, Y3_PICKED,Y31_PICKED,Y32_PICKED
DOUBLE PRECISION ,DIMENSION(1:3):: PRange, P1Range, P2Range,
    alpha2Range,alpha1Range, alpha0Range,alpha3Range,
    gamma1Range,gamma2Range,lambda1Range, lambda2Range,
    lambda3Range,lambda4Range,deltaRange,gamma3Range,
    theta1Range,theta2Range
DOUBLE PRECISION,DIMENSION(1:12) :: old_para
DOUBLE PRECISION tmyP, tmyalpha2, tmygamma1, tmylambda(2),
    tmydelta,chisq
REAL ctrl(12), pikaia_fit2_x(n_fit2),pikaia_fit2_f, fit2out,
     normalized_para(1:14)
INCLUDE 'Wilms85_3_2_stage.FI'
DO i = 1, MaxAge
    Y(i) = Y_Orig(i)
```

```
   N(i) = N_Orig(i)
END DO
open (unit=5,file='output.dat',status='unknown')
open (unit=10,file='output_pick.dat',status='unknown')
iseed=12345
CALL RNSET(iseed)
CALL rninit(iseed)

DO n_loop= 1, 50
 DO i = 1, MaxAge
 DO kk=1, NB
  iseed=iseed+1
 CALL NGENERNOR_01(N(i),myP1,myP2,ISEED,N1Temp,N2Temp,N3Temp
     )
 N1(i) = N1Temp
 N2(i) = N2Temp
 N3(i) = N3Temp
 N1B(kk,i)=N1Temp
 N2B(kk,i)=N2Temp
 N3B(kk,i)=N3Temp
 CALL NGENERNOR(N3(i),myalpha3,ISEED,N32Temp,N31Temp)
 N31(i) = N31Temp
 N32(i) = N32Temp
 N31B(kk,i)=N31Temp
 N32B(kk,i)=N32Temp
 END DO
 END DO

 DO i = 1, MaxAge
 IF (Y(i)==0) THEN
 K_PICKED(i) = 1
 Y1_PICKED(i) = 0
 Y2_PICKED(i) = 0
 Y3_PICKED(i) = 0
 N1_PICKED(i) = N1B(K_PICKED(i),i)
 N2_PICKED(i) = N2B(K_PICKED(i),i)
 N3_PICKED(i) = N3B(K_PICKED(i),i)
 Y31_PICKED(i) = 0
 Y32_PICKED(i) = 0
 N31_PICKED(i) = N31B(K_PICKED(i),i)
 N32_PICKED(i) = N32B(K_PICKED(i),i)
 CYCLE
 END IF
 DO j = 1, NB
 CALL CalculateQ(myalpha2, mygamma1, mylambda, myalpha3,
```

```
     mygamma2, mygamma3, mytheta2, N1B(j,i), N2B(j,i), N3B(j,
     i), N31B(j,i), N32B(j,i), i, Q(j), Q1(j), Q2(j), Q3(j),
     Q31(j), Q32(j), qstatus)
  END DO

  IF(qstatus .EQ. 1) THEN
  mygamma1 = old_para(1)
  mylambda(1) = old_para(2)
  mylambda(2) = old_para(3)
  mylambda(3) = old_para(4)
  myalpha2 = old_para(5)
  myP1 = old_para(6)
  mylambda(4) = old_para(7)
  myalpha3 = old_para(8)
  mygamma2 = old_para(9)
  mygamma3 = old_para(10)
  mytheta2 = old_para(11)
  myP2 = old_para(12)
  GOTO 300
  END IF

  DO j = 1, NB
  CALL YGENER(Y(i),Q1(j),Q2(j),Q3(j),ISEED,Y1(j),Y2(j),Y3(j))
  CALL Y3GENER(Y3(j),Q31(j),Q32(j),ISEED,Y31(j),Y32(j))
  END DO

  CALL PICK(Y(i),Y1,Y2,Y31,Q1, Q2,Q31,Q32,K)
  K_PICKED(i) = K

  IF(K_PICKED(i) == 0)
   K_PICKED(i) = 1
   Y1_PICKED(i) = Y1(K)
   Y2_PICKED(i) = Y2(K)
   Y3_PICKED(i) = Y3(K)
   N1_PICKED(i) = N1B(K_PICKED(i),i)
   N2_PICKED(i) = N2B(K_PICKED(i),i)
   N3_PICKED(i) = N3B(K_PICKED(i),i)
   Y31_PICKED(i) = Y31(K)
   Y32_PICKED(i) = Y32(K)
   N31_PICKED(i) = N31B(K_PICKED(i),i)
   N32_PICKED(i) = N32B(K_PICKED(i),i)
  END DO

 DO 10 i=1,12
  ctrl(i) = -1
```

```
10 continue
ctrl(1)=100
ctrl(2)=10000
ctrl(6)=0.005
ctrl(7)=0.0005
ctrl(8)=0.2

old_para(1) = mygamma1
old_para(2) = mylambda(1)
old_para(3) = mylambda(2)
old_para(4) = mylambda(3)
old_para(5) = myalpha2
old_para(6) = myp1
old_para(7) = mylambda(4)
old_para(8) = myalpha3
old_para(9) = mygamma2
old_para(10) = mygamma3
old_para(11) = mytheta2
old_para(12) = myp2

CALL pikaia(Fit2,n_fit2,ctrl,pikaia_fit2_x,pikaia_fit2_f,
    pikaia_status)
CALL Q_from_pikaia(DBLE(pikaia_fit2_x),n_fit2,N,
    Q_from_pikaia_output)
mygamma1 = DenormalizeX(DBLE(pikaia_fit2_x(1)),gamma1Range)
mylambda(1) = DenormalizeX(DBLE(pikaia_fit2_x(2)),
    lambda1Range)
mylambda(2) = DenormalizeX(DBLE(pikaia_fit2_x(3)),
    lambda2Range)
mylambda(3) = DenormalizeX(DBLE(pikaia_fit2_x(4)),
    lambda3Range)
myalpha2 = DenormalizeX(DBLE(pikaia_fit2_x(5)),alpha2Range)
myp1 = DenormalizeX(DBLE(pikaia_fit2_x(6)),p1Range)
mylambda(4) = DenormalizeX(DBLE(pikaia_fit2_x(7)),
    lambda4Range)
myalpha3 = DenormalizeX(DBLE(pikaia_fit2_x(8)),alpha3Range)
mygamma2 = DenormalizeX(DBLE(pikaia_fit2_x(9)),gamma2Range)
mygamma3 = DenormalizeX(DBLE(pikaia_fit2_x(10)),gamma3Range)
mytheta2 = DenormalizeX(DBLE(pikaia_fit2_x(11)),theta2Range)
myp2 = DenormalizeX(DBLE(pikaia_fit2_x(12)),p2Range)

y0_pred=n0*myp2*myalpha2
write (5, *) y0_pred
do i = 1,MaxAge
write (5,*) Q_from_pikaia_output(i)
```

```
end do
 write (5,'(A,I5,$)') ' n_loop ',n_loop
 write (5,'(A,ES14.6)') ' Fit2 ',pikaia_fit2_f
 write (5,'(A,ES14.6,$)') ' p1 ',myp1
 write (5,'(A,ES14.6,$)') ' p2 ',myp2
 write (5,'(A,ES14.6,$)') ' alpha2 ',myalpha2
 write (5,'(A,ES14.6,$)') ' gamma1 ',mygamma1
 write (5,'(A,ES14.6,$)') ' gamma2 ',mygamma2
 write (5,'(A,ES14.6,$)') ' lambda1 ',mylambda(1)
 write (5,'(A,ES14.6,$)') ' lambda2 ',mylambda(2)
 write (5,'(A,ES14.6,$)') ' lambda3 ',mylambda(3)
 write (5,'(A,ES14.6,$)') ' lambda4 ',mylambda(4)
 write (5,'(A,ES14.6,$)') ' alpha3 ',myalpha3
 write (5,'(A,ES14.6,$)') ' gamma3 ',mygamma3
 write (5,'(A,ES14.6,$)') ' theta2 ',mytheta2
 write (5,'(A,ES14.6,$)') ' Fit2 ',pikaia_fit2_f
 write (5,*) 'End'
 write (*,*) 'Loop'
 300 END DO ! End Main loop

CONTAINS
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

!Subroutine: NGENERNOR_01
!Description: Generate (n1,n2,n3) from multinomial
!             distribution. Since n is very large and p is
!             very small, the normal approximation is
   applied
!Input: N = population at each age period
!       P1 = the proportion for I1 people in population
!       P2 = the proportion for I2 people in population
!       ISEED = seed for generating random number
!Output: N1 = the number of I1 people
!        N2 = the number of I2 people
!        N3 = the number of normal people
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

SUBROUTINE NGENERNOR_01(N,P1,P2,ISEED,N1,N2,N3)
  DOUBLE PRECISION, INTENT(IN) :: P1,P2
  INTEGER, INTENT(IN) ::ISEED,N
  INTEGER, INTENT(OUT) :: N1,N2,N3
  REAL NOR1,NOR2
  REAL IR(1)
  DO
  CALL RNNOR(IR)
```

```
  NOR1=IR(1)
  CALL RNNOR(IR)
  NOR2=IR(1)
  IF (NOR1 < 3 .AND. NOR1 > -3) EXIT
  IF (NOR2 < 3 .AND. NOR2 > -3) EXIT
  END DO
  N1=NOR1 * sqrt(N*p2*(1-p2))+ N*p2
  N2=NOR2 * sqrt((N-N1)*p1*(1-P1-P2)/(1-P2)/(1-P2))+ (N-N1)*
     p1/(1-p2)
  N3=N-N1-N2
  IF(N1 < 0) THEN
    N1 = 1e-7
  END IF
  RETURN
END SUBROUTINE NGENERNOR_01

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

!Subroutine: NGENERNOR_01
!Description: Generate (n31,n32) from binomial
!             distribution. Since n is very large and p is
!             very small, the normal approximation is
   applied
!Input: N = the number of people in population who are
   normal
!           at the embryo stage at each age period
!       P1 = the proportion for normal people in population
!            who develop tumor by 2-stage pathway
!       ISEED = seed for generating random number
!Output: N3 = the number of normal people at risk of
   developing
!              tumor by 3-stage pathway
!        N2 = the number of normal people at risk of
   developing
!              tumor by 2-stage pathway
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

SUBROUTINE NGENERNOR(N,P,ISEED,N2,N3)
  IMPLICIT NONE
  DOUBLE PRECISION, INTENT(IN) :: P
  INTEGER, INTENT(IN) ::ISEED,N
  INTEGER, INTENT(OUT) :: N2, N3
  REAL NOR2,NOR3
  REAL IR(1)
  DO
```

```
   CALL RNNOR(IR)
   NOR2=IR(1)
   IF (NOR2 < 3 .AND. NOR2 > -3) EXIT
   END DO
   N2=NOR2*sqrt(N*(2*p)*(1-p)/(1+p)/(1+p))+N*2*p/(1+p)
   N3= N-N2
   RETURN
END SUBROUTINE NGENERNOR

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!Subroutine: YGENER
!Description: Generate (y1,y2,y3) from multinomial
!             distribution with parameters {Y;Q1n/Qn,Q2n/Qn}
!Input: Y = the observed cancer cases at each age period
!       Q1n = the product of N1 and the probability of
!             developing tumor during each age period in
!             people who have genotype I1 at embryo stage
!       Q2n = the product of N2 and the probability of
!             developing tumor during each age period in
!             people who have genotype I2 at embryo stage
!       Q3n = the product of N3 and the probability of
!             developing tumor during each age period in
!             people who normal people at embryo stage
!       ISEED = seed for generating random number
!Output: Y1 = number of cancer cases generated by people
!             who have genotype I1 at embryo stage
!        Y2 = number of cancer cases generated by people
!             who have genotype I2 at embryo stage.
!        Y3 = number of cancer cases generated by people
!             who are normal people at embryo stage
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
SUBROUTINE YGENER(Y,Q1n,Q2n,Q3n,ISEED,Y1,Y2,Y3)
   IMPLICIT  NONE
   DOUBLE PRECISION, INTENT(IN) :: Q1n,Q2n,Q3n
   INTEGER, INTENT(IN) ::ISEED, Y
   INTEGER, INTENT(OUT) :: Y1,Y2,Y3
   REAL P1,P2,P3,Qn
   INTEGER IR(1)
   Qn = Q1n +Q2n + Q3n
   P1=Q1n/Qn
   P2=Q2n/Qn
   !     CALL RNSET(ISEED)
   if (P1==0) then
   Y1=0
   else if (P1==1) then
```

```
   Y1=Y
   else
   CALL RNBIN(Y,P1,IR)
   Y1=IR(1)
   end if
   P3=P2/(1-P1)
   if (P3==0) then
   Y2=0
   else if (P3==1) then
   Y2=Y-Y1
   else if (Y-Y1  == 0) then
   Y2 = 0
   Y3 = 0
   else
   CALL RNBIN(Y-Y1,P3,IR)
   Y2=IR(1)
   end if
   Y3=Y-Y2-Y1
   RETURN
END SUBROUTINE YGENER

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!Subroutine: Y3GENER
!Description: Generate (y31,y32) from binomial
!             distribution with parameters {Y3;Q31n/Q3n}
!Input: Y = the observed cancer cases at each age period
!       Q31n = the product of N31 and the probability of
!              developing tumor during each age period in
!              people who are normal people at embryo stage
!              and develop tumor through 3-stage pathway
!       Q32n = the product of N32 and the probability of
!              developing tumor during each age period in
!              people who are normal people at embryo stage
!              and develop tumor through 2-stage pathway
!       ISEED = seed for generating random number
!Output: Y31 = the number of cancer cases generated by
   people
!               who are normal people at embryo stage and
!               develop tumor through 3-stage pathway
!        Y32 = the number of cancer cases generated by
   people
!               who are normal people at embryo stage and
!               develop tumor through 2-stage pathway
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
SUBROUTINE Y3GENER(Y3,Q31n,Q32n,ISEED,Y31,Y32)
```

```
  IMPLICIT  NONE
  DOUBLE PRECISION, INTENT(IN) :: Q31n,Q32n
  INTEGER, INTENT(IN) ::ISEED, Y3
  INTEGER, INTENT(OUT) :: Y31,Y32
  REAL P31,Q3n
  INTEGER IR(1)
  Q3n = Q31n +Q32n
  P31=Q31n/Q3n
  !CALL RNSET(ISEED)
  if (P31==0) then
  Y31=0
  else if (P31==1) then
  Y31=Y3
  else if (Y3 == 0) then
  Y31 = 0
  else
  CALL RNBIN(Y3,P31,IR)
  Y31=IR(1)
  end if
  Y32=Y3-Y31
  RETURN
END SUBROUTINE Y3GENER

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

!Subroutine: PICK
!Description: Select the k-th (n1, n2, n3) from 500 samples
!             through the Weighted Bootstrap Method
!Input: Y = the number of cancer cases at each age period
!       Y1 = the number of cancer cases generated by people
!            who have genotype I1 at embryo stage
!       Y2 = the number of cancer cases generated by people
!            who have genotype I2 at embryo stage
!       Y31 = the number of cancer cases generated by people
!             who are normal people at embryo stage and
!             develop tumor through 3-stage pathway
!       Q1 = the product of N1 and the probability of
!             developing tumor during each age period in
!             people who have genotype I1 at embryo stage
!       Q2 = the product of N2 and the probability of
!             developing tumor during each age period in
!             people who have genotype I2 at embryo stage
!       Q31 = the product of N31 and the probability of
!             developing tumor during each age period in
!             people who are normal people at embryo stage
```

```
!              and develop tumor through 3-stage pathway
!Output: K = the selected number from 1 to 500 through
!            weighted bootstrap method
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
SUBROUTINE PICK(Y,Y1,Y2,Y31,Q1,Q2,Q31,Q32,K)
  IMPLICIT  NONE
  DOUBLE PRECISION, INTENT(IN) ::Q1(1:NB),Q2(1:NB),Q31(1:NB)
     ,Q32(1:NB)
  INTEGER, INTENT(IN) :: Y,Y1(1:NB),Y2(1:NB),Y31(1:NB)
  INTEGER, INTENT(OUT) :: K
  DOUBLE PRECISION W(NB),G(NB)
  DOUBLE PRECISION SW, U, SG,LogFactorialY1(1:NB),
     LogFactorialY2(1:NB),LogFactorialY31(1:NB),
     LogFactorialY32(1:NB),w1(1:NB), w2(1:NB), w31(1:NB),
     w32(1:NB),wt(1:NB), wtmean
  REAL R(1)
  INTEGER T,j,Y32(1:NB), l
  Y32=Y-Y1-Y2-Y31
  LogFactorialY1 = 0.0
  LogFactorialY2 = 0.0
  LogFactorialY32 = 0.0
  SW = 0.0
  DO j=1,NB
   W(j)=0.0
  END DO
  DO j=1,NB
  DO l = 1, Y1(j)
    LogFactorialY1(j) = LogFactorialY1(j) + log(REAL(l))
  END DO
  DO l = 1, Y2(j)
    LogFactorialY2(j) = LogFactorialY2(j) + log(REAL(l))
  END DO
  DO l = 1, Y31(j)
    LogFactorialY31(j) = LogFactorialY31(j) + log(REAL(l))
  END DO
  DO l = 1, Y32(j)
    LogFactorialY32(j) = LogFactorialY32(j) + log(REAL(l))
  END DO
  w1(j) = (-Q1(j)+Y1(j)*log(Q1(j))-LogFactorialY1(j))
  w2(j) = (-Q2(j)+Y2(j)*log(Q2(j))-LogFactorialY2(j))
  w31(j) = (-Q31(j)+Y31(j)*log(Q31(j))-LogFactorialY31(j))
  w32(j) = (-Q32(j)+Y32(j)*log(Q32(j))-LogFactorialY32(j))
  wt(j) = w1(j) +w2(j) + w31(j)+ w32(j)
  END DO
  wtmean = sum(wt)/NB
```

```
  DO j = 1,NB
  W(j) = exp(wt(j) - wtmean)
  SW=SW+W(j)
  END DO
  DO j=1,NB
  IF (SW.NE.0) G(j)=W(j)/SW
  END DO
  CALL RNUN(1,R)
  U=R(1)
  SG=0.0
  j=0
  5  j=j+1
  SG=SG+G(j)
  IF(SG.LT.U) THEN
  GOTO 5
  END IF
  K=j
  RETURN
END SUBROUTINE PICK

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

!Function: Fit2
!Description: the function is called in pikaia as a fitness
!             function
!Input: m, X
!Parameter: m = the number of parameters
!           X = an array of parameters with m elements
!Return: Fit2 = the fitness function which is the negative
   of
!               the deviance for the conditional posterior
!               distribution
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

REAL FUNCTION Fit2(m, X)
  IMPLICIT NONE
  INTEGER, INTENT(IN) :: m
  REAL, INTENT(IN) :: X(m)
  INTEGER j, status
  DOUBLE PRECISION:: gamma1, lambda(4), alpha2, p1,p2,alpha1
     ,alpha3,gamma2,gamma3,alpha0,theta1,theta2
  DOUBLE PRECISION P030(1:MaxAge), P130(1:MaxAge), P131(1:
     MaxAge), P231(1:MaxAge),P232(1:MaxAge),P020(1:MaxAge),
     phi2s(1:MaxAge), phi2b(1:MaxAge),phi22s(1:MaxAge),
     phi22b(1:MaxAge), phi3s(1:MaxAge), phi3b(1:MaxAge),
```

```
    phi1s(1:MaxAge), phi1b(1:MaxAge),Q1(1:MaxAge),Q2(1:
    MaxAge),Q3(1:MaxAge),Q31(1:MaxAge), Q32(1:MaxAge),Q(1:
    MaxAge)
DOUBLE PRECISION D0,Dev1,Dev2,Dev3
Dev1 = 0
Dev2 = 0
IF ( m /= n_fit2 ) THEN
Fit2 = 0
RETURN
END IF
gamma1 = DenormalizeX(DBLE(X(1)),gamma1Range)
lambda(1) = DenormalizeX(DBLE(X(2)),lambda1Range)
lambda(2) = DenormalizeX(DBLE(X(3)),lambda2Range)
lambda(3) = DenormalizeX(DBLE(X(4)),lambda3Range)
alpha2 = DenormalizeX(DBLE(X(5)),alpha2Range)
p1 = DenormalizeX(DBLE(X(6)),p1Range)
lambda(4) = DenormalizeX(DBLE(X(7)),lambda4Range)
alpha3 = DenormalizeX(DBLE(X(8)),alpha3Range)
gamma2 = DenormalizeX(DBLE(X(9)),gamma2Range)
gamma3 = DenormalizeX(DBLE(X(10)),gamma3Range)
theta2 = DenormalizeX(DBLE(X(11)),theta2Range)
p2 = DenormalizeX(DBLE(X(12)),p2Range)

DO j = 1, MaxAge
 CALL CalculateQ(alpha2, gamma1, lambda, alpha3,gamma2,
     gamma3,theta2,N1_PICKED(j),N2_PICKED(j), N3_PICKED(j),
      N31_PICKED(j),N32_PICKED(j),j, Q(j),Q1(j),Q2(j),Q3(j)
     , Q31(j),Q32(j),qstatus)
END DO

D0 = ( n0*p2*alpha2- y0) - y0*log((n0*p2*alpha2)/y0)
DO j = 1, MaxAge
IF ((N1_PICKED(j) > 1e-20)) THEN
  Dev1 = Dev1 + N1_PICKED(j)*log(N1_PICKED(j)/N(j))
  Dev1 = Dev1 - N1_PICKED(j)*log(p2)
END IF
IF ((N2_PICKED(j) > 1e-20)) THEN
  Dev1 = Dev1 + N2_PICKED(j)*log(N2_PICKED(j)/N(j))
  Dev1 = Dev1 - N2_PICKED(j)*log(p1)
END IF
IF ((N3_PICKED(j) > 1e-20)) THEN
 Dev1 = Dev1 + N3_PICKED(j)*log(N3_PICKED(j)/N(j))
  Dev1 = Dev1 - N3_PICKED(j)*log(1-p1-p2)
END IF
IF ((N31_PICKED(j) > 1e-20)) THEN
```

```
      Dev3 = Dev3 + N31_PICKED(j)*log(N31_PICKED(j)/N3_PICKED(
         j))
      Dev3 = Dev3 - N31_PICKED(j)*log(1-alpha3)
    END IF
    IF ((N32_PICKED(j) > 1e-20)) THEN
      Dev3 = Dev3 + N32_PICKED(j)*log(N32_PICKED(j)/N3_PICKED(
         j))
      Dev3 = Dev3 - N32_PICKED(j)*log(alpha3)
    END IF
    Dev2 = Dev2 + Q1(j) - Y1_PICKED(j)
    IF ((Q1(j) > 1e-20) .AND. (Y1_PICKED(j) > 1e-20)) THEN
      Dev2 = Dev2 - Y1_PICKED(j)*log(Q1(j)/Y1_PICKED(j))
    END IF
    Dev2 = Dev2 + Q2(j) - Y2_PICKED(j)
    IF ((Q2(j) > 1e-20) .AND. (Y2_PICKED(j) > 1e-20)) THEN
      Dev2 = Dev2 - Y2_PICKED(j)*log(Q2(j)/Y2_PICKED(j))
    END IF
    Dev2 = Dev2 + Q31(j) - Y31_PICKED(j)
    IF ((Q31(j) > 1e-20) .AND. (Y31_PICKED(j) > 1e-20)) THEN
      Dev2 = Dev2 - Y31_PICKED(j)*log(Q31(j)/Y31_PICKED(j))
    END IF
    Dev2 = Dev2 + Q32(j) - Y32_PICKED(j)
    IF ((Q32(j) > 1e-20) .AND. (Y32_PICKED(j) > 1e-20)) THEN
      Dev2 = Dev2 - Y32_PICKED(j)*log(Q32(j)/Y32_PICKED(j))
    END IF
    END DO
    Fit2 = -1.0 * (D0+Dev1+Dev2+Dev3)
END FUNCTION Fit2

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!Subroutine:  CalculateQ
!Description: the function is used to calculate the
!             probability of developing tumor during each
!             age period in people who have different
!             genotype and develop tumor through different
!             pathways
!Input: alpha2,gamma1,lambda,alpha3,gamma2,gamma3,theta2
!       are parameters in the model
!       N1 = the number of I1 people at embryo stage
!       N2 = the number of I2 people at embryo stage
!       N3 = the number of normal people at embryo stage
!       N31 = the number of normal people at risk who
   develop
!             tumor through 3-stage pathway
```

```
!       N32 = the number of normal people at risk who
   develop
!             tumor through 2-stage pathway
!       i = the age period
!Output: Q = Q1+Q2+Q3
!        Q1 = the product of N1 and the probability of
!             developing tumor during each age period in
!             people who have genotype I1 at embryo stage
!        Q2 = the product of N2 and the probability of
!             developing tumor during each age period in
!             people who have genotype I2 at embryo stage
!        Q3 = the product of N3 and the probability of
!             developing tumor during each age period in
!             people who are normal people at embryo stage
!        Q31 = the product of N31 and the probability of
!             developing tumor during each age period in
!             people who are normal people at embryo stage
!             and develop tumor through 3-stage pathway
!        Q32 = the product of N32 and the probability of
!             developing tumor during each age period in
!             people who are normal people at embryo stage
!             and develop tumor through 2-stage pathway
!        status return 0 if correct,
!               return 1 if Q1<0 or Q2<0 or Q31 < 0
!                        or Q32 < 0
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
SUBROUTINE CalculateQ(alpha2,gamma1,lambda,alpha3,gamma2,
   gamma3,theta2,N1, N2, N3,N31,N32,i,Q,Q1,Q2,Q3,Q31,Q32,
   status)
IMPLICIT NONE
DOUBLE PRECISION, INTENT(IN) :: alpha2,alpha3, gamma1,gamma2
   ,gamma3,theta2, lambda(4), N1B,N2B, N3B, N31B, N32B
INTEGER, INTENT(IN) :: i
DOUBLE PRECISION, INTENT(OUT) :: Q,Q1,Q2,Q3,Q31,Q32
INTEGER, INTENT(OUT) :: status
DOUBLE PRECISION P030, P130,P31,P131, P231,P232,P020, phi2s
   , phi2b,phi22s,phi22b,phi3b, phi3s, phi1s,phi1b
INTEGER j, temp
status = 0
phi1s = ((1+gamma2)**(dt*(i-1))-1)
phi1b = ((1+gamma2)**(dt*i-1)-1)
phi2s = gamma1*gamma2*(((1+gamma1)**(dt*(i-1))-1)/(gamma1-
   gamma2)/gamma1 +((1+gamma2)**(dt*(i-1))-1)/(gamma2-gamma1
   )/gamma2)
phi2b = gamma1*gamma2*(((1+gamma1)**(dt*i-1)-1)/(gamma1-
```

```
    gamma2)/gamma1+((1+gamma2)**(dt*i-1)-1)/(gamma2-gamma1)/
    gamma2)
phi3s = gamma1*gamma2*(((1+gamma1)**(dt*(i-1))-1-gamma1*dt*(
    i-1))/(gamma1-gamma2)/gamma1/gamma1+((1+gamma2)**(dt*(i
    -1))-1-gamma2*dt*(i-1))/(gamma2-gamma1)/gamma2/gamma2)
phi3b = gamma1*gamma2*(((1+gamma1)**(dt*i-1)-1-gamma1*(dt*i
    -1))/(gamma1-gamma2)/gamma1/gamma1+((1+gamma2)**(dt*i-1)
    -1-gamma2*(dt*i-1))/(gamma2-gamma1)/gamma2/gamma2)
phi22s = ((1+gamma3)**(dt*(i-1))-1-gamma3*(dt*(i-1)))
phi22b = ((1+gamma3)**(dt*i-1)-1-gamma3*(dt*i-1))

P030 = exp(-lambda(3)*phi3s) - exp(-lambda(3)*phi3b)
P232 = exp(-lambda(1)*phi1s) - exp(-lambda(1)*phi1b)
P020 = exp(-lambda(4)*phi22s) - exp(-lambda(4)*phi22b)

Q1 = N1B*P232*(1-alpha2)
Q2 = N2B*(exp(-theta2*phi1s-lambda(2)*phi2s)-exp(-theta2*
    phi1b-lambda(2)*phi2b))
Q31=N31B*P030
Q32=N32B*P020
Q3=Q31+Q32
IF(Q1 < 0 .OR.Q2 < 0 .OR. Q31 < 0.OR. Q32 < 0) THEN
status = 1
END IF
Q = Q1+ Q2 + Q3
END SUBROUTINE CalculateQ

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!Function: DenormalizeX
!Description: Scale parameter value from (0,1) range to
!             actual range
!Parameter: x = the parameter in the model
!           xrange = the range of parameter in the model
!Return: DenormalizeX
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
DOUBLE PRECISION FUNCTION DenormalizeX(x,xrange)
IMPLICIT NONE
DOUBLE PRECISION, INTENT(IN) :: x,xrange(1:3)
DenormalizeX = x*(xrange(2)-xrange(1))+xrange(1)
END FUNCTION DenormalizeX

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!Subroutine: Q_from_pikaia
!Description: to obtain the predicted cancer cases
!Input: m = the number of parameters
```

```
!        X = an array of parameters with m elements
!Output: Q = the predicted cancer cases at each age period
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

SUBROUTINE Q_from_pikaia(X, m, N, Q)
IMPLICIT NONE
INTEGER, INTENT(IN) :: m,N(1:MaxAge)
DOUBLE PRECISION, INTENT(IN) :: X(m)
DOUBLE PRECISION, INTENT(OUT) :: Q(1:MaxAge)
DOUBLE PRECISION p,p1,p2, gamma1, lambda(4), alpha2,gamma2,
   alpha1,alpha3,alpha0,gamma3,theta1,theta2
DOUBLE PRECISION Q1(1:MaxAge), Q2(1:MaxAge),Q3(1:MaxAge),
   Q31(1:MaxAge),Q32(1:MaxAge),N1_pred(1:MaxAge), N2_pred
   (1:MaxAge), N3_pred(1:MaxAge),N31_pred(1:MaxAge),N32_pred
   (1:MaxAge)
INTEGER status
gamma1 = DenormalizeX(DBLE(X(1)),gamma1Range)
lambda(1) = DenormalizeX(DBLE(X(2)),lambda1Range)
lambda(2) = DenormalizeX(DBLE(X(3)),lambda2Range)
lambda(3) = DenormalizeX(DBLE(X(4)),lambda3Range)
alpha2 = DenormalizeX(DBLE(X(5)),alpha2Range)
p1 = DenormalizeX(DBLE(X(6)),p1Range)
lambda(4) = DenormalizeX(DBLE(X(7)),lambda4Range)
alpha3 = DenormalizeX(DBLE(X(8)),alpha3Range)
gamma2 = DenormalizeX(DBLE(X(9)),gamma2Range)
gamma3 = DenormalizeX(DBLE(X(10)),gamma3Range)
theta2 = DenormalizeX(DBLE(X(11)),theta2Range)
p2 = DenormalizeX(DBLE(X(12)),p2Range)
 DO j = 1, MaxAge
 N1_pred(j)=N(j)*p2
 N2_pred(j)=N(j)*p1
 N3_pred(j)=N(j)-N1_pred(j)-N2_pred(j)
 N31_pred(j)=N3_pred(j)*(1-alpha3)
 N32_pred(j)=N3_pred(j)- N31_pred(j)
 CALL CalculateQ(alpha2, gamma1, lambda, alpha3,gamma2,
    gamma3,theta2,N1_pred(j),N2_pred(j), N3_pred(j),
    N31_pred(j), N32_pred(j),j, Q(j),Q1(j),Q2(j),Q3(j),Q31(j
    ),Q32(j), qstatus)
 END DO
END SUBROUTINE Q_from_pikaia
END PROGRAM WimlsTumor
```

APPENDIX C

PROGRAM CODE TO FIT A MODEL OF RENAL CELL CARCINOMA

```
PROGRAM RenalCarcinomas
USE RNNOR_INT
USE RNBIN_INT
USE UMACH_INT
USE RNSET_INT
USE FAC_INT
IMPLICIT NONE
INTEGER, PARAMETER:: NB=500, n_fit2 = 15 , MaxAge = 84, dt=4
INTEGER SEED(NB),ISEED,K,control,Y(MaxAge),Y1(1:NB),Y2(1:NB)
    ,Y3(1:NB),N(MaxAge),N1(MaxAge),N2(MaxAge),N3(MaxAge),i,j,
    pikaia_status, n_loop,N1Temp,N2Temp,N3Temp,kk,Y_Orig(
    MaxAge), N_Orig(MaxAge),qstatus,T
DOUBLE PRECISION Q1(1:NB),Q2(1:NB), Q3(1:NB), P22(MaxAge),
    P33(MaxAge), Q(1:NB), Q_from_pikaia_output(1:MaxAge),
    N1_pred(MaxAge), N2_pred(MaxAge),N3_pred(MaxAge),
    myalpha1,myalpha2,kidney_03_mylambda,kidney_03_mygamma
    (1:2), kidney_mygamma(1:3),kidney_gamma3_delta_s,
    kidney_gamma3_delta_b, kidney_mylambda,kidney_mydelta,
    kidney_05_mygamma(1:4),kidney_05_mylambda,fit2_result,
    sum_of_H,chisq
DOUBLE PRECISION,DIMENSION(1:NB,1:MaxAge)::H, N1B, N2B, N3B
DOUBLE PRECISION,DIMENSION(1:MaxAge)::N1_PICKED,N2_PICKED,
    N3_PICKED, K_PICKED,Y1_PICKED, Y2_PICKED, Y3_PICKED
DOUBLE PRECISION,DIMENSION(1:3):: alpha1Range,alpha2Range,
    kidney_03_gamma1Range, kidney_03_gamma2Range,
    kidney_03_lambdaRange,kidney_gamma1Range,
    kidney_gamma2Range, kidney_gamma3Range,kidney_lambdaRange
    ,kidney_deltaRange,kidney_05_gamma1Range,
    kidney_05_gamma2Range,kidney_05_gamma3Range,
    kidney_05_gamma4Range, kidney_05_lambdaRange,
    kidney_05_deltaRange
DOUBLE PRECISION, PARAMETER :: min_real = 1e-37
DOUBLE PRECISION,DIMENSION(1:15) :: old_para
REAL ctrl(12), pikaia_fit2_f, pikaia_fit2_x(n_fit2), fit2out
REAL normalized_para(1:15)
!!! Read data
INCLUDE 'RenalCarcinomas_3_4_5_stage_ICCC.FI'
DO i = 1, MaxAge
Y(i) = Y_Orig(i)
N(i) = N_Orig(i)
END DO
```

```
open (unit=5,file='output.dat',status='unknown')
iseed=12345
CALL RNSET(iseed)
CALL rninit(iseed)

!!! Main loop
DO n_loop= 1, 50
DO  i = 1, MaxAge
Do kk=1, NB
iseed=iseed+1
CALL NGENERNOR_0(N(i),myalpha1,myalpha2,ISEED,N1Temp,N2Temp,
    N3Temp)
N1(i) = N1Temp
N2(i) = N2Temp
N3(i) = N3Temp
N1B(kk,i)=N1Temp
N2B(kk,i)=N2Temp
N3B(kk,i)=N3Temp
END DO
END DO

DO i = 1, MaxAge
IF (Y(i)==0) THEN
K_PICKED(i) = 1
Y1_PICKED(i) = 0
Y2_PICKED(i) = 0
Y3_PICKED(i) = 0
N1_PICKED(i) = N1B(K_PICKED(i),i)
N2_PICKED(i) = N2B(K_PICKED(i),i)
N3_PICKED(i) = N3B(K_PICKED(i),i)
CYCLE
END IF

DO j = 1, NB
CALL CalculateQ (myalpha1,myalpha2,kidney_03_mygamma,
    kidney_03_mylambda,kidney_mygamma, kidney_mylambda,
    kidney_mydelta,kidney_05_mygamma,kidney_05_mylambda,N1B(j
    ,i),N2B(j,i),N3B(j,i),i,Q(j),Q1(j),Q2(j),Q3(j),qstatus)
END DO

IF(qstatus .EQ. 1) THEN
 myalpha1 = old_para(1)
 myalpha2=old_para(2)
 kidney_03_mygamma(1) =old_para(3)
 kidney_03_mygamma(2)=old_para(4)
```

```
 kidney_03_mylambda=old_para(5)
 kidney_mygamma(1) =  old_para(6)
 kidney_mygamma(2)  = old_para(7)
 kidney_mygamma(3)  = old_para(8)
 kidney_mylambda  =  old_para(9)
 kidney_mydelta  =  old_para(10)
 kidney_05_mygamma(1) =old_para(11)
 kidney_05_mygamma(2) =  old_para(12)
 kidney_05_mygamma(3) = old_para(13)
 kidney_05_mygamma(4) =  old_para(14)
 kidney_05_mylambda  = old_para(15)
END IF

DO j = 1, NB
CALL YGENER(Y(i),Q1(j),Q2(j),Q3(j),ISEED,Y1(j),Y2(j),Y3(j))
END DO

CALL PICK(Y(i),Y1,Y2,Q1, Q2,Q3,K)
K_PICKED(i) = K

IF(K_PICKED(i) == 0)
K_PICKED(i) = 1

Y1_PICKED(i) = Y1(K)
Y2_PICKED(i) = Y2(K)
Y3_PICKED(i) = Y3(K)
N1_PICKED(i) = N1B(K_PICKED(i),i)
N2_PICKED(i) = N2B(K_PICKED(i),i)
N3_PICKED(i) = N3B(K_PICKED(i),i)
END DO

do 10 i=1,12
ctrl(i) = -1
10 continue
ctrl(1)=100
ctrl(2)=10000
ctrl(6)=0.005
ctrl(7)=0.0005
ctrl(8)=0.2

old_para(1) = myalpha1
old_para(2) = myalpha2
old_para(3) = kidney_03_mygamma(1)
old_para(4) = kidney_03_mygamma(2)
old_para(5) = kidney_03_mylambda
```

```
old_para(6) = kidney_mygamma(1)
old_para(7) = kidney_mygamma(2)
old_para(8) = kidney_mygamma(3)
old_para(9) = kidney_mylambda
old_para(10) = kidney_mydelta
old_para(11) = kidney_05_mygamma(1)
old_para(12) = kidney_05_mygamma(2)
old_para(13) = kidney_05_mygamma(3)
old_para(14) = kidney_05_mygamma(4)
old_para(15) = kidney_05_mylambda

CALL pikaia(Fit2,n_fit2,ctrl,pikaia_fit2_x,pikaia_fit2_f,
    pikaia_status)
CALL Q_from_pikaia(DBLE(pikaia_fit2_x),n_fit2,N,
    Q_from_pikaia_output)

myalpha1 = DenormalizeX(DBLE(pikaia_fit2_x(1)),alpha1Range)
myalpha2 = DenormalizeX(DBLE(pikaia_fit2_x(2)),alpha2Range)
kidney_03_mygamma(1) = DenormalizeX(DBLE(pikaia_fit2_x(3)),
    kidney_03_gamma1Range)
kidney_03_mygamma(2) = DenormalizeX(DBLE(pikaia_fit2_x(4)),
    kidney_03_gamma2Range)
kidney_03_mylambda = DenormalizeX(DBLE(pikaia_fit2_x(5)),
    kidney_03_lambdaRange)
kidney_mygamma(1) = DenormalizeX(DBLE(pikaia_fit2_x(6)),
    kidney_gamma1Range)
kidney_mygamma(2) = DenormalizeX(DBLE(pikaia_fit2_x(7)),
    kidney_gamma2Range)
kidney_mygamma(3) = DenormalizeX(DBLE(pikaia_fit2_x(8)),
    kidney_gamma3Range)
kidney_mylambda = DenormalizeX(DBLE(pikaia_fit2_x(9)),
    kidney_lambdaRange)
kidney_mydelta = DenormalizeX(DBLE(pikaia_fit2_x(10)),
    kidney_deltaRange)
kidney_05_mygamma(1) = DenormalizeX(DBLE(pikaia_fit2_x(11)),
    kidney_05_gamma1Range)
kidney_05_mygamma(2) = DenormalizeX(DBLE(pikaia_fit2_x(12)),
    kidney_05_gamma2Range)
kidney_05_mygamma(3) = DenormalizeX(DBLE(pikaia_fit2_x(13)),
    kidney_05_gamma3Range)
kidney_05_mygamma(4) = DenormalizeX(DBLE(pikaia_fit2_x(14)),
    kidney_05_gamma4Range)
kidney_05_mylambda = DenormalizeX(DBLE(pikaia_fit2_x(15)),
    kidney_05_lambdaRange)
```

```
do i = 1,MaxAge
write (5,*) Q_from_pikaia_output(i)
end do
write (5,'(A,I5,$)') ' n_loop ',n_loop
write (5,'(A,ES14.6,$)') ' Fit2 ',pikaia_fit2_f
write (5,'(A,ES14.6,$)') ' alpha1 ',myalpha1
write (5,'(A,ES14.6,$)') ' alpha2 ',myalpha2
write (5,'(A,ES14.6,$)')  ' kidney_03_gamma1 ',
   kidney_03_mygamma(1)
write (5,'(A,ES14.6,$)')  ' kidney_03_gamma1 ',
   kidney_03_mygamma(2)
write (5,'(A,ES14.6,$)')  ' kidney_03_lambda1 ',
   kidney_03_mylambda
write (5,'(A,ES14.6,$)')  ' kidney_mygamma1 ',kidney_mygamma
   (1)
write (5,'(A,ES14.6,$)')  ' kidney_mygamma2 ',kidney_mygamma
   (2)
write (5,'(A,ES14.6,$)')  ' kidney_mygamma3 ',kidney_mygamma
   (3)
write (5,'(A,ES14.6,$)') ' kidney_mydelta ',kidney_mydelta
write (5,'(A,ES14.6,$)')  ' kidney_mylambda ',
   kidney_mylambda
write (5,'(A,ES14.6,$)')  ' kidney_05_mygamma1 ',
   kidney_05_mygamma(1)
write (5,'(A,ES14.6,$)')  ' kidney_05_mygamma2 ',
   kidney_05_mygamma(2)
write (5,'(A,ES14.6,$)')  ' kidney_05_mygamma3 ',
   kidney_05_mygamma(3)
write (5,'(A,ES14.6,$)')  ' kidney_05_mygamma4 ',
   kidney_05_mygamma(4)
write (5,'(A,ES14.6,$)')  ' kidney_05_mylambda ',
   kidney_05_mylambda
write (5,*) 'End'
write (*,*) 'Loop'
END DO
!!! End Main loop

CONTAINS
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

!Subroutine: NGENERNOR_0
!Description: Generate (n1,n2,n3) from multinomial
!             distribution. Since nj is very large and p is
!             very small, normal approximation is applied.
!Input: N = population at each age period
```

```
!       P1 = the proportion for 3-stage pathway in
   population
!       P2 = the proportion for 4-stage pathway in
   population
!       ISEED = seed for generating random number
!Output: N1 = the number of people at risk who develop
   cancer
!             through 3-stage pathway
!        N2 = the number of people at risk who develop
   cancer
!             through 4-stage pathway
!        N3 = the number of people at risk who develop
   cancer
!             through 5-stage pathway
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

SUBROUTINE NGENERNOR_0(N,P1,P2,ISEED,N1,N2,N3)
IMPLICIT  NONE
DOUBLE PRECISION, INTENT(IN) :: P1,P2
INTEGER, INTENT(IN) ::ISEED,N
INTEGER, INTENT(OUT) :: N1,N2,N3
REAL NOR1,NOR2
REAL IR(1)
DO
CALL RNNOR(IR)
NOR1=IR(1)
CALL RNNOR(IR)
NOR2=IR(1)
IF (NOR1 < 3 .AND. NOR1 > -3) EXIT
IF (NOR2 < 3 .AND. NOR2 > -3) EXIT
END DO
N1=NOR1 * sqrt(N*p1*(1-p1))+ N*p1
N2=NOR2 * sqrt((N-N1)*p2*(1-P1-P2)/(1-P1)/(1-P1))+ (N-N1)*p2
   /(1-p1)
N3=N-N1-N2
IF(N1 < 0) THEN
N1 = 1e-7
END IF
RETURN
END SUBROUTINE NGENERNOR_0

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!Subroutine: YGENER
!Description: Generate (y1, y2, y3) from multinomial with
!             parameters {Y; p1, p2}
```

```
!Input: Y = the observed cancer cases at each age period
!       Q1n = the product of N1 and the probability of
!             developing tumor during each age period in
!             people who develop kidney cancer through
!             3-stage pathway
!       Q2n = the product of N2 and the probability of
!             developing tumor during each age period in
!             people who develop kidney cancer through
!             4-stage pathway
!       Q3n = the product of N3 and the probability of
!             developing tumor during each age period in
!             people who develop kidney cancer through
!             5-stage pathway
!       ISEED = seed for generating random number
!Output: Y1 = number of cancer cases generated by people
!             developing cancer through 3-stage pathway
!        Y2 = number of cancer cases generated by people
!             developing cancer through 4-stage pathway
!        Y3 = number of cancer cases generated by people
!             developing cancer through 5-stage pathway
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
SUBROUTINE YGENER(Y,Q1n,Q2n,Q3n,ISEED,Y1,Y2,Y3)
IMPLICIT  NONE
DOUBLE PRECISION, INTENT(IN) :: Q1n,Q2n,Q3n
INTEGER, INTENT(IN) ::ISEED, Y
INTEGER, INTENT(OUT) :: Y1,Y2,Y3
REAL P1,P2,P3,Qn
INTEGER IR(1)
Qn = Q1n+Q2n+Q3n
P1=Q1n/Qn
P2=Q2n/Qn
if (P1==0) then
Y1=0
else if (P1==1) then
Y1=Y
else
CALL RNBIN(Y,P1,IR)
Y1=IR(1)
end if
P3=P2/(1-P1)
if (P3==0) then
Y2=0
else if (P3==1) then
Y2=Y-Y1
else if (Y-Y1==0) then
```

```
Y2 = 0
Y3 = 0
else
CALL RNBIN(Y-Y1,P3,IR)
Y2=IR(1)
end if
Y3=Y-Y2-Y1
RETURN
END SUBROUTINE YGENER

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

!Subroutine: PICK
!Description: Select the k-th (n1, n2, n3) from 500 samples
!             through the Weighted Bootstrap Method
!Input: Y = the number of cancer cases at each age period
!       Y1 = the number of cancer cases generated by people
!            who develop cancer by 3-stage pathway
!       Y2 = the number of cancer cases generated by people
!            who develop cancer by 4-stage pathway
!       Q1 = the product of N1 and the probability of
!            developing tumor during each age period in
!            people who develop cancer by 3-stage pathway
!       Q2 = the product of N2 and the probability of
!            developing tumor during each age period in
!            people who develop cancer by 4-stage pathway
!       Q3 = the product of N3 and the probability of
!            developing tumor during each age period in
!            people who develop cancer by 5-stage pathway
!Output: K = the selected number from 1 to 500 through
!            weighted bootstrap method
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
SUBROUTINE PICK(Y,Y1,Y2,Q1,Q2,Q3,K)
IMPLICIT NONE
DOUBLE PRECISION, INTENT(IN) ::Q1(1:NB),Q2(1:NB),Q3(1:NB)
INTEGER, INTENT(IN) :: Y,Y1(1:NB),Y2(1:NB)
INTEGER, INTENT(OUT) :: K
DOUBLE PRECISION W(NB),G(NB),SW,U,SG,LogFactorialY1(1:NB),
   LogFactorialY2(1:NB),LogFactorialY3(1:NB),w1(1:NB),w2(1:
   NB),w3(1:NB),wt(1:NB),wtmean
REAL R(1)
INTEGER T,j,Y3(1:NB), l
Y3=Y-Y1-Y2
LogFactorialY1 = 0.0
LogFactorialY2 = 0.0
```

```
LogFactorialY3 = 0.0
SW = 0.0
DO j=1,NB
 W(j)=0.0
END DO
DO j=1,NB
DO l = 1, Y1(j)
LogFactorialY1(j) = LogFactorialY1(j) + log(REAL(l))
END DO
DO l = 1, Y2(j)
LogFactorialY2(j) = LogFactorialY2(j) + log(REAL(l))
END DO
DO l = 1, Y3(j)
LogFactorialY3(j) = LogFactorialY3(j) + log(REAL(l))
END DO
w1(j) = (-Q1(j)+Y1(j)*log(Q1(j))-LogFactorialY1(j))
w2(j) = (-Q2(j)+Y2(j)*log(Q2(j))-LogFactorialY2(j))
w3(j) = (-Q3(j)+Y3(j)*log(Q3(j))-LogFactorialY3(j))
wt(j) = w1(j) +w2(j) + w3(j)
END DO
wtmean = sum(wt)/NB
DO j = 1,NB
W(j) = exp(wt(j) - wtmean)
SW = SW + W(j)
END DO
DO j=1,NB
IF (SW.NE.0) G(j)=W(j)/SW
END DO
CALL RNUN(1,R)
U=R(1)
SG=0.0
j=0
5 j=j+1
SG = SG+G(j)
IF(SG.LT.U) THEN
GOTO 5
END IF
K=j
RETURN
END SUBROUTINE PICK

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

!Function: Fit2
!Description: the function is called in pikaia as fitness
```

```
!              function
!Input: m, X
!Parameter: m = the number of parameters
!           X = an array of parameters with m elements the
!Return: Fit2= the fitness function which is the negative of
!              deviance for the conditional posterior
!              distribution
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

REAL FUNCTION Fit2(m, X)
IMPLICIT NONE
INTEGER, INTENT(IN) :: m
REAL, INTENT(IN) :: X(m)
INTEGER j, status
DOUBLE PRECISION:: kidney_03_gamma(2), kidney_03_lambda,
   alpha2, alpha1, kidney_gamma(3),kidney_lambda,
   kidney_delta, kidney_05_gamma(4), kidney_05_lambda
DOUBLE PRECISION P22(1:MaxAge), P33(1:MaxAge), phi22s(1:
   MaxAge), phi22b(1:MaxAge), phi33s(1:MaxAge), phi33b(1:
   MaxAge), Q2(1:MaxAge),Q3(1:MaxAge), Q(1:MaxAge)
DOUBLE PRECISION phi04s(1:MaxAge), phi04b(1:MaxAge)
DOUBLE PRECISION phi03s(1:MaxAge), phi03b(1:MaxAge)
DOUBLE PRECISION phi05s(1:MaxAge), phi05b(1:MaxAge)
DOUBLE PRECISION Dev1,Dev2
Dev1 = 0
Dev2 = 0
IF ( m /= n_fit2 ) THEN
Fit2 = 0
RETURN
END IF
alpha1 = DenormalizeX(DBLE(X(1)),alpha1Range)
alpha2 = DenormalizeX(DBLE(X(2)),alpha2Range)
kidney_03_gamma(1) = DenormalizeX(DBLE(X(3)),
   kidney_03_gamma1Range)
kidney_03_gamma(2) = DenormalizeX(DBLE(X(4)),
   kidney_03_gamma2Range)
kidney_03_lambda = DenormalizeX(DBLE(X(5)),
   kidney_03_lambdaRange)
kidney_gamma(1) = DenormalizeX(DBLE(X(6)),kidney_gamma1Range
   )
kidney_gamma(2) = DenormalizeX(DBLE(X(7)),kidney_gamma2Range
   )
kidney_gamma(3) = DenormalizeX(DBLE(X(8)),kidney_gamma3Range
   )
kidney_lambda = DenormalizeX(DBLE(X(9)),kidney_lambdaRange)
```

```
kidney_delta = DenormalizeX(DBLE(X(10)),kidney_deltaRange)
kidney_05_gamma(1) = DenormalizeX(DBLE(X(11)),
   kidney_05_gamma1Range)
kidney_05_gamma(2) = DenormalizeX(DBLE(X(12)),
   kidney_05_gamma2Range)
kidney_05_gamma(3) = DenormalizeX(DBLE(X(13)),
   kidney_05_gamma3Range)
kidney_05_gamma(4) = DenormalizeX(DBLE(X(14)),
   kidney_05_gamma4Range)
kidney_05_lambda = DenormalizeX(DBLE(X(15)),
   kidney_05_lambdaRange)
DO j = 1, MaxAge
CALL CalculateQ(alpha1,alpha2, kidney_03_gamma,
   kidney_03_lambda,kidney_gamma, kidney_lambda,
   kidney_delta,kidney_05_gamma, kidney_05_lambda, N1_PICKED
   (j), N2_PICKED(j), N3_PICKED(j), j, Q(j), Q1(j), Q2(j),Q3
   (j),status)
END DO
DO j = 1, MaxAge
IF ((N1_PICKED(j) > 1e-20)) THEN
Dev1 = Dev1 + N1_PICKED(j)*log(N1_PICKED(j)/N(j))
Dev1 = Dev1 - N1_PICKED(j)*log(alpha1)
END IF
IF ((N2_PICKED(j) > 1e-20)) THEN
Dev1 = Dev1 + N2_PICKED(j)*log(N2_PICKED(j)/N(j))
Dev1 = Dev1 - N2_PICKED(j)*log(alpha2)
END IF
IF ((N3_PICKED(j) > 1e-20)) THEN
Dev1 = Dev1 + N3_PICKED(j)*log(N3_PICKED(j)/N(j))
Dev1 = Dev1 - N3_PICKED(j)*log((1-alpha1-alpha2))
END IF
Dev2 = Dev2 + Q1(j) - Y1_PICKED(j)
IF ((Q1(j) > 1e-20) .AND. (Y1_PICKED(j) > 1e-20)) THEN
Dev2 = Dev2 - Y1_PICKED(j)*log(Q1(j)/Y1_PICKED(j))
END IF
Dev2 = Dev2 + Q2(j) - Y2_PICKED(j)
IF ((Q2(j) > 1e-20) .AND. (Y2_PICKED(j) > 1e-20)) THEN
Dev2 = Dev2 - Y2_PICKED(j)*log(Q2(j)/Y2_PICKED(j))
END IF
Dev2 = Dev2 + Q3(j) - Y3_PICKED(j)
IF ((Q3(j) > 1e-20) .AND. (Y3_PICKED(j) > 1e-20)) THEN
Dev2 = Dev2 - Y3_PICKED(j)*log(Q3(j)/Y3_PICKED(j))
END IF
END DO
Fit2 = -1.0 *(Dev1+Dev2)
```

```
END FUNCTION Fit2

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!Subroutine:  CalculateQ
!Description: the function is used to calculate the
!             probability of developing tumor during each
!             age period in people who are normal at birth
!             and develop kidney cancer through 3-stage,
!             4-stage or 5-stage pathway.
!Input: alpha1, alpha2, kidney_03_gamma, kidney_03_lambda,
!       kidney_gamma, kidney_lambda,kidney_delta,
!       kidney_05_gamma,kidney_05_lambda are parameters in
!       the model
!       N1 = the number of people at risk who develop kidney
!            cancer through 3-stage pathway
!       N2 = the number of people at risk who develop kidney
!            cancer through 4-stage pathway
!       N3 = the number of people at risk who develop kidney
!            cancer through 5-stage pathway
!        i = the age period
!Output: Q = Q1+Q2+Q3
!        Q1 = the product of N1 and the probability of
!             developing tumor during each age period in
!             people who develop kidney cancer through
!             3-stage pathway.
!        Q2 = the product of N2B and the probability of
!             developing tumor during each age period in
!             people who develop kidney cancer through
!             4-stage pathway.
!        Q3 = the product of N3B and the probability of
!             developing tumor during each age period in
!             people who develop kidney cancer through
!             5-stage pathway.
!        status return 0 if correct,
!               return 1 if Q1<0 or Q2<0 or Q3<0
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
SUBROUTINE CalculateQ(alpha1,alpha2,kidney_03_gamma,
   kidney_03_lambda,kidney_gamma,kidney_lambda,kidney_delta,
   kidney_05_gamma,kidney_05_lambda,N1,N2,N3,i,Q,Q1,Q2,Q3,
   status)
IMPLICIT NONE
DOUBLE PRECISION, INTENT(IN) :: alpha1,alpha2,
   kidney_03_gamma(2),kidney_03_lambda, kidney_gamma(3),
   kidney_lambda,kidney_delta,kidney_05_gamma(4),
   kidney_05_lambda,N1,N2,N3
```

```
INTEGER, INTENT(IN) :: i
DOUBLE PRECISION, INTENT(OUT) :: Q,Q1,Q2,Q3
INTEGER, INTENT(OUT) :: status
DOUBLE PRECISION phi3s,phi3b,P03,phi05s,phi05b,P05,phi04s,
   phi04b,P04, kidney_gamma3_delta_s,kidney_gamma3_delta_b
INTEGER j,temp
status = 0
phi3s = ((1+kidney_03_gamma(1))**(dt*(i-1))-1-
   kidney_03_gamma(1)*dt*(i-1))/(kidney_03_gamma(1)-
   kidney_03_gamma(2))/kidney_03_gamma(1)/kidney_03_gamma(1)
phi3s = phi3s +((1+kidney_03_gamma(2))**(dt*(i-1))-1-
   kidney_03_gamma(2)*dt*(i-1)) /(kidney_03_gamma(2) -
   kidney_03_gamma(1))/kidney_03_gamma(2)/kidney_03_gamma(2)
phi3s = phi3s * kidney_03_gamma(1) * kidney_03_gamma(2)
phi3b = kidney_03_gamma(1) * kidney_03_gamma(2)*(((1+
   kidney_03_gamma(1))**(dt*i-1)-1-kidney_03_gamma(1) *(dt*i
   -1))/(kidney_03_gamma(1)-kidney_03_gamma(2))/
   kidney_03_gamma(1)/ kidney_03_gamma(1)+((1+
   kidney_03_gamma(2))**(dt*i-1)-1-kidney_03_gamma(2)*(dt*i
   -1))/ (kidney_03_gamma(2)-kidney_03_gamma(1))/
   kidney_03_gamma(2)/kidney_03_gamma(2))
if (phi3b< 0) then
phi3b = 1e-20
end if
P03=exp(-kidney_03_lambda*phi3s)-exp(-kidney_03_lambda*phi3b
   )
kidney_gamma3_delta_s = kidney_gamma(3)*exp(-kidney_delta*dt
   *(i-1))
phi04s = ((1+kidney_gamma(1))**(dt*(i-1))-1-dt*(i-1)*
   kidney_gamma(1))/(kidney_gamma(1)**2)/(kidney_gamma(1)-
   kidney_gamma3_delta_s)/(kidney_gamma(1)-kidney_gamma(2))
phi04s = phi04s+((1+kidney_gamma(2))**(dt*(i-1))-1-dt*(i-1)*
   kidney_gamma(2))/(kidney_gamma(2)**2)/(kidney_gamma(2)-
   kidney_gamma(1))/( kidney_gamma(2)-kidney_gamma3_delta_s)
phi04s = phi04s+((1+kidney_gamma3_delta_s)**(dt*(i-1))-1-dt
   *(i-1)*kidney_gamma3_delta_s)/ (kidney_gamma3_delta_s**2)
   /(kidney_gamma3_delta_s-kidney_gamma(1))/(
   kidney_gamma3_delta_s-kidney_gamma(2))
phi04s = kidney_gamma(1)*kidney_gamma(1)*kidney_gamma(2)*
   kidney_gamma(2)*phi04s
kidney_gamma3_delta_b = kidney_gamma(3)*exp(-kidney_delta*(
   dt*i-1))
phi04b = ((1+kidney_gamma(1))**(dt*i-1)-1-(dt*i-1)*
   kidney_gamma(1))/(kidney_gamma(1)**2)/ (kidney_gamma(1)-
   kidney_gamma3_delta_b)/(kidney_gamma(1)-kidney_gamma(2))
```

```
phi04b = phi04b+((1+kidney_gamma(2))**(dt*i-1)-1-(dt*i-1)*
    kidney_gamma(2))/(kidney_gamma(2)**2)/(kidney_gamma(2)-
    kidney_gamma(1))/(kidney_gamma(2)-kidney_gamma3_delta_b)
phi04b = phi04b+((1+kidney_gamma3_delta_b)**(dt*i-1)-1-(dt*i
    -1)*kidney_gamma3_delta_b)/(kidney_gamma3_delta_b**2)/(
    kidney_gamma3_delta_b-kidney_gamma(1))/(
    kidney_gamma3_delta_b-kidney_gamma(2))
phi04b = kidney_gamma(1)*kidney_gamma(1)*kidney_gamma(2)*
    kidney_gamma(2)*phi04b
P04 = exp(-kidney_lambda*phi04s)-exp(-kidney_lambda*phi04b)
phi05s = ((1+kidney_05_gamma(1))**(dt*(i-1))-1-dt*(i-1)*
    kidney_05_gamma(1))/ (kidney_05_gamma(1)**2)/(
    kidney_05_gamma(1)-kidney_05_gamma(2))/(kidney_05_gamma
    (1)-kidney_05_gamma(3))/(kidney_05_gamma(1)-
    kidney_05_gamma(4))
phi05s = phi05s +((1+kidney_05_gamma(2))**(dt*(i-1))-1-dt*(i
    -1)*kidney_05_gamma(2))/ (kidney_05_gamma(2)**2)/(
    kidney_05_gamma(2)-kidney_05_gamma(1))/(kidney_05_gamma
    (2)-kidney_05_gamma(3))/(kidney_05_gamma(2)-
    kidney_05_gamma(4))
phi05s = phi05s +((1+kidney_05_gamma(3))**(dt*(i-1))-1-dt*(i
    -1)*kidney_05_gamma(3))/ (kidney_05_gamma(3)**2)/(
    kidney_05_gamma(3)-kidney_05_gamma(1))/(kidney_05_gamma
    (3)-kidney_05_gamma(2))/(kidney_05_gamma(3)-
    kidney_05_gamma(4))
phi05s = phi05s +((1+kidney_05_gamma(4))**(dt*(i-1))-1-dt*(i
    -1)*kidney_05_gamma(4))/ (kidney_05_gamma(4)**2)/(
    kidney_05_gamma(4)-kidney_05_gamma(1))/(kidney_05_gamma
    (4)-kidney_05_gamma(2))/(kidney_05_gamma(4)-
    kidney_05_gamma(3))
phi05s = kidney_05_gamma(1)*kidney_05_gamma(1)*
    kidney_05_gamma(2)*kidney_05_gamma(2)* kidney_05_gamma(3)
    *kidney_05_gamma(3)*phi05s
phi05b = ((1+kidney_05_gamma(1))**(dt*i-1)-1-(dt*i-1)*
    kidney_05_gamma(1))/ (kidney_05_gamma(1)**2)/(
    kidney_05_gamma(1)-kidney_05_gamma(2))/(kidney_05_gamma
    (1)-kidney_05_gamma(3))/(kidney_05_gamma(1)-
    kidney_05_gamma(4))
phi05b = phi05b +((1+kidney_05_gamma(2))**(dt*i-1)-1-(dt*i
    -1)*kidney_05_gamma(2))/ (kidney_05_gamma(2)**2)/(
    kidney_05_gamma(2)-kidney_05_gamma(1))/(kidney_05_gamma
    (2)-kidney_05_gamma(3))/(kidney_05_gamma(2)-
    kidney_05_gamma(4))
phi05b = phi05b +((1+kidney_05_gamma(3))**(dt*i-1)-1-(dt*i
    -1)*kidney_05_gamma(3))/ (kidney_05_gamma(3)**2)/(
```

```
   kidney_05_gamma(3)-kidney_05_gamma(1))/(kidney_05_gamma
   (3)-kidney_05_gamma(2))/(kidney_05_gamma(3)-
   kidney_05_gamma(4))
phi05b = phi05b +((1+kidney_05_gamma(4))**(dt*i-1)-1-(dt*i
   -1)*kidney_05_gamma(4))/ (kidney_05_gamma(4)**2)/(
   kidney_05_gamma(4)-kidney_05_gamma(1))/(kidney_05_gamma
   (4)-kidney_05_gamma(2))/(kidney_05_gamma(4)-
   kidney_05_gamma(3))
phi05b = kidney_05_gamma(1)*kidney_05_gamma(1)*
   kidney_05_gamma(2)*kidney_05_gamma(2)* kidney_05_gamma(3)
   *kidney_05_gamma(3)*phi05b
if (phi05b< 0) then
phi05b = 1e-20
end if
P05=exp(-kidney_05_lambda*phi05s)-exp(-kidney_05_lambda*
   phi05b)
Q1=  N1*P03
Q2 = N2*P04
Q3 = N3*P05
IF(Q1 < 0 .OR.Q2 < 0 .OR. Q3 < 0) THEN
status = 1
END IF
Q = Q1+ Q2 + Q3
END SUBROUTINE CalculateQ

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!Function: DenormalizeX
!Description: Scale parameter value from (0,1) range to
!             actual range
!Parameter: x = parameter in the model
!           xrange = range of parameter in the model
!Return: DenormalizeX
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
DOUBLE PRECISION FUNCTION DenormalizeX(x, xrange)
IMPLICIT NONE
DOUBLE PRECISION, INTENT(IN) :: x,xrange(1:3)
DenormalizeX = x*(xrange(2)-xrange(1))+xrange(1)
END FUNCTION DenormalizeX

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!Subroutine: Q_from_pikaia
!Description: to obtain the predicted cancer cases
!Input: m = the number of parameters
!       X = an array of parameters with m elements
!Output: Q = the predicted cancer cases at each age period
```

```
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

SUBROUTINE Q_from_pikaia(X, m, N, Q)
IMPLICIT NONE
INTEGER, INTENT(IN) :: m,N(1:MaxAge)
DOUBLE PRECISION, INTENT(IN) :: X(m)
DOUBLE PRECISION, INTENT(OUT) :: Q(1:MaxAge)
DOUBLE PRECISION alpha1,kidney_03_gamma(2),kidney_03_lambda,
    alpha2,kidney_gamma(3),kidney_lambda,kidney_delta,
    kidney_05_gamma(4),kidney_05_lambda
DOUBLE PRECISION Q1(1:MaxAge), Q2(1:MaxAge),Q3(1:MaxAge),
    N1_pred(1:MaxAge),N2_pred(1:MaxAge), N3_pred(1:MaxAge)
INTEGER status
alpha1 = DenormalizeX(DBLE(X(1)),alpha1Range)
alpha2 = DenormalizeX(DBLE(X(2)),alpha2Range)
kidney_03_gamma(1) = DenormalizeX(DBLE(X(3)),
    kidney_03_gamma1Range)
kidney_03_gamma(2) = DenormalizeX(DBLE(X(4)),
    kidney_03_gamma2Range)
kidney_03_lambda = DenormalizeX(DBLE(X(5)),
    kidney_03_lambdaRange)
kidney_gamma(1) = DenormalizeX(DBLE(X(6)),kidney_gamma1Range
    )
kidney_gamma(2) = DenormalizeX(DBLE(X(7)),kidney_gamma2Range
    )
kidney_gamma(3) = DenormalizeX(DBLE(X(8)),kidney_gamma3Range
    )
kidney_lambda = DenormalizeX(DBLE(X(9)),kidney_lambdaRange)
kidney_delta = DenormalizeX(DBLE(X(10)),kidney_deltaRange)
kidney_05_gamma(1) = DenormalizeX(DBLE(X(11)),
    kidney_05_gamma1Range)
kidney_05_gamma(2) = DenormalizeX(DBLE(X(12)),
    kidney_05_gamma2Range)
kidney_05_gamma(3) = DenormalizeX(DBLE(X(13)),
    kidney_05_gamma3Range)
kidney_05_gamma(4) = DenormalizeX(DBLE(X(14)),
    kidney_05_gamma4Range)
kidney_05_lambda = DenormalizeX(DBLE(X(15)),
    kidney_05_lambdaRange)
DO j = 1, MaxAge
N1_pred(j)=N(j)*alpha1
N2_pred(j)=N(j)*alpha2
N3_pred(j)=N(j)-N1_pred(j)-N2_pred(j)
! write (5,'(ES14.6,A,ES14.6,A,ES14.6)')N1_pred(j),' ',
    N2_pred(j),' ',N3_pred(j)
```

```
CALL CalculateQ(alpha1,alpha2, kidney_03_gamma,
    kidney_03_lambda,kidney_gamma, kidney_lambda,
    kidney_delta, kidney_05_gamma, kidney_05_lambda, N1_pred(
    j),N2_pred(j), N3_pred(j), j, Q(j), Q1(j),Q2(j),Q3(j),
    status)
END DO
END SUBROUTINE Q_from_pikaia
END PROGRAM RenalCarcinomas
```