Modeling carcinogenesis and cancer stages for latency estimation in dynamic geographic systems: Linking genomic, cellular, individual and population levels

(1) Pancreatic cancer

Author: Geoffrey M. Jacquez[1,2]

Affiliations: 1 BioMedware 2; Department of Geography, University at Buffalo

BioMedware research report

Presented at the 2013 meetings of the North American Association of Central Cancer Registries, Austin Texas, June 12, 2013

Citation: Jacquez, G. M. 2013. "Modeling carcinogenesis and cancer stages for latency estimation in dynamic geographic systems: Linking genomic, cellular, individual and population levels. (1) Pancreatic cancer". BioMedware Research Report. 46 pages. Published online www.biomedware.com.

Dedication: This work is dedicated to our colleague and friend Dr. Jawaid Rasul, who died of pancreatic cancer in May 2011 approximately one year after diagnosis.

Abstract: Understanding latency – the delay between a cause and its effect – is critical when modeling dynamic geographic systems in general and in particular for human health outcomes such as cancer. This paper uses compartmental models to estimate residence times in states defining the progression of pancreatic cancer. Two models are developed. The first models carcinogenesis and how cancer evolves in an organ (the pancreas) based on the cascade of mutations and cellular changes that lead to metastatic cancer. The second is a model of cancer stages as defined by diagnostic criteria for the progression from early to late stage cancers. The unit of observation for the carcinogenesis model is the pancreatic cell; for the stage-based model it is the individual cancer patient. These models are linked using a logical mapping of the molecular and cellular characteristics of pancreatic cancer cells to the stage at diagnosis. The resulting system provides the following: (1) Empirically-based, biologically reasonable estimates of the distribution of residence times in cellular states and pancreatic cancer stages; (2) Conditions for metastasis and remission; (3) Estimates of the total burden of yet to be diagnosed cancers (we call this the silent cancer burden); (4) Maps of the geographic distribution of the silent cancer burden. The modeling approach developed in this paper may be used in disease surveillance and disease clustering to reveal where people lived when they were vulnerable to exposures that could have caused their disease. It also may be an important advance in our understanding of the latencies for the progression from one disease state to another, both for individuals as well as populations. Finally, the modeling approach links our emerging knowledge of cancer genomics to cancer progression at the cellular level, to individuals and the stage of their cancer at diagnosis, and finally to population-level outcomes describing geographic distributions of cancer in extant populations.

Introduction

In post-industrialized nations the rise of antibiotics in the 1940's, coupled with public health advances including vaccines and the establishment of infrastructure and policies for the wide provision of clean water, sanitation and food, led to an era of chronic diseases that began in the 1950's and continues to the present day (Brownson and Bright 2004). These public health advances greatly reduced childhood mortality and the major causes of population-based mortality switched from respiratory and other infections to chronic diseases including heart disease and cancers. When modeling the geographic aspect and spread of disease, whether chronic, outbreak, epidemic or endemic, there is a potential disconnect whenever the time between the action of causal factors promoting disease and the diagnosis event is long. This time between causal factors and diagnosis is called a *latency*, and the disconnect arises when latency obscures geographic patterns in the action of the causal factors. For chronic diseases such as cancer we often use the place of residence at diagnosis or death to record the health event. But where people reside at time of diagnosis may be far removed from where they lived when causative exposures occurred. This disconnect is widely recognized, yet techniques for estimating appropriate sampling distributions for latencies for the geographic modeling of human diseases that are biologically reasonable, based on observable disease states, and that incorporate knowledge of disease progression are seldom available. This paper addresses this need.

In addition, as our understanding of the genetic bases of disease increases, the need for *systems biology* approaches that integrate across genetic, cellular, organ, individual and population-level scales is increasingly recognized. How can we incorporate knowledge, for example, of the cascade of genetic mutations leading to pancreatic cancer into our understanding of cancer latency, and how might this impact estimates of the burden of cancer at the population level? How do changes manifested in pancreatic cells as a result of mutations translate into cancer progression, and can we use this information to better understand conditions of metastasis and remission? This paper addresses this need by linking a model of carcinogenesis at the cellular level with a model of cancer stages at the individual and population level.

We begin with an introduction to the approach for the modeling and analysis of dynamic geographic systems using process-based temporal lags. This is followed by a brief background on a range of latency estimation approaches that motivate the use of residence times in compartmental systems. A primer on compartmental analysis is presented, followed by a simple three stage model of disease, and results for distribution of residence times. Next, the specific example of pancreatic cancer is considered, and a five state model of carninogenesis is developed along with its biological foundation. A second model of progression through cancer stages based on diagnostic criteria used by the American Cancer Association follows. These models are linked using knowledge of the mapping of stage of diagnosis with progression of tumor growth and metastatic capacity. This is applied to data from the Michigan cancer registry on stage at diagnosis for all incident pancreatic cancers in white males from 1985 to 2005 in the Detroit metropolitan area. Potential applications of this approach and next steps are then discussed.

The analysis of dynamic geographic systems involving cancer

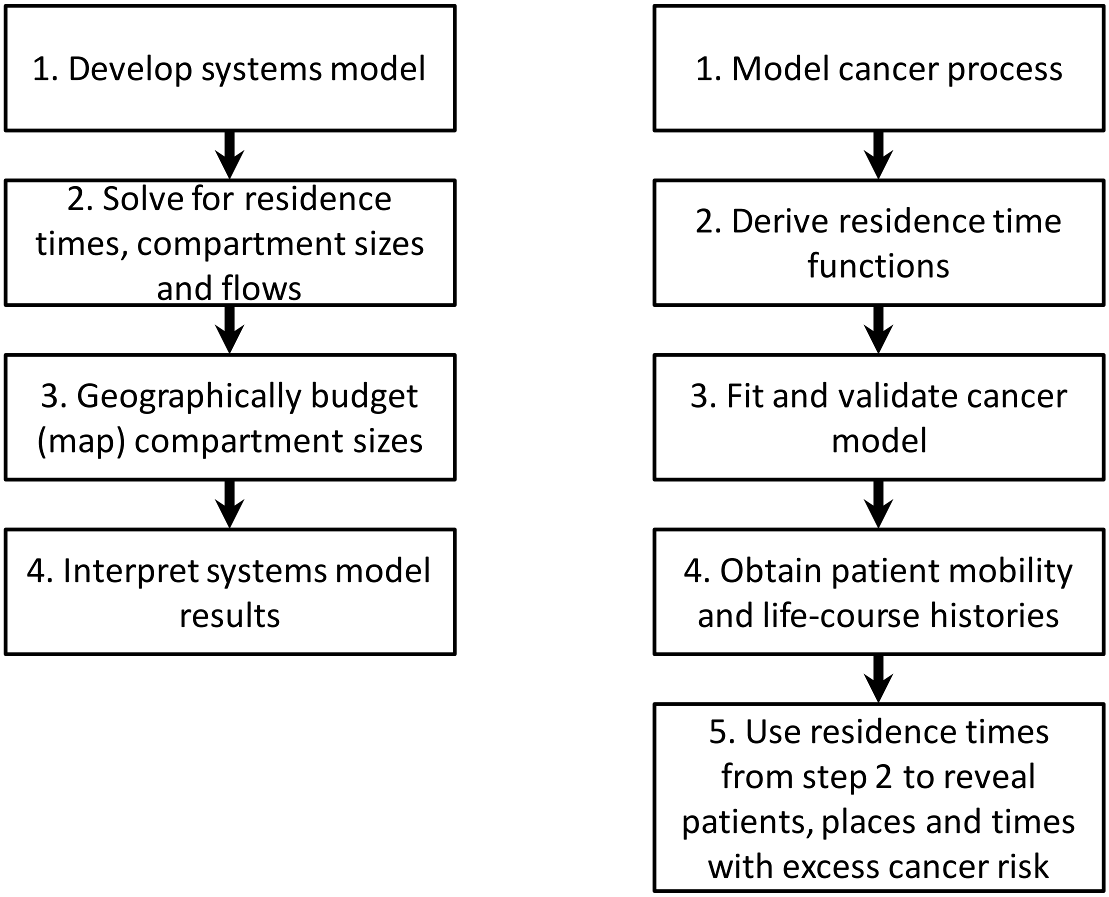

1. Develop systems model
2. Solve for residence times, compartment sizes and flows
3. Geographically budget (map) compartment sizes
4. Interpret systems model results
1. Model cancer process
2. Derive residence time functions
3. Fit and validate cancer model
4. Obtain patient mobility and life-course histories
5. Use residence times from step 2 to reveal patients, places and times with excess cancer risk

Figure 1. Steps in dynamic geographic systems analysis (left) and specific application to cancer using knowledge of residential history to budget excess risk (right).

A schematic of the modeling approach is shown in Figure 1. The generalized approach (Figure 1, left) can be applied to any geographic system amenable to a compartmental or systems model representation. Here the emphasis is on the development of a minimally sufficient but mechanistically reasonable systems model. A specific example illustrating application to a dynamic geographic systems model of cancer using residence times to estimate the space-time lag is shown in Figure 1, right. The remainder of this paper will deal specifically with the development and application of systems models of disease, but the approach is generalizable to other geographic systems that arise in sociology, economics and geomorphology.

We desire systems models with several characteristics. First, they must be biologically reasonable and capture relevant aspects of disease etiology and natural history (e.g. known disease states). Second, they must provide estimates of the distributions of disease latency. These are key in order to quantify space-time lags in geographic dynamic systems. Third, they must be estimable from empirical data, so that we can derive latency distributions from observable measures and based on the current state of knowledge of the disease. The derivation of biologically-based estimates of disease latency is a difficult problem, and we next consider alternative approaches to latency estimation.

Temporal lag estimation for disease

Several techniques exist for modeling disease latency, including representations of cohort exposures, developmental stages of vulnerability, models of empirical induction periods, and compartmental models. We briefly summarize each of these before focusing on residence times in compartmental models.

Cohort exposures arise when a common exposure is hypothesized for a population or group of individuals, resulting in an overall increase in disease risk. Salient examples are exposure to ionizing radiation from the Hiroshima and Nagasaki bombs; and the release of radionucleotides from the Chernobyl nuclear plant meltdown with airborne transport in Russia and northern Europe. Here the temporal lag between the causal event and later health outcomes known to be associated with the causal event are directly observable, especially when the background risk for the health outcome is small in the absence of the causal event. For Chernobyl, radioactive iodide was released over Belarus and led to an increase in pediatric thyroid cancers. The latent period for tumor development was 4–6 years, with a mean of 5.8 years (Nikiforov and Gnepp 1994).

Developmental stages of vulnerability arise when the timing and characteristics of biological stages of development are associated with increased risk of an adverse health event in later years. Consider breast cancer, which has known genetic and environmental risk factors. Here, the inherited genetic risk for breast cancer accounts for approximately 10-15% of breast cancer cases, and the timing of reproductive events (windows of vulnerability) may be critical. Of particular importance are the exposures that occur before a woman's first birth, and during the development of breast tissues (Colditz and Frazier 1995). Since the timing of the developmental stage is often observable, an average latency and its distribution may be estimated as the time from the developmental stage to disease diagnosis.

The *Empirical Induction Period* (EIP) models the lag between initial onset and manifestation of disease as the sum of induction and latent periods. These are the periods between causal action and disease initiation (induction), and between disease initiation and detection (latent). The combined length of the induction and latent periods is the empirical induction period. The length of the induction period is not estimable except in relation to specific etiologic factors, since different exposures (for example) might have different levels of effect in terms of disease expression (Rothman 1981). An important result is that invalid assumptions regarding the duration of the empirical induction period result in misclassification errors and bias toward the null. Rothman recommended that sensitivity analyses varying the length of the empirical induction period can be used to minimize misclassification error, a useful finding whenever the induction period is not directly estimable.

Residence times in compartmental models of disease may be obtained directly from the model itself, that is, when one has a compartmental model and parameter values in hand, the mean residence time and distribution of residence times in each compartment are known. This result has been demonstrated for both deterministic and stochastic compartmental models, but to this authors knowledge has yet to be used in geographic models of human disease. When the models are constructed to correspond to stages of disease known to comprise the latency process, the residence times in these compartments may be used as estimates of disease latency. Compartmental models (defined below) are best constructed such that the compartments correspond to known disease states (e.g. are biologically reasonable), and so that the coefficients governing transitions between disease stages are formulated in terms of known biological and infection processes (e.g. the mechanics are process-based). Once the parameters of the model are identified, whether or not the model is estimable may be determined for any given set of inputs. Residence times from compartmental models thus convey the characteristics required at the beginning of this section; (1) they may be formulated in a biologically reasonable fashion. (2) They provide estimates of the distribution of latencies. (3) Whether a given model, and hence its residence times, is estimable is known once the model and observable measures are identified. The remainder of this paper employs compartmental models.

Primer on compartmental modeling

This section provides an introduction to compartmental models, beginning with their representations and founding assumptions. A simple 1-state model is then presented, along with the system equations and derivation of residence times for that simple model. Considerations in model formulation, including tradeoffs between simplicity/abstraction versus complexity, and the handling of real-world heterogeneity, are next, followed by means for accomplishing the transition from deterministic models to their stochastic analogs.

Model Representation and Definitions There is a substantial literature on compartmental models and compartmental systems; their behavior, properties and theory are well known for both deterministic and stochastic formulations. Compartmental analysis is both powerful and flexible, and diverse approaches including Markov chains, linear, non-linear and systems of ordinary as well as partial differential equations often may be expressed as compartmental systems.

A compartment is a quantity of some material considered to be homogeneous in two respects. First, additions to the compartment are instantaneously and uniformly mixed throughout the compartment. Second, small amounts of material leaving the compartment have the same chance of leaving as any other small amount of material. This second property makes it possible to derive closed form solutions to compartment residence times – how long it takes, on average, for a small amount of material to leave the compartment.

A compartmental system is constructed from compartments, the flows between them, and the inflows from and outflows to the environment. The flows into and out of compartments are labeled with transfer coefficients, the product of a transfer coefficient and the size of the compartment it is leaving gives the rate of transfer (exit) of the material from that compartment.

Diagrams of compartmental systems may be drawn using boxes, circles or nodes to represent compartments, and directed line segments or arcs to represent flows between compartments. Inflows and outflows from outside are usually indicated with arrows indicating external sources and sinks (e.g. Figure 2). Such diagrams are useful for developing a cognitive model of the system under scrutiny, and later guide formulation of the system equations.

Assumptions: The term compartment has been called a “kinetic construct” in that it does not necessarily refer to a mechanical construct, such as a container or bucket. The flow of water through the plumbing in a house may be represented as a compartmental system with flows between mechanical constructs such as the reservoir, water basins, and bathtubs. The flow of water between the kinetic constructs in a lake of the bottom layer (Hypolimnion), thermocline (Metalimnion), and upper layer (Epilimnion) may also be analyzed as a compartmental system. In this paper we use kinetic constructs defined to be pancreatic cells in various states of carcinogenesis, and individuals in different diagnostic stages of pancreatic cancer.

The homogeneity and instantaneous mixing properties are strong assumptions when one extends compartmental analysis to systems where the material itself is heterogeneous in some regard. For applications involving enzyme kinetics and the decay of radioisotopes the homogeneity and instantaneous mixing assumptions seem reasonable, these assumptions warrant more careful consideration when the compartments correspond to stages of a disease such as cancer. Here there can be substantial differences in the speed of disease progression depending on an individual's genetic susceptibility and the genetics of the tumor itself. These have been handled in stage-based compartmental models of cancer by introducing sub-compartments corresponding to finer tumor stages reflecting the makeup of the cancer clones. In practice mixing is never instantaneous, what matters are the magnitude of the flows between compartments relative to the size of the compartment and the speed with which actual mixing occurs.

For deterministic compartmental models, a final assumption is the treatment of flows between compartments as continuous. This is reasonable when the number of particles comprising specific flows between compartments is large. But when flows represent individuals (e.g. as in a population growth, infectious disease or migration model) and the flows are small the discrete nature of the individuals may be lost. For example, the notion of a flow of 0.12 persons is not realistic since individuals are not divisible. Then the system should be treated as a stochastic compartmental model, with stochasticity introduced by the probabilistic treatment of discrete events (e.g. whether the transition for an individual from 1 disease state to another transpires is evaluated probabilistically).

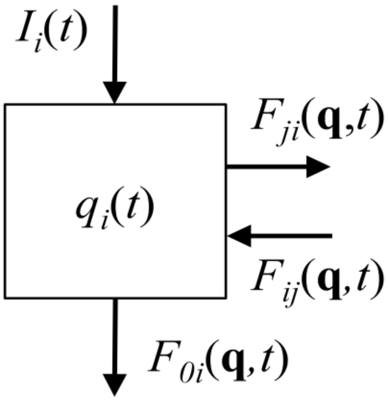


Figure 2. i^{th} compartment of a compartmental system with flows to ($F_{0i}(\mathbf{q},t)$) and from ($I_i(t)$) outside the system. Flows to and from other compartments are $F_{ji}(\mathbf{q},t)$ and $F_{ij}(\mathbf{q},t)$, respectively. The size of the i^{th} compartment at time *t* is $q_i(t)$.

Single compartment of an n-compartment system: Now let's turn our attention to a general model to illustrate the derivation of the system equations and residence times. In this section we use the notation and summarize relevant results presented by Jacquez (Jacquez 1996). Figure 2 illustrates a compartment in a larger system with flows from and to the environment ($I_i(t)$ and $F_{Oi}(\mathbf{q},t)$) and to ($F_{ji}(\mathbf{q},t)$) and from ($F_{ij}(\mathbf{q},t)$) the jth compartment. These flows are shown as functions (hence the notation in capital letters for F and I). The functions are evaluated at time t, and the ones from the system compartments depend on the sizes of the compartments, $\mathbf{q}$. Here the bold notation indicates the vector of compartment sizes $\mathbf{q} = \{q_1, q_2, .., q_n\}$. We then can write the instantaneous rate of change of the size of the ith compartment as:

$$\frac{dq_i}{dt} = \sum_{j\neq i}^{n} \left(F_{ij}(\mathbf{q},t) - F_{ji}(\mathbf{q},t)\right) + I_i(t) - F_{Oi}(t). \qquad \text{(Eqn 1)}$$

This maintains conservation of mass such that the change in size of the compartment is equal to the sum of the inflows minus the sum of outflows. It is useful to rewrite the flows between compartments in terms of the product of *transfer coefficients* and compartment sizes:

$$F_{ij}(\mathbf{q},t) \equiv f_{ij}(\mathbf{q},t) q_j(t). \qquad \text{(Eqn 2)}$$

Here the transfer coefficient may be thought of as that proportion of the source compartment j that flows to the destination compartment i in some small time dt. We then rewrite Equation 1 in terms of transfer coefficients, compartment sizes, and dropping the arguments yielding:

$$\frac{dq_i}{dt} = I_i + \sum_{j\neq i}^{n} f_{ij} q_j - q_i\left(f_{Oi} + \sum_{j\neq i}^{n} f_{ji}\right) \qquad \text{(Eqn 3)}$$

This is a budget for the changes in size of compartment i in terms of inflows from outside and from other compartments, and the flows to outside and to other compartments. This maintains a mass balance condition such that the materials in the compartments cannot be created or disappear from the system of compartments. Linear compartmental systems arise when all of the transfer coefficients are constant or functions only of time. When the transfer coefficients are functions of compartment sizes then it is a nonlinear compartmental system. This has implications for the residence times, as we will see later.

Notation: Before considering residence times it is useful to summarize the notation used in the remainder of this paper.

q_i: Generally used to indicate the size of compartment i, in this paper it is the number of persons in the ith disease state or cancer stage. Units are number of persons.

x_i: It is sometimes useful to standardize by population size or in terms of number of cases in all stages of disease. In these instances the units are (cases in stage i)/population-at-risk, or (cases in stage i)/(cases in all stages).

F_{ij}: The instantaneous rate of flow to compartment i from compartment j, usually a function of compartment sizes and time (e.g. $F_{ij}(\mathbf{q}, t)$). We continue to write the arguments only as necessary. For us, this is the number of persons moving from cancer stage j to stage i. Units are persons/time.

F_{0i}: The instantaneous rate of flow out of the system from compartment i. This is the number of persons exiting cancer stage *i*, usually by remission or death. The units are persons/time.

f_{ij}: The transfer coefficient for the flow to compartment i from compartment j, $f_{ij}q_j = F_{ij}$. The units are t^{-1}.

f_{0j}: The transfer coefficient for the flow to the outside from compartment j, $f_{0j}q_j = F_{0j}$. The units are t^{-1}.

I_i: The rate of flow from outside of the system into the i^{th} compartment. May be time-dependent or a function of some compartment sizes.

k_{ij}: Denotes a constant transfer coefficient for the flow from compartment j to compartment i.

t: Time

Residence times: Residence time can be defined as the average time required for a particle to enter and then exit a compartment. Ideally, compartments are constructed to correspond to meaningful states of the system being modeled, in which case compartment residence times provide an important means of model validation. That is, the residence times from the model are compared to the corresponding empirical residence time in the system being modeled. For linear compartmental models with constant transfer coefficients the residence times are inverse exponential functions, and a complete theory for calculating the probability density functions (pdf's) is in place (Jacquez 1996). In deterministic non-linear compartmental systems the distributions of residence times are functions of the state variables, and hence of the compartments sizes. The probability density functions of linear stochastic models are the same as for their deterministic analog. However, the probability density functions of residence times for non-linear stochastic systems differ from those of their non-linear deterministic counterpart (Jacquez 2002). In this paper we present results for linear deterministic stage-based models of cancer, which should apply to their linear stochastic counterparts (Table 1).

Model Type	Transfer coefficients /rate laws	Residence times	Probability Density Function of residence times
Linear deterministic	Constant transfer coefficients	Depend only on compartments where material is injected	Negative exponential
Non-linear deterministic	Transfer coefficients are functions of compartment size	Functions of state variables; depends on occupancies of system compartments	Erlang distribution
Linear stochastic	Constant rate laws	Same as linear deterministic analog	Negative exponential
Non-linear stochastic	Non-linear rate laws	Different from analogous non-linear deterministic system	Erlang distribution

Table 1. Characteristics of probability density functions of residence times for deterministic and stochastic compartmental models. Summarized from Jacquez (2002).

Simplicity versus complexity, and implications for residence times: When constructing models there is a tension between simplicity, which makes models more easy to understand and mathematically tractable, and complexity, which seeks to incorporate the nuances and details of a complex reality. In compartmental models, simplicity may correspond to a representation with fewer compartments; an implicit combining of compartments that has implications for the modeling of residence times. When the residence times for a compartment in a model are too short, the creation of sub-compartments to represent that compartment can be used to obtain longer average residence times (Jacquez and Simon 2002). Correspondence of residence times to those observed in the system under scrutiny thus can be used as a diagnostic for model over-simplification and misspecification.

Transition from deterministic to stochastic compartmental models: This paper presents results from a linear deterministic compartmental model of pancreatic cancer stages. As shown in Table 1 (above), the formulae for the probability density functions of residence times from the linear deterministic model should apply to its stochastic analog. The transition from a linear deterministic to its linear stochastic analog is accomplished using methods of *model transition sensitivity analysis*, and is not presented here. See Koopman et al for details (Koopman, Jacquez et al. 2001).

Modeling carcinogenesis

For the geographic modeling of disease we are interested in identifying those places and sub-populations characterized by an excess of cancer for individuals in those states of carcinogenesis when exposures to mutagens might have been causal. That is, we are looking for the geographic signature of the actions of past environmental exposures that gave rise to cancers. To do this we require biologically reasonable models of carcinogenesis (e.g. the biological events that have cancer as their sequelae) and cancer stages (how cancers progress once they have started). We begin with carcinogenesis.

The initial biological event leading to cancer is damage to DNA. Such damage occurs on one DNA strand, and repair mechanisms can reverse that damage. Whether the damage is maintained among daughter cells depends on the timing of replication and repair. If replication occurs before repair then the damaged DNA strand is passed on to the daughter cells (a mutation). Notice that only some of these mutations are deleterious and lead to cancers. The mechanism of replication differs between meiosis and mitosis. Mitosis is the cell division that is undertaken for growth of the organism and replacement of aging cells (e.g. "life cycle" replication); meiosis is a special kind of cell division necessary for the production of gametes in sexual reproduction. In mitosis the DNA is copied and passed on to the daughter cells. The daughter cells are genetically identical to the parent line save for mutations and errors in DNA replication. In meiosis recombination occurs and each resulting sperm or egg has only ½ of the genetic complement of the parent cell. These details have implications for models of carcinogenesis. Here we are concerned with life cycle replication, a model for meiosis is given in the Appendix.

The usual approach for modeling carcinogenesis is to treat irreversible steps in the chain of mutations leading to cancer as comprised of sub-states with reversible damage attributable to DNA repair mechanisms (Kopp-Schneider, Portier et al. 1991; Jacquez 1999). The last few years have seen dramatic advances in our understanding of tumor genetics, and it now is possible to sequence the genomes sampled from cancer tumors to elucidate the sequence of mutations that lead to cancer. The specific mutations may vary from one tumor to another and from one patient to another, but the steps of mutation, repair, and fixation of deleterious mutations via replication events are largely the same. The sub-states of a model of carcinogenesis thus should be constructed to correspond to the observed tumor morphological characteristics, with flows corresponding to state transitions from mutation, repair, and replication.

Model of pancreatic cancer carcinogenesis

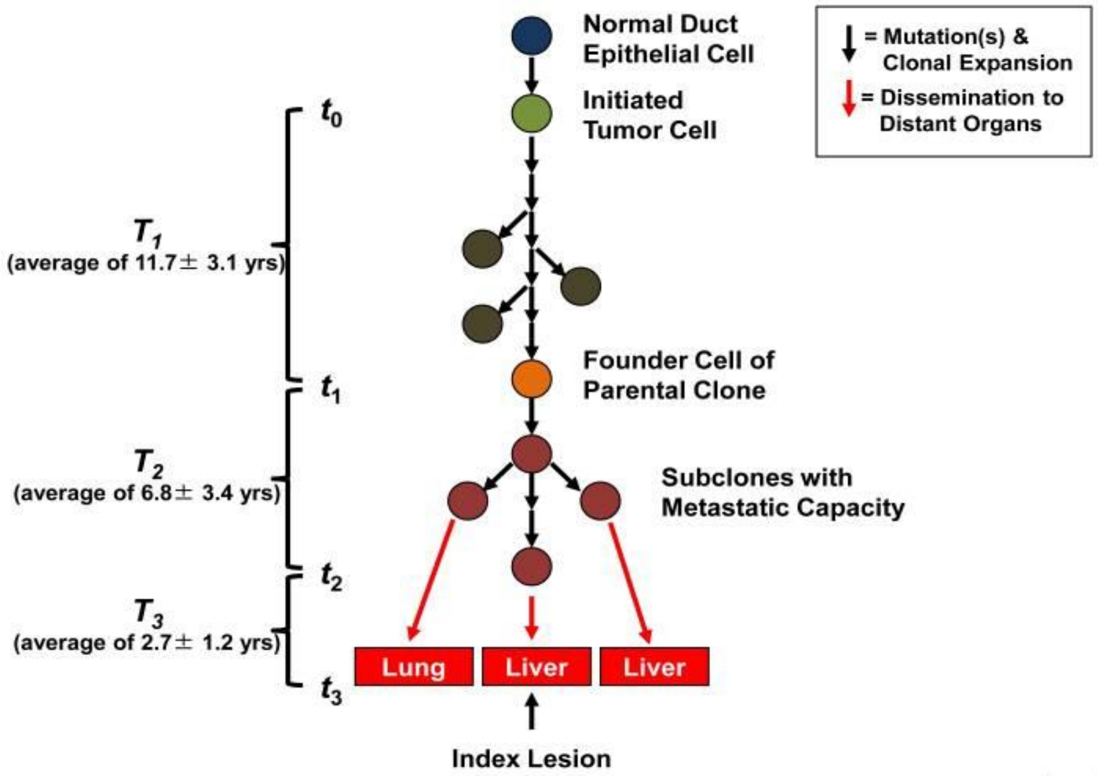

Normal Duct Epithelial Cell
Initiated Tumor Cell
= Mutation(s) & Clonal Expansion
= Dissemination to Distant Organs
t_0
T_1 (average of 11.7 ± 3.1 yrs)
t_1
Founder Cell of Parental Clone
T_2 (average of 6.8 ± 3.4 yrs)
Subclones with Metastatic Capacity
t_2
T_3 (average of 2.7 ± 1.2 yrs)
t_3
Lung
Liver
Liver
Index Lesion

Figure 3. Schematic of the evolution of pancreatic cancer. Normal pancreatic duct epithelial cells undergo mutation events to become an initiated tumor cell. Additional mutations and clonal expansions lead eventually to a founder cell of the index pancreatic cancer clone. These produce subclones with metastatic capacity, eventually leading to dissemination to distant organs such as the liver. Times shown are the empirical residence times in each system state. Adapted from Yachida, Jones et al. (2010).

Consider pancreatic cancer carcinogenesis (Figure 3) and its corresponding compartmental model (Figure 4). A recent study sequenced the genome of pancreatic cancer tumors in 24 patients, allowing the reconstruction of mutation events leading to pancreatic cancer, and estimation of the durations (residence times) associated with each cancer state (Yachida, Jones et al. 2010). Here we abstract only those biological features and mutation events critical to modeling pancreatic cancer, refer to Yachida, Jones et al. (2010) for information on the site-specific mutations. There are a host of driver genes of pancreatic carcinogenesis including KRAS, CDKN2A, TP53 and SMAD4, among others, and a cascade of specific mutation events appear to be responsible for pancreatic cancer, although these differ from one patient to another (Maitra and Hruban 2008). Precursor lesions include the mucinous cycstic neoplasm (MCN), the intraductal papillary mucinous neoplasm (IPMN) and the pancreatic intraepithelial neoplasia (PanIN). Here we consider the PanIN pathway, which is thought responsible for the majority of pancreatic cancers. How this PanIN-focused model may be generalized to include the MCN and IPMN pathways is described in the Discussion.

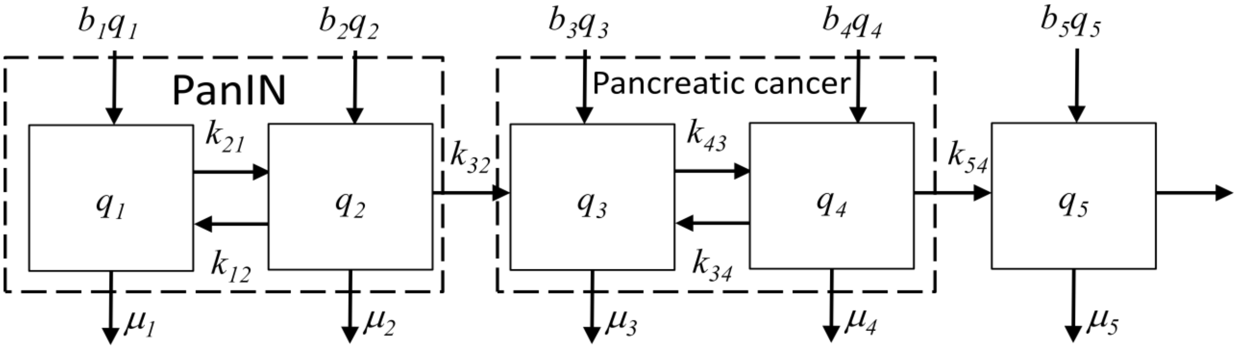


Figure 4. Proposed model of pancreatic cancer carcinogenesis.

Carcinogenesis is initiated by a mutation in a normal cell that leads to accelerated cell proliferation. Waves of clonal expansion along with additional mutations progress to pancreatic intraepithelial neoplasia (PanIN) during time T_1 (Figure 3). This corresponds to the sub-states q_1 (normal cell) and q_2 (PanIN) in the model in Figure 4. One founder cell from a PanIN lesion will start the parental clone that will initiate an infiltrating carcinoma; this is indicated by the irreversible flow (k_{32}) from q_2 to q_3 in Figure 4. Here sub-state q_3 is the parental clone, and sub-state q_4 indicates sub clones with metastatic capacity. The flow k_{32} to sub-state q_3 thus represents that replication that gives rise to the index pancreatic cancer lesion (the cells in q_3), along with the mutation events that confer metastatic capacity (resulting in the cells in q_4) . The empirical residence time in sub-states q_3 and q_4 is T_2. The irreversible flow k_{54} indicates a proliferation and spreading of cells with metastatic capacity, with metastases (state $q_{5\,)}$ to other organs such as the liver occurring in time T_3. The observed average times in each model state are T1=11.7 years, T2=6.8 years , and T3=2.7 years (Campbell, Yachida et al. 2010). These are the empirical residence times in those states describing pancreatic carcinogenesis, and were estimated from tumor histology and tumor genetics. It is worth noting this model is consistent with recent findings regarding mechanisms of pancreatic tumorigenesis. For example, inflammation and injury has been implicated as a precursor event in some pancreatic cancers, leading to acinar-to-ductal metaplasia (ADM). ADM is reversible, but an oncogenic mutation in KRAS prevents this, and the injured cells enter the pathway to pancreatic intraepithelial neoplasia (PanIN). Additional mutation events then can result in pancreatic ductal adenocarcinoma (Seton-Rogers 2012), represented by the “pancreatic cancer” meta-compartment in Figure 4.

There are five states that were described in the preceding paragraph, q_1,…, q_5. State q_1 corresponds to normal cells that divide and give rise to new normal cells, this is the input b_1q_1.

State q_2 is those cells in which one strand of DNA has been damaged, that occurs at rate $k_{21}q_1$ and is repaired at rate $k_{12}q_2$. Cell division splits the DNA into one strand each of damaged and normal DNA, these are then copied and healthy cells with one normal and one damaged strand results in the inflow of cells (proliferation through cell division) into q_2 with rate b_2q_2. A second mutation or set of mutations results in an irreversible promotion event to become the founder pancreatic cancer cells and their descendants in state q_3. Additional DNA damage occurs at rate $k_{43}q_3$ giving rise to the cells in state q_4 that have metastatic capacity. Proliferation and spread of these metastatic cells beyond the primary site results in metastases (state q_5) occurring at rate $k_{54}q_4$. The model in Figure 4 has two irreversible steps, one from q_2 to q_3 and the second from q_4 to q_5. The per cell death rates for each state are μ_1,... μ_5. The cell proliferation and death rates are given by the b_i and μ_i terms. The transfer coefficients and their underlying biological mechanisms are:

Coefficient	Biological Mechanism
k_{21}	Initiating DNA damage of normal pancreatic cancer cell
k_{34}, k_{12}	DNA repair
k_{32}	Promotion to pancreatic cancer cell, by additional mutation and/or gene expression
k_{43}	Promotion to pancreatic cancer with metastatic capacity, by additional mutation and/or gene expression
k_{54}	Formation of metastases; spread of primary cancer to distant sites

System equations: The system equations for the pancreatic cancer model are given in Equation 4.

$$dq_1/_{dt} = q_1(b_1 - k_{21} - \mu_1) + q_2k_{12} \qquad \text{(Eqn 4)}$$

$$dq_2/_{dt} = q_1k_{21} + q_2(b_2 - k_{12} - k_{32} - \mu_2)$$

$$dq_3/_{dt} = q_2k_{32} + q_3(b_3 - k_{43} - \mu_3) + q_4k_{34}$$

$$dq_4/_{dt} = q_3k_{43} + q_4(b_4 - k_{34} - k_{54} - \mu_4)$$

$$dq_5/_{dt} = q_4k_{54} + q_5(b_5 - \mu_5)$$

Equilibrium conditions: Equilibrium conditions should obtain only for the normal cells (q_1) and those normal cells with 1 damaged DNA strand (q_2), since the number of normal pancreatic cancer cells in a person does not change to a large extent over the life course. Since pancreatic cancer is defined by rapid cell proliferation, equilibrium conditions are not expected to obtain in states q_3, q_4 and q_5.

Since equilibrium obtains for the normal cells we write the condition that the birth rate is equal to the sum of the death rate plus progression to cancer cells through another mutation event:

$$\frac{d(q_1+q_2)}{dt} = 0, \qquad \text{(Eqn 5)}$$

$$b_1 q_1 + b_2 q_2 = \mu_1 q_1 + q_2(k_{32} + \mu_2).$$

This states that the net inflows into the PanIN compartment on the left hand side of the equation equal the net outflows on the right hand side of the equation. This simplifies to

$$q_1(b_1 - \mu_1) = q_2(k_{32}+\mu_2 - b_2), \qquad \text{(Eqn 6)}$$

Yielding the relationship between the equilibrium compartment sizes and transfer coefficients

$$\frac{q_1}{q_2} = \frac{(k_{32}+\mu_2-b_2)}{b_1-\mu_1}. \qquad \text{(Eqn 7)}$$

This states that the ratio of normal pancreatic cells (q_1) to cells with damaged DNA (q_2) equals the sum of transfer coefficients governing the number of cells with damaged DNA ($k_{32}+\mu_2 - b_2$) divided by the transfer coefficients governing the number of normal pancreatic cells ($b_1 - \mu_1$).

Using Eqn 7 we can write q_2 in terms of q_1.

$$q_2 = q_1(b_1 - \mu_1)/(k_{32}+\mu_2 - b_2) \qquad \text{(Eqn 8)}$$

Condition for cancer proliferation and metastasis to distant sites: When pancreatic cancer is progressing the number of cells in states q_3, q_4 and q_5 will be changing and there is no closed form solution for the sizes of these compartments based on equilibrium conditions. The number of pancreatic cancer cells (q_4 + q_5) will be increasing and the following condition holds

$$\frac{d(q_3+q_{4+}q_5)}{dt} > 0. \qquad \text{(Eqn 9)}$$

This states that the sum of the number of pancreatic cancer cells in the founding lesion plus their daughter clones with metastatic capacity plus the number of cells in metastases are increasing through time. Hence cancer cell creation and proliferation must be greater than the loss to cell death, meaning

$$q_3(b_3 + k_{32}) + q_4 b_4 + q_5 b_5 > q_3\mu_3 + q_4\mu_4 + q_5\mu_{5.} \qquad \text{(Eqn 10)}$$

Condition for proliferation of the parental clone: In the early stages of carcinogenesis the parental clone represented by state q_3 must grow. That is, tumor growth occurs through the rapid proliferation of cancer cells, meaning

$$\frac{dq_3}{dt} \gg 0. \qquad \text{(Eqn 11)}$$

Furthermore

$$q_4 k_{34} + q_3 b_3 + q_2 k_{32} \gg q_3(\mu_3 + k_{43}). \qquad \text{(Eqn 12)}$$

The first term is the repair of cells with metastatic capacity, and likely is small; the second term $q_3 b_3$ is the proliferation of the tumor; the third term is mutation to the cancer ($q_2 k_{32}$) and should be small. Together, these must substantially exceed the death and mutation/progression to metastatic capacity $q_3(\mu_3 + k_{43})$. Assuming the rates of incoming cancer cells due to mutation events ($q_2 k_{32}$) and repair of progressed cells ($q_4 k_{34}$) is small, the condition for cancer proliferation is

$$\sim q_3 b_3 \gg q_3(\mu_3 + k_{43}). \qquad \text{(Eqn 13)}$$

This states that for pancreatic cancer to progress the proliferation of the cancer cells in the parental clone must be much larger than their loss by death and mutation.

Condition for metastatic capacity: The condition for growth of metastatic capacity is

$$\frac{dq_4}{dt} > 0. \qquad \text{(Eqn 14)}$$

For this to occur the inflow of cells with metastatic capacity must exceed their loss due to cell death, progression to distant metastases and repair, specifically

$$q_3 k_{43} + b_4 q_4 > q_4(k_{34} + k_{54} + \mu_4). \qquad \text{(Eqn 15)}$$

This gives a threshold condition for establishment and maintenance of metastatic capacity such that

$$q_3 k_{43} > q_4(k_{34} + k_{54} + \mu_4 - b_4). \qquad \text{(Eqn 16)}$$

Here the flow of cells attaining metastatic capacity $q_3 k_{43}$ through mutation events must be greater than the net number of cells with metastatic capacity that are lost due to cell repair (k_{34}), metastasis to distant sites (k_{54}), and cell death (μ_4), and that are not replaced by proliferation (b_4).

Condition for growth of metastatic cancer at distant sites: The condition for emergence and continued growth of cancers that have metastasized to distant sites is

$$\frac{dq_5}{dt} \gg 0. \qquad \text{(Eqn 17)}$$

Metastasis is usually typified by aggressive, rapid cancers, which is why the much greater than condition is used. To achieve this inequality the condition

$$q_4k_{54} + b_5q_5 \gg q_5\mu_5 \qquad \text{(Eqn 18)}$$

must hold such that the inflows of metastatic cancer cells from proliferation and establishment of new metastases must exceed metastatic cancer cell death. This yields a requirement for the establishment and growth of metastatic cancers,

$$q_4k_{54} \gg q_5(\mu_5 - b_5). \qquad \text{(Eqn 19)}$$

This states that the number of cancer cells with metastatic capacity that migrate from the pancreas (q_4k_{54}) to establish new metastases and replenish existing ones is much greater than the net loss in the number of cancer cells in distant metastases due to the difference between cell death and proliferation, $q_5(\mu_5 - b_5)$. This implies that treating metastases without also reducing the migration of cells with metastatic capacity from the primary site is not an effective treatment strategy.

Condition for remission: The establishment and maintenance of remission implies that metastases as well as tumors at the primary site are absent or shrinking. For metastases at distant sites this is expressed as

$$\frac{dq_5}{dt} < 0, \qquad \text{(Eqn 20)}$$

yielding the condition

$$q_4k_{54} < q_5(\mu_5 - b_5). \qquad \text{(Eqn 21)}$$

Hence the migration of cells with metastatic capacity to existing and new distant sites must be less than the net loss of cancer cells in existing metastases to cell death and proliferation. This suggest a possible explanation to the problem in cancer treatment of metastatic latency, where metastases go into remission but then recur: treatment of metastases targets the right hand side of equation 21 and reduces the term $q_5(\mu_5 - b_5)$. This potentially invalidates the "less than" term, and the condition for remission no longer holds. Equation 21 implies that cancers at the primary site must be treated at least as effectively as the metastases for remission to be maintained. For cancers at the primary site to shrink

$$\frac{d(q_3+q_4)}{dt} < 0 \qquad \text{(Eqn 22)}$$

must be true. Hence to reduce the primary tumor and maintain remission treatment must achieve this condition:

$$q_2 k_{32} + q_3 b_3 + q_4 b_4 < q_3 \mu_3 + q_4(\mu_4 + k_{54})_. \qquad \text{(Eqn 23)}$$

This states that mutations generating new cancer cells plus the proliferation of the primary cancer ($q_2 k_{32} + q_3 b_3 + q_4 b_4$) must be less than the loss of cancer cells due to cell death and migration to distant sites $q_3 \mu_3 + q_4(\mu_4 + k_{54})_.$ Treatments that increase the death rate of cancer cells, decrease their proliferation, or decrease the mutation rates leading to new cancers are effective means of achieving this condition.

Simplifying assumptions: Pancreatic cancer involves a cascade of genetic mutations in inherited (germ-line) and somatic cells, including KRAS2, p16/CDKN2A, TP53, SMAD4/DPC4, and other genes. These changes are accompanied by genomic and transcriptomic alterations that lead to invasion, metastases, cell cycle deregulation, and enhanced cancer cell survival (Maitra and Hruban 2008). As a point of departure and to make modeling tractable we now make simplifying assumptions regarding cell replication, death, mutation and DNA repair. These simplifying assumptions have to do with homogeneity in cell replication rates, cell senescence rates, DNA mutation rates, and DNA repair rates:

$$b' = b_1 = b_2;\; b'' = b_3 = b_4 \qquad \text{(Eqn 20)}$$

$$\mu' = \mu_1 = \mu_2;\; \mu'' = \mu_3 = \mu_4 \qquad \text{(Eqn 21)}$$

$$k' = k_{21} = k_{43} \qquad \text{(Eqn 22)}$$

$$k'' = k_{12} = k_{34} \qquad \text{(Eqn 23)}$$

Here b' is the replication of normal cells, and b'' is the replication of in situ pancreatic cancer cells. Senescence and death for normal and in situ pancreatic cancer cells are μ' and μ''. Mutation events for normal pancreatic cells initiating to the pre-cancerous condition are k'. DNA repair is k''.

Parameter	Description	Units	Estimate	Reference/note
b'	replication of cells in PanIN	Cell divisions per cell per unit time	1 replication / 2.3 cell days	(Yachida, Jones et al. 2010)
b''	replication of pancreatic cancer cells	Cell divisions per cell per unit time	1 replications / 2.3 days per cell division	(Yachida, Jones et al. 2010)
b_5	Replication of metastatic cancer cells	Cell divisions per cell per unit time	1 replication / 56 days	(Yachida, Jones et al. 2010)
μ'	Normal cell death	Deaths per cell per unit time	1 death / 2.3 cell days	#1

μ''	Death of pancreatic cancer cells	Deaths per cell per unit time	$\mu'' \quad < b'' - k_{54}$ ~0.75 * 1/2.3 deaths / cell day	#2
k'	Mutation/initiation to reversible pre-cancerous or cancerous condition	Mutations per cell per unit time	*k'*=6.957*10^{-3}	#3
k''	DNA repair to normal or earlier cancer state	Repair to prior cell state per cell per unit time	*k''*=6.887*10^{-3}	#4

Table 2. Model parameter estimates. Notes. #1: Cell death rate equals the birth rate in a normal pancreas. #2: For carcinogenesis the death rate of pancreatic cancer cells must be less than their death rate. As a point of departure we assume the death rate for cancerous cells is 0.75 the replication rate. #3: Using an assumed mutation rate per base pair per generation of 5 x 10^{-10}, Yachida and Jones et al. (2010) estimated the mutation rate per cell generation to be 0.016. We require the mutation rate per cell per unit time, and hence estimate k' to be 0.016 mutations / cell-replication * 1 cell-replication / 2.3 cell days. #4 We set the repair rate to be equal to 99% of the mutation rate per cell per unit time; k''=0.99 x 0.16/2.3

Model parameter estimates: What should the values of the pancreatic cell birth, death, mutation and repair rates, be? Table 2 summarizes the parameter estimates obtained from the current knowledge of the cellular biology of the pancreas and of pancreatic cancer, as described below.

Replication of normal cells, b': The parameter b' is the replication of normal pancreatic cancer cells (into compartment q_1) and of initiated precursors of cancer (compartment q_2). We assume a 2.3 day cell cycle time per mitosis/replication event as was used by Yachida et al. (2010).

Replication of pancreatic cancer cells, b'': Amikura et al (Amikura, Kobari et al. 1995) report an average cell doubling time for pancreatic cancers of 2.3 days, which was used by Yachida et al (Yachida, Jones et al. 2010) in their model of pancreatic tumor evolution. A characteristic of several of the mutations that typify pancreatic cancer is increased cell replication rate, and this value of 2.3 may be an underestimate.

Replication of metastatic pancreatic cancer cells, b_5: The median doubling time of pancreatic cancer metastases is estimated to be about 56 days (Amikura, Kobari et al. 1995). We assume the number of cell divisions per unit time for metastatic pancreatic cancer cells to be 1 replication per 56 cell-days.

Normal pancreatic cell death, μ': We assume the number of cells in the normal adult pancreas does not change appreciably over the life course, hence the per cell death rate must be about the same as the per cell replication rate of 1 death per 2.3 cell-days.

Pancreatic cancer cell death, μ'': For pancreatic cancer to occur the replication rate of pancreatic cancer cells must be greater than the death rate of pancreatic cancer cells, as per Equation 13. We thus assume pancreatic cell death is ¾ of the birth rate of pancreatic cancer cells, or ~0.75 * 1/2.3 with units deaths / cell-day.

Mutation to reversible pre-cancerous or cancerous condition, k': For our model k' is the mutation rate per cell per unit time. We recognize that cancer arises when damage to DNA is retained through cell replication (fixing a mutation), and that specific mutations can lead to rapid cell proliferation – a cancer. We further recognize that DNA is subject to *background mutations*, mutations that arise due to chance errors in DNA replication; and to environmental exposures that can be considered to be about the same everywhere (e.g. such as cosmic rays (Juckett 2009)). This has been called the *spontaneous mutation rate* and is estimated to be about 2.5x10-8 mutations per nucleotide site or ~175 mutations per diploid genome per generation (Nachman and Crowell 2000). Mutations also may arise due to exposures to mutagenic agents; including ionizing radiation (e.g. X-rays, gamma rays, and alpha particles), ultraviolet radiation, and radioactive decay of isotopes of the elements (e.g. Carbon 14) incorporated into DNA strands; DNA reactive chemicals and their metabolites (e.g. polycyclic aromatic hydrocarbons, aromatic amines, reactive oxygen species such as hydroxyl radicals, benzene, alkylating agents and others); DNA base analogs that can substitute for constituent bases when DNA is replicated; intercalcating agents that cause frame shift mutations by inserting themselves into a DNA strain; metals such as arsenic, cadmium and chromium; and biological agents including transposons and viruses that insert materials into DNA strands. Such *exogenous mutations* are in addition to the background mutations, and the total mutation rate is then the sum of the background mutation rate and the exogenous mutation rate. Most exogenous mutagens induce mutations such that higher exposures are associated with higher mutation rates (a dose-response effect). For modeling purposes we represent the observed mutation rate as the sum of a background mutation rate and the exogenous mutation rate (Equation 23).

$$k' \propto (m_b + m_e) \qquad \text{(Eqn 23)}$$

This states that mutations of pancreatic cell DNA that initiate or promote pancreatic cancer is proportional to the background mutation rate (m_b) plus the exogenous mutation rate (m_e). This model thus can capture changes in the underlying mutation rate attributable to specific exposures characterizing the exposome. We assume it is "proportional to" since only certain mutations will initiate and promote pancreatic cancers. Recall the *background mutation* is estimated to be about $2.5x10^{-8}$ mutations per nucleotide site (Nachman and Crowell 2000). We are interested in the mutation rate per cell replication event and that occur at those base pairs associated with pancreatic cancer genes. Yachida, Jones et al. (2010) sequenced 31.7×10^{6} base pairs in their study of the pancreatic cancer genome. Using an assumed mutation rate per base pair per generation of 5×10^{-10}, they estimated the mutation rate per cell generation to be 0.016. We require the mutation rate per cell per unit time, and hence estimate k' to be 0.016 mutations / cell-replication * 1 cell-replication / 2.3 cell days.

DNA repair to normal or earlier cancer state, k'': DNA repair to a normal, non-cancerous state must be frequent, otherwise cancer would be very common and occur early in life. For damaged cells, either in states q_2 or q_4, we require an estimate of the rate at which they are repaired to an earlier state (either healthy, q_1, or early cancer, q_3) per unit time. We thus set the repair rate to be equal to 99% of the mutation rate per cell per unit time; k''=0.99 x 0.16/2.3.

Initial Conditions: The initial conditions specify the compartment sizes before carcinogenesis begins and are given in Table 3.

State	Description	Initial value	Rationale and notes
q_1	Number of normal cells in the human pancreas	$7.5x10^{7}$	This is the number of cells in the normal pancreas (1)
q_2	Number of cells with an initiating but reversible mutation to pancreatic cancer	0	We assume no initiation events have transpired before time 0
q_3	Number of cells with an additional irreversible mutation to pancreatic cancer	0	We assume pancreatic cancer is absent before time 0
q_4	Number of cells in pancreas with metastatic capacity	0	We assume pancreatic cancer is absent before time 0
q_5	Number of cells in metastases at distant sites (e.g. lung, liver)	0	We assume there are no metastases before time 0.

Table 3. Model initial conditions. We assume only normal pancreatic cancer cells are present at the beginning. (1) The islets of Langerhans only make up about 1-2% of the total pancreas cells although the average human pancreas has about one million of them. Therefore we estimate 50-100 million cells in the human pancreas, with a mean of 75 million cells. From http://www.elp.manchester.ac.uk/pub_projects/2000/mnby7lc2/pancreas.htm

Residence times

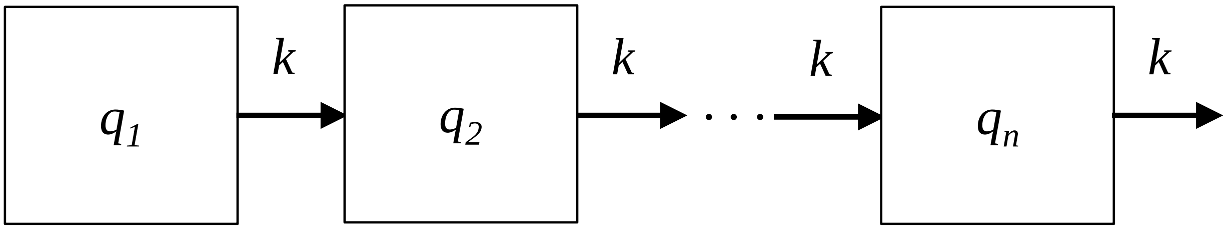


Figure 5. Outflow connected *n* compartment system useful for solving for the probability density function and cumulative distribution function of residence times.

For an outflow connected system without inflow and comprised of *n* compartments (Figure 5), the compartment sizes and density function of residence times, given an input of 1 unit at *t*=0 into compartment 1, are known to be Equations 24 and 25, respectively (Jacquez 2002).

$$q_n = \frac{k^{n-1}t^{n-1}}{(n-1)!}e^{-kt} \qquad \text{Eqn 24}$$

$$\theta(\rho, t) = \frac{k^{n}t^{n-1}}{(n-1)!}e^{-kt} \qquad \text{Eqn 25}$$

Here *ρ* specifies the proportions of particles in the *n* compartments such that the first compartment has size 1, and the others have size 0. This means the initial conditions specify that all particles at time 0 are in compartment 1. These equations may be applied to solve for the density function of residence times in the compartmental model of pancreatic cancer (Figure 4) in the subsystems PanIN, pancreatic cancer, and metastatic pancreatic cancer, given certain simplifying assumptions.

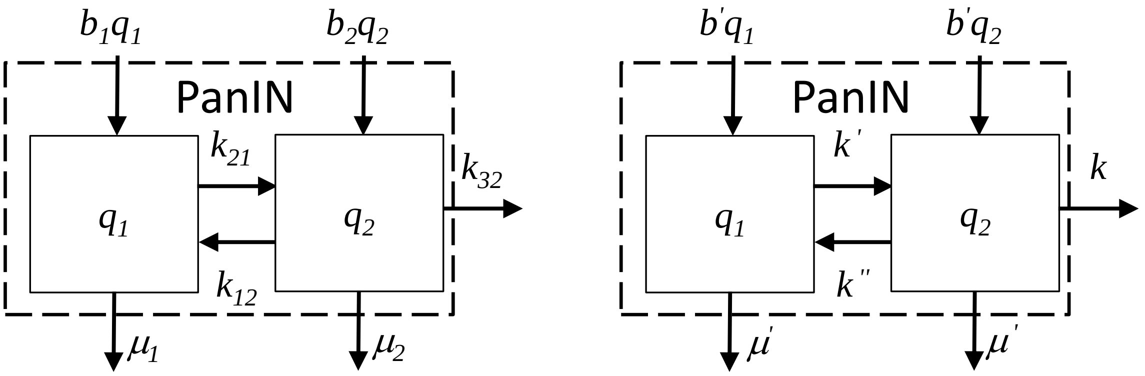


Figure 6. Representation of the PanIN model subsystem for estimation of distributions of residence times. Original subsystem model (left); simplified model used for calculation of distributions of residence times (right).

For PanIN we have the original model on the left of Figure 6 and its simplified form on the right, using the parameterization from Table 2 “Model parameter estimates”. We notice from our initial parameter estimates that $b' = \mu'$ and that $k' > k''$. For residence time calculations and for the time being assume $k = (k' - k'') = k_{32}$. We can then represent the PanIN subsystem model as shown in Figure 7.

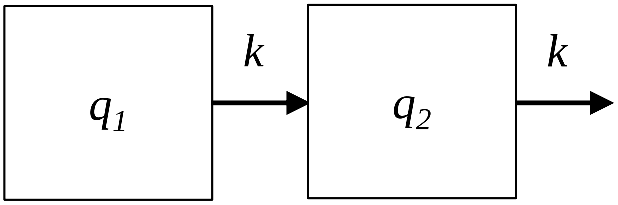


Figure 7. Simplified version of the PanIN subsystem model used for calculation of distributions of residence times.

The probability density function of residence times for an initial input of 1 unit into compartment q_1 at *t*=0 (all healthy cells) is then

$$\theta(\rho, t) = \frac{k^2 t}{1!} e^{-kt} = k^2 t e^{-kt} \qquad \text{Eqn 26}$$

This is the probability density function of the Erlang distribution for parameters *n* and *k*. For the parameter values k=0.00007, t=4,270.5 days, we obtain $\theta(\rho, t) = 1.55 \cdot 10^{-5}$; this is the likelihood (probability density) of a residence time of *t*=4,270.5 days in the PanIN stage. Examining the Erlang distribution as a function of residence time (Figure 8), we find the cumulative distribution at residence time of 4,270.5 days to be 0.0367. The mean residence time of 11.7 years for cells and the descendants of cells in the stage PanIN before exiting to pancreatic cancer is a relatively rare event.

Solved, another way, we can ask how long it would take for an average healthy pancreatic cell or its descendants to exit the PanIN compartment by progressing to pancreatic cancer. This mean residence time is 28,571.43 days, or 78.28 years. According to this model, that is the time required, on average, for a healthy pancreatic cell to be promoted to an irreversible cancer (e.g. enter compartment q_3). This is consistent with the observation that pancreatic cancers occur later in life.

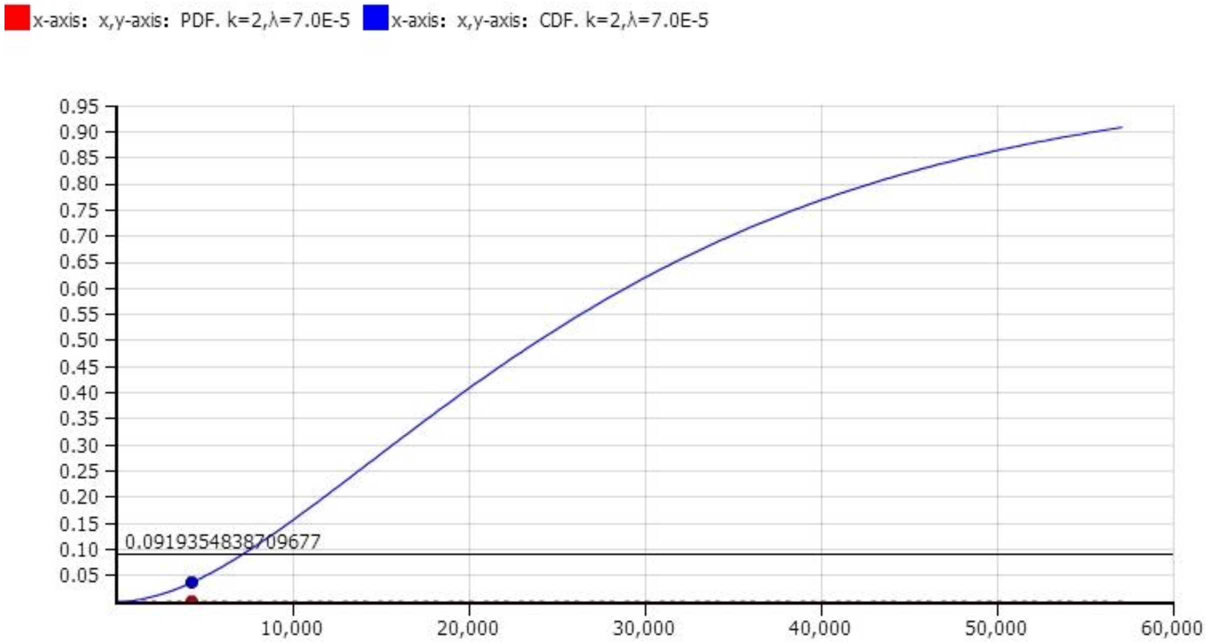


Figure 8. Distribution of residence times in PanIN. X-axis, days; y-axis, cumulative probability. At t= 4,270.5 days the cdf=0.0367 and the pdf=$1.55 \cdot 10^{-5}$. The mean time for progression from PanIN to pancreatic cancer of 4,270.5 days observed by Yachida, Jones et al. (2010) is thus a relatively rare event.

For the pre-metastatic cancer stages q_3 and q_4 we are concerned with the residence times for cells entering q_3 and exiting to metastases q_5. We employ the simplifying assumption that the cell replication rate b'' is approximately equal to the cell death rate μ''. We also assume the rate of progression to metastases is approximately equal to the net forward rate k'-k''. We then again impose the condition $k=k_{54}$=k'-k''=0.00007 events per cell day. This yields a similar system of equations for the pdf and cdf of residence times as for PanIN. Using the observed time in pancreatic stage T_2=6.8 years or 2,482 days we obtain:

$$\theta(\rho, T_2) = 1.02 \cdot 10^{-5}$$

$$\mathrm{CDF}(T_2) = 0.0135.$$

Under this model the probability of a healthy cancer cell progressing through PanIN and becoming a metastatic cancer cell in less than T_1=4,270.5 days plus T_2=2,482 days is CDF(T_1) x CDF(T_2)=4.9 x 10^{-4}.

This leaves the residence times for metastatic pancreatic cancer, compartment q_5. For metastatic cancers in particular replication must exceed cell death, $b_5>\mu_5$. However, to solve for the residence times we consider an initial pulse of cells have founded the metastases; $q_5>0$. Further, we note the replication of metastatic cells is 1 replication/56 days, much slower than healthy cells. For the metastases to grow cell death must be less than replication, and we impose $\mu_5 = 0.75b_5$. What then are the density function of residence times and cumulative distribution function of T_3=2.7 years? For the cells comprising the founding clone, what is the mean residence time μ_5, and what is the probability of having died at T_3. For this one compartment system:

$$\theta(\rho, T_3) = ke^{-kt} = 4.42 \cdot 10^{-7}$$

$$\text{CDF}(T_3) \approx 0.0.$$

None of the cells in the original founding clone are typically still present in the metastases at diagnosis, and the mean residence time of the metastatic cancer cells is 74.67 days before cell death.

Modeling cancer stages

We now present a model of pancreatic cancer stages. Here, the unit of observation is the cancer patient, and we observe counts of patients in early and late stage pancreatic cancer, before and after diagnosis (Figure 9). Counts of people in early and late stages prior to diagnosis are represented by q_6 and q_7. Compartment q_8 is comprised of patients who have been diagnosed, either in early or late stage cancer. The flow of the number of health individuals entering early stage cancer is F_{60}. The rate of progression from early to late stage cancer is k_{76}. Diagnosis events from early and late stage are given by the rates k_{86} and k_{87}. Death from the compartments is represented by μ_6, μ_7, and μ_8.

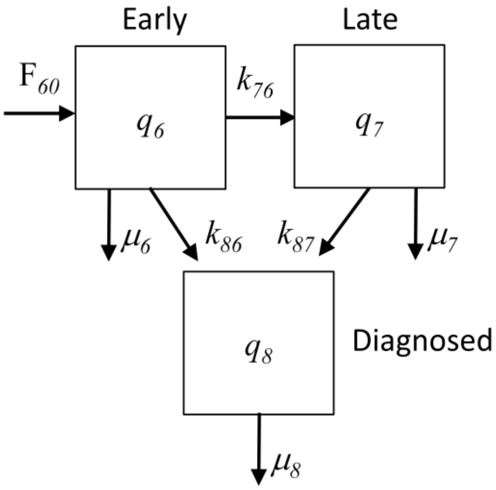


Figure 9. Stage-based model of pancreatic cancer. Here the compartment sizes are number of patients with early (q_6) and late stage cancers (q_7) prior to diagnosis, and the number diagnosed (q_8).

System Equations: The system equations for this stage-based model of pancreatic cancer are

$$\frac{dq_6}{dt} = \mathrm{F}_{60} - q_6(k_{76} + k_{86} + \mu_6) \qquad \text{Eqn 27}$$

$$\frac{dq_7}{dt} = q_6 k_{76} - q_7(k_{87} + \mu_7)$$

$$\frac{dq_8}{dt} = q_6 k_{86} + q_7 k_{87} - q_8 \mu_8.$$

Equilibrium: Equilibrium occurs under the following conditions.

$$0 = \frac{dq_6}{dt} = \frac{dq_7}{dt} = \frac{dq_8}{dt} \qquad \text{Eqn 28}$$

$$q_6 = \frac{\mathrm{F}_{60}}{(k_{76}+k_{86}+\mu_6)}$$

$$q_7 = \frac{q_6 k_{76}}{(k_{87}+\mu_7)}$$

$$q_8 = \frac{q_6 k_{86}+q_7 k_{87}}{\mu_8}$$

Estimation: The number of incident early and late stage cancers (compartment q_8) are directly observable in most of the states comprising the United States from cancer registry data. The flows $q_6 k_{86}$ and $q_7 k_{87}$ are observable as the number in a defined time period of early and late stage diagnoses. The mortality rate $q_8 \mu_8$ is directly observable as the number of diagnosed pancreatic patients who die in a defined time period. The quantities $q_6 \mu_6$ and $q_7 \mu_7$ are the number of deaths of people with early and late stage, but undiagnosed, pancreatic cancer. The estimation of parameter values and the number of yet to be diagnosed cancer cases will be demonstrated below in the example of pancreatic cancer in Southeast Michigan.

Carcinogenesis and stage-based model of pancreatic cancer

The carcinogenesis model deals with pancreatic cancer cells in histological and genetic states as compartment members, whereas the stage-based model uses individuals and the stage of their pancreatic cancer to define compartment membership. The model of carcinogenesis informs the stage model through an equivalence of residence times and model states (Table 4).

Stage model compartment	Carcinogenesis model compartment	Description	American Joint Committee on Cancer (AJCC) staging	Residence time
q_6	q_3 , q_4	Insitu, local, not diagnosed	Insitu: AJCC Tis, N0, M0 Local: AJCC IA, IB, N0,M0	T_2 : 6.8 ± 3.4 yrs
q_7	q_5	Regional, distant, not	Regional: AJCC IIA, IIB Distant: AJCC IV	T_3 : 2.7 ± 1.2 yrs

		diagnosed		
q_8	-	Diagnosed pancreatic cancer	May be *in situ*, local, regional, or distant; in most cases pancreatic cancer is diagnosed at an advanced stage	T_4 : 0.5 ± 0.25 yrs (2011 five-year survival rate < 6% and average life expectancy after diagnosis is 3 to 9 months.

Table 4. Equivalence of model states and residence times between carcinogenesis- and stage-based models, using diagnostic pancreatic cancer staging according to the American Cancer Society American Joint Committee on Cancer (Edge, Byrd et al. 2010).

Model stage	**AJCC stage**	**Prognostic Groups**			**Diagnosed**
Early (q_6)	Stage 0	Tis	N0	M0	N
	Stage IA	T1	N0	M0	N
	Stage IB	T2	N0	M0	N
Late (q_7)	Stage IIA	T3	N0	M0	N
	Stage IIB	T1-3	N1	M0	N
	Stage III	T4	Any N	M0	N
	Stage IV	Any T	Any N	M1	N
Diagnosed (q_8)	Any stage	Any T	Any N	Any M	Y

Table 5. Correspondence of modeled cancer stages to anatomic stages from the American Joint Committee on Cancer. Primary Tumor (T) coding is Tis: Carcinoma in situ; T1: Tumor limited to pancreas, 2cm or less in greatest dimension; T2: Tumor limited to pancreas, more than 2cm in greatest dimension;T3: Tumor extends beyond the pancreas but without involvement of the celiac axis or the superior mesenteric artery; T4: Tumor involves the celiac axis or the superior mesenteric artery (unresectable primary tumor). Regional lymph nodes (N) coding is N0: No regional lymph node metastasis; N1: Regional lymph node metastasis. Distant metastasis (M) coding is M0: No distant metastasis; M1: Distant metastasis.

Application: Pancreatic cancer in Southeast Michigan

To demonstrate the approach we apply the stage-based model to incident pancreatic cancer cases in southeastern Michigan. We employ the four steps illustrated in Figure 1, customized to this specific application.

Step 1: Develop the minimally sufficient biologically reasonable systems model

Step 2: Solve for residence times, compartment sizes and flows

Step 3: Map the data to identify local populations with excess risk

Step 4: Interpret the results

Background and Data: An analysis of pancreatic cancer mortality in white males in Michigan counties in two time periods from 1950-70 and 1970-95 found statistically significant clusters that persisted in Wayne county in both time periods and that expanded to include adjacent Macomb county in 1970-95 (Jacquez 2009). This finding was confirmed using more recent incidence and mortality data from the Surveillance Epidemiology and End Results program, SEER (Ries, Harkens et al. 2007). 17 registry/areas are included in the SEER program, including Atlanta, rural Georgia, California (Bay Area, San Francisco-Oakland, San Jose-Monterey, Los Angeles and Greater California), Connecticut, Hawaii, Iowa, Kentucky, Louisiana, New Jersey, New Mexico, Seattle-Puget Sound, Utah and Detroit. In 2000-2004 Detroit had the highest age-adjusted incidence rate for white males at 15.0 cases per 100,000 out of all of the 17 registry/areas, and the second highest mortality rate at 12.9 deaths per 100,000. In contrast, the SEER-wide averages for white males in this period were 12.8 incident cases and 12.0 deaths per 100,000. Notice the incidence is nearly equal to the deaths for the SEER-wide averages (12.8 vs 12.0), but the incident cases in Detroit exceed the mortality rate by a larger difference (15.0 vs. 12.9). This is consistent with the observation that pancreatic cancer incidence in Detroit is increasing, and that the Detroit system may not be in equilibrium. In terms of our compartmental model, it appears the flows in (F_{06}) exceed the flows out due to mortality ($q_8\mu_8$). Notably, the Detroit registry pancreatic cancer mortality for white males in 2000-2004 increased on average 0.9% per year (Calculated by SEER*Stat from the National Vital Statistics System public use data file). The population covered by the Detroit registry in this period was 1,365,315 white males. The finding of excess pancreatic mortality with increasing incidence was thus independently confirmed by data from SEER and found to persist from 1950 through 2004 (Jacquez 2009).

As a follow-up to this study and to explore the hypotheses that H1: cancer incidence is not in equilibrium with cancer mortality, and that H2: pancreatic cancer incidence is increasing, we obtained annual incidence data from the Michigan Cancer Registy for the period 1985-2005. The Michigan Cancer Registry is a gold-standard registry whose completeness and accuracy is certified on an annual basis. The variable descriptions and coding are in Table 6.

Variable	Description	Coding
sfnum	Report ID	
tract2000	Census 2000 Tract Code	
block2000	Census 2000 Block Code	
tract	Census 1990 Tract Code	
block	Census 1990 Block Code	
longitude	Longitude	
latitude	Latitude	
mappedmcd	Fips code for mapped minor civil division	
mappedcty	Fips code for mapped county	

fipscty	Fips code for reported county	
zip	Reported zip at diagnosis	
age_diag	Age at diagnosis	
sex	Sex	1 = Male 2 = Female 3= Transgender 9 = Unknown
seerrace1	Race code	01 = white 02 = black 03 = American Indian 04-32,96,97 = Asian 90 = Multiracial 98 = Other 99 = Unknown
icdoii	Primary site code (ICD-O III)	C250 = Head of pancreas; C251 = Body of pancreas; C252 = Duct; C254 = Islets of Langerhans; C257 = Other Pancreas; C258 = Overlapping regions; C259 = Pancreas NOS
pctcb	Morphology and tumor behavior	See ICDO III
stagedis	Stage at diagnosis	01=insitu; 02=local; 03=regional; 04=distant; 05= unknown
yeardiag	Year diagnosed	
reg_num	Patient ID	
reg_seq	Primary tumor sequence	

Table 6. Variable description and coding for incident pancreatic cancer cases in the Detroit metropolitan area, 1985-2005.

Data cleaning and processing: The geocoding budget and numbers of observations are as follows (Table 7). A total of 11,068 pancreatic cancer cases were diagnosed between January 1, 1985 through December 31, 2005. Of these, 192 addresses of place of residence at diagnosis failed to geocode, leaving 10,876 cases with known places of residence at diagnosis. Stage at diagnosis (insitu, local, regional, distant and unknown) was recorded as unknown for 2,250 of these, leaving 8,826 cases with known place of residence and known stage at diagnosis. The head of pancreas and pancreas not otherwise specificed were the most frequent primary sites, with 4,496 and 1,621 respectively. Males accounted for 4,202 cases and females 4,424. By race, 6,356 cases were whites, 2,192 blacks, and the balance American Indian (8 cases), Asian (61) and other or unknown groups (9).

Desciption	Count	Subtotal		Note
Total number of cases	11,068	11,068		
Failed to geocode	192	10,876		
Stage at diagnosis unknown	2250	8,626		656 cases with stage unknown in 2004 and 2005
Stage at diagnosis				
Insitu	25			
local	947			
regional	2558			
distant	5096	8626	8626	
Primary site				
Head of pancreas	4496			Code C250
Body of pancreas	733			Code C251
Duct	891			Code C252
	64			Code C253

	Islets of Langerhans	8		Code C254
	Other pancreas	42		Code C257
	Overlapping regions	771		Code C258
	Pancreas NOS	1621	8626	Code C259
Sex				
	Male	4202		Code 1
	Female	4424		Code 2
	Transgender	0		Code 3
	Unknown	0	8626	Code 9
Race				
	White	6356		Code 01
	Black	2192		Code 02
	American Indian	8		Code 03
	Asian	61		Code 04-32, 96, 97
	Multiracial	0		Code 90
	Other	5		Code 98
	Unknown	4	8626	Code 99

Table 7. Pancreatic cancer data budget.

Stage ascertainment by case was under-recorded for 2004 and 2005 (Figure 10), and for that reason these years were excluded from certain analyses. When staging was not required for an analysis we retained the data for years 2004 and 2005 since the total number of incident cases appeared consistent with earlier years. For estimating inflows to compartment q_8 we used data from 2003 and earlier. For the case-clustering analysis of early versus late-stage diagnosis we used data from 2003 and earlier, since stage ascertainment was incomplete for 2004-2005. The frequency distributions by year diagnoses and by age at diagnosis are shown in Figure 11.

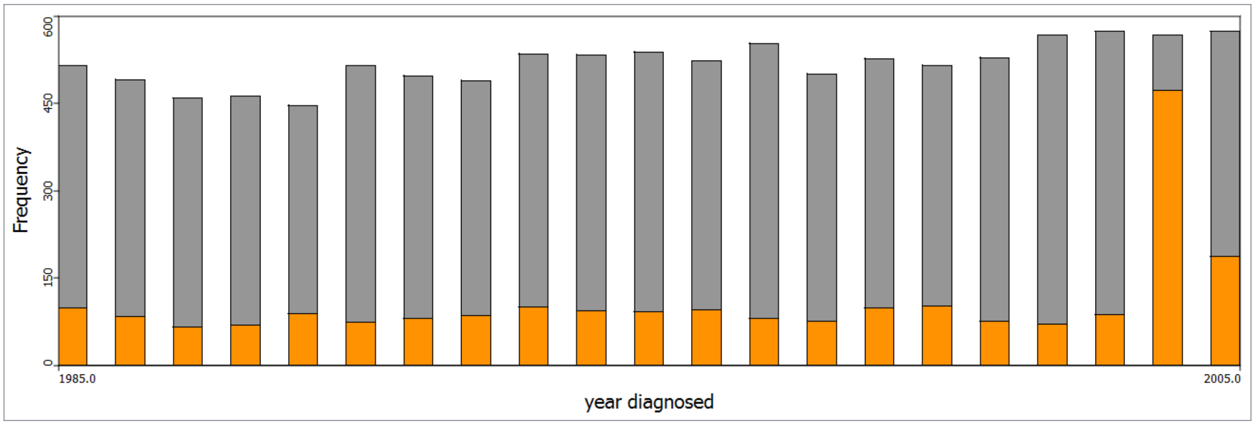


Figure 10. Number of incident cases by year. Gold color indicates observations with stage unknown. Created using the SpaceStat software.

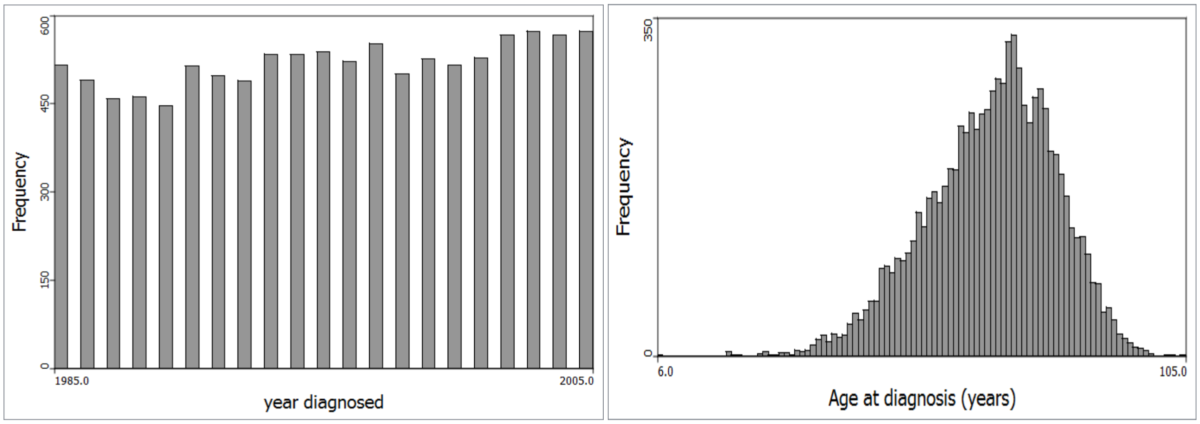


Figure 11. Frequency distributions of year diagnosed (left) and age at diagnosis (right). Mean age at diagnosis was 68.7 years. Created using the SpaceStat software.

We now proceed through each of the 4 steps needed to construct and apply the model.

Step 1: Describe the model. We employ the model of pancreatic cancer stages in Figure 9, system equations in Eqn 27.

Step 2: Estimate flows, compartment sizes and residence times. The quantities directly observable are the incident flows into compartment 8 from early and late stage but not diagnosed cancers. We use the data for all incident pancreatic cases, whether they geocoded or not, and whether the stage at diagnosis was known or unknown. Let o_e be the total number of cases from 1985 through 2005 observed in the early stage, o_L be the number late stage, and o_u be the number in unknown stage. Y is the number of years over which the observations accrued (21 years). We can then estimate the flows into compartment q_8 for early and late stage cancers as

$$\widehat{q_6 k_{86}} = \frac{(o_e + (^{o_e}/_{(o_e+o_L)})o_u)}{Y} = \frac{1246.8}{21} = 59.37$$

$$\widehat{q_7 k_{87}} = \frac{(o_L + (^{o_L}/_{(o_e+o_L)})o_u)}{Y} = \frac{1246.8}{21} = 467.67.$$

The units on these are number of cases in the given stage diagnosed per year. According to the American Cancer Society, for all stages of pancreatic cancer combined, the one-year relative survival rate is 20%, and the five-year rate is 4%. If we assume $\mu_8 = 0.8$ deaths/diagnosed case-year, and assuming the equilibrium condition in equation 28, we can then estimate the size of compartment q_8 as

$$\widehat{q_8} = \frac{q_6 k_{86} + q_7 k_{87}}{\mu_8} = 658.81.$$

This is the average number of diagnosed and surviving (not yet deceased) pancreatic cancer cases.

For the late stage but not diagnosed cases in compartment q$_7$ we note that at equilibrium

$$q_7(k_{87} + \mu_7) = q_6 k_{76}.$$

The rate μ_7 is deaths of late-stage but not diagnosed cases that are not diagnosed after the death event, and thus do not flow into compartment q$_8$ (they would have to be diagnosed to enter this compartment). We impose $\mu_7 = 0$, under the assumption that all of the late-stage pancreatic cancer cases are diagnosed (this assumption can be relaxed but seems reasonable since late stage pancreatic cancers are by definition advanced and metastatic). Hence deaths for late stage but not yet diagnosed cases are diagnosed after they decease. This then yields

$$q_7 k_{87} = q_6 k_{76} = 467.67.$$

Again, the units here are number of cases per year. Since $q_7 k_{87} = q_6 k_{76}$ and $q_6 k_{86} = 59.37$,

$$\widehat{q_7} = 59.37 \frac{k_{76}}{(k_{86}+k_{87})}.$$

The age-adjusted annual mortality rate from all causes in Michigan in 2010 was 764.2 deaths per 100,000 (Miniño and Murphy 2012), and has decreased from 1,027.10 deaths per 100,000 in 1985 (MDCH 2011). We therefore estimated the background mortality rate from 1985-2005 as the sum of the age-adjusted death rates for all races and sexes divided by the number of years being considered, yielding a 21 year average of 924.05 deaths per 100,000. We set person-specific annual death rate $\mu_6 = 0.00924$ and using the equilibrium condition for compartment q$_6$ obtain

$$q_6 = \frac{\mathrm{F}_{60}}{(k_{76}+k_{86}+\mu_6)}$$

$$\mathrm{F}_{60} = q_6(k_{76} + k_{86} + \mu_6)$$

$$\widehat{F_{60}} = 467.67 + 59.37 + q_6\mu_6$$

$$\widehat{F_{60}} = 467.67 + 59.37 + \frac{59.37}{k_{86}}\mu_6 = 527.04 + \frac{0.5486}{k_{86}}, \text{ and finally}$$

$$\widehat{q_6} = \frac{F_{60} - 527.04}{0.009241}.$$

Earlier we demonstrated an equivalence between residence times in early and late stage cancer stages (q$_6$ and q$_7$) and residence times in the carcinogenetic model of PanIN and its sequelae. Then the residence time in q$_6$ is T$_2$, and in q$_7$ it is T$_3$. It still remains to solve for the residence time in q$_8$, T$_4$. Consider a pulse of newly diagnosed cases entering q_8 either from q$_6$ (diagnosed in early stage) or q_7 (diagnosed in late stage). Recall the median survival after diagnosis is 6 months, and that the one year survival rate is about 20%. Expressing time in days, we wish to fit the Erlang distribution such that

CDF(182.5 days)=0.5, and CDF(365 days)=0.8.

We solved this using the formulation for a one compartment system with μ_8 as the exit. At a daily mortality rate of $\mu_8 = 0.0038$ we find

CDF(182.5 days)=0.5002, and CDF(365 days)=0.7502.

Put another way, this states that for a pulse of cases diagnosed on the same day, about 50% will be alive after 182.5 days, and about 25% will be alive after 1 year. This indicates our fairly simple model of compartment q_8 is reasonably complete, at least in terms of its ability to represent observed 6 months and 1 year survival statistics.

Now that we have estimated μ_8 we use the relationship

$$q_8 = \frac{\widehat{q_6 k_{86}} + \widehat{q_6 k_{86}}}{\mu_8}$$

to solve for the size of compartment 8 yielding 379.98.

$$\widehat{q_8} = 379.98.$$

This is the estimate of the average number of diagnosed but not deceased pancreatic cancer cases in the study area.

Earlier we solved for q_6 and q_7 using observed quantities such as incident early and late stage pancreatic cancer case diagnoses. It is interesting to note for q_6 that an alternative solution is to use the observed residence time in early stage, T_2, to then solve for q_6. This provides a validation of the estimate.

Define *k'* to be the sum of the outflow coefficients from compartment q_6

$$k' = k_{76} + k_{86} + \mu_6.$$

Notice we can now estimate k' using the methods developed earlier for the residence time of the Erlang distribution. Specifically, solve for k' for a 1 compartment system such that the mean residence time is T_2. This yields an estimate of k'

$$\widehat{k'} = 0.00028,$$

which is the per case daily rate of exit from early stage but not-yet diagnosed pancreatic cancer, attributable to background mortality, progression to advanced cancer, and diagnosis.

Multiplying by q_6 and using hat notation to indicate values we can estimate from the observed data yields

$$q_6\widehat{k'} = \widehat{q_6 k_{76}} + \widehat{q_6 k_{86}} + q_6\widehat{\mu_6}.$$

We now divide through by q_6, rearrange and have an estimator for q_6 as

$$\widehat{q_6} = \frac{\widehat{q_6 k_{76}} + \widehat{q_6 k_{86}}}{\widehat{k'} - \widehat{\mu_6}}.$$

Using the values obtained earlier yields (written using annual time orientation)

$$\widehat{q_6} = \frac{467.67+59.39}{0.1022-0.00924} = 5669.75.$$

This is the estimated number of early stage cancers that are in the population but not yet diagnosed. We now use a similar approach to solve for the estimated number of undiagnosed advanced cancers, q_7. Recall at equilibrium the inflows into this compartment must equal the outflows, hence $q_7 k_{87} = q_6 k_{76}$. This is estimated as the observed number of diagnosed advanced stage cancers, and for our system $\widehat{q_7 k_{87}} = \widehat{q_6 k_{76}} = 467.67$. Solving for $\widehat{q_7}$

$$\widehat{q_7} = \frac{\widehat{q_6 k_{76}}}{\widehat{k_{87}}}.$$

Again, we estimate $\widehat{k_{87}}$ using the Erlang distribution of residence times. Specifically, solve for $\widehat{k_{87}}$ for a 1 compartment system such that the mean residence time is T_3. This gives an estimate of $\widehat{k_{87}}$

$$\widehat{k_{87}} = 0.000703.$$

This is the estimated daily diagnosis rate per person with advanced stage pancreatic cancer. Using annual values we now estimate

$$\widehat{q_7} = \frac{\widehat{q_6 k_{76}}}{\widehat{k_{87}}} = \frac{467.67}{0.257} = 1{,}822.6.$$

This is the number of individuals with undiagnosed advanced-stage pancreatic cancer.

Step 3: **Map undiagnosed early and late stage pancreatic cancers; assess clustering of advanced stage cancers in age 55 and younger**

We now estimated the numbers of undiagnosed cancers in total, and for both early and late stages. We define the estimated relative risks for total undiagnosed (TRR), early stage undiagnosed (ERR), and late stage undiagnosed (LRR) as the proportion of cases in each of these groups (total undiagnosed, early stage undiagnosed, late stage undiagnosed) relative to the total number of diagnosed cases,

$$\widehat{TRR} = \frac{\widehat{q_6}+\widehat{q_7}}{\widehat{q_8}} = \frac{5{,}669.75+1{,}822.6}{379.6} = 19.72$$

$$\widehat{ERR} = \frac{\widehat{q_6}}{\widehat{q_8}} = \frac{5{,}669.75}{379.6} = 14.92$$

$$L\widehat{RR} = \frac{\widehat{q_7}}{\widehat{q_8}} = \frac{1{,}822.6}{379.6} = 4.80.$$

We find the total number of silent (yet to be diagnosed) case is more than 19 times the number diagnosed. Hence, for each case that is diagnosed we estimate there are 19 pancreatic cancer cases in the at-risk population that have yet to be diagnosed. Of these, almost 15 are in the early stages of pancreatic cancer, and nearly 5 are advanced. *This means that application of a screen for early stage pancreatic cancer could dramatically reduce pancreatic cancer mortality, since such a large proportion of undiagnosed cases are in the early stages.*

The geographic distribution of pancreatic cancer cases is shown in Figure 12, displaying the stage at diagnosis. The map and the frequency distribution of the estimated count of silent (yet to be diagnosed) cases are in figure 13.

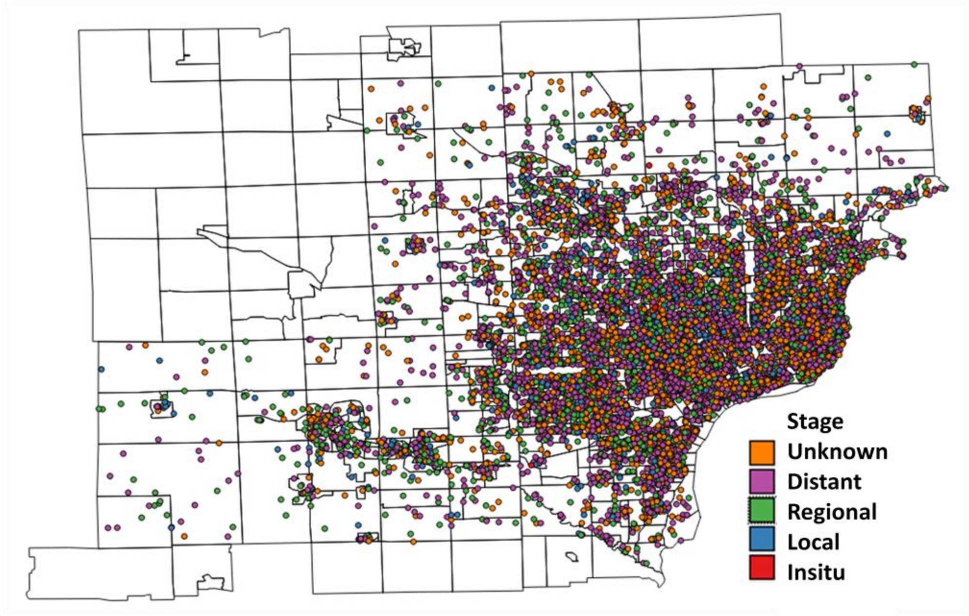


Figure 12. Locations of incident cases of pancreatic cancer in southeast Michigan, 1985-2005. Created using the SpaceStat software.

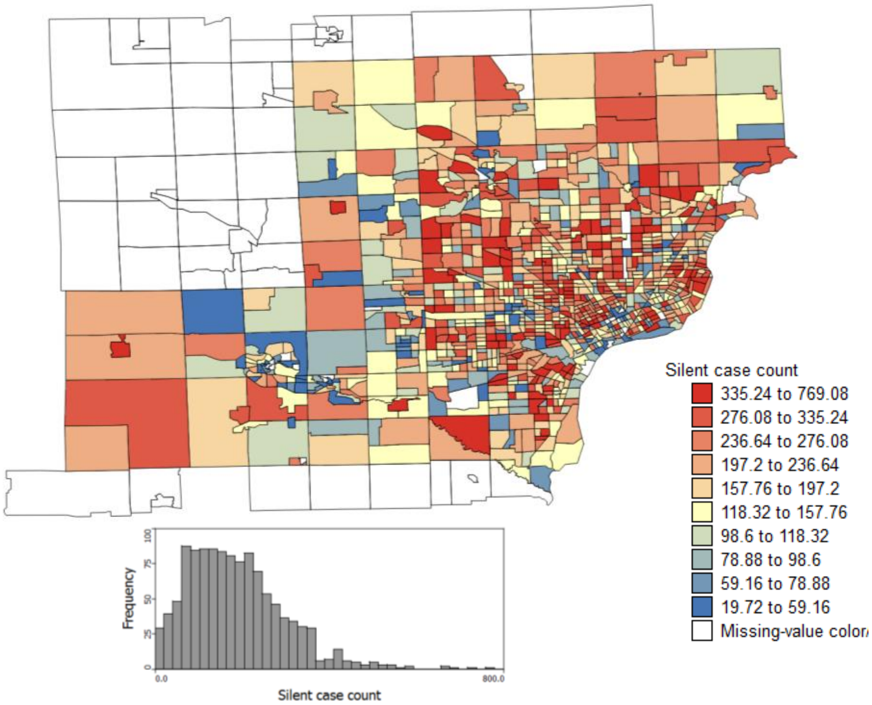


Figure 13. Map and frequency histogram of silent (yet to be diagnosed) pancreatic cancer cases in the greater Detroit metropolitan area. Created using the SpaceStat software.

Recall the results from the SEER program show pancreatic incidence in the study area increasing about 0.9%/year. Further, inspection of temporal trend in late-stage diagnoses in cases 55 years of age and younger suggests such diagnoses are increasing (Figure 14). This might be consistent with a change in the timing of cancer onset or aggressiveness over the life course.

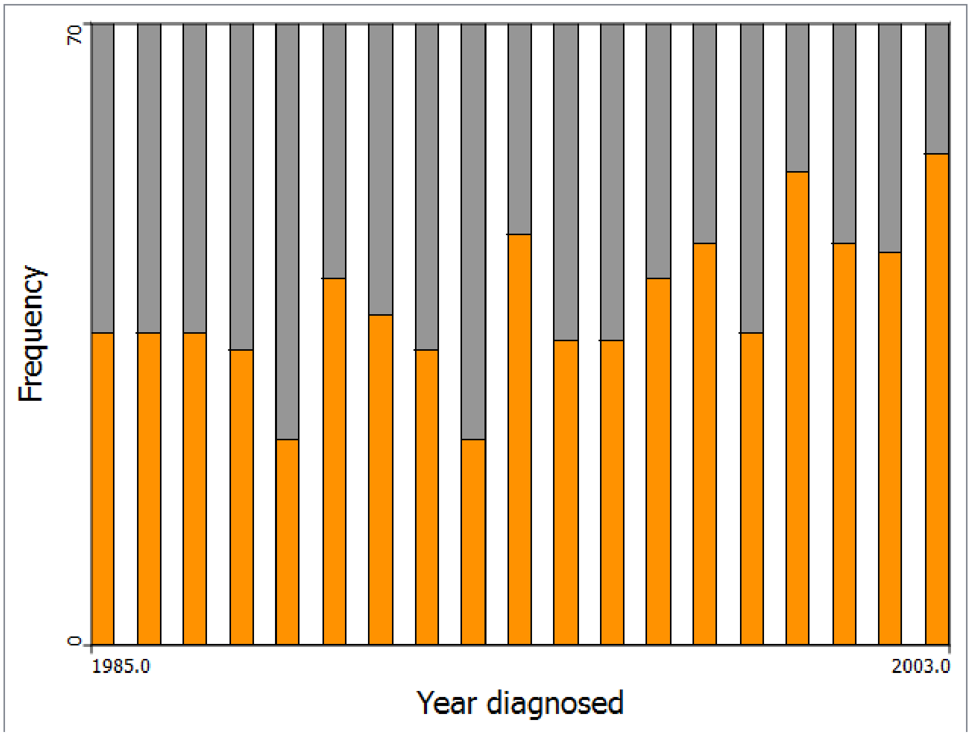


Figure 14. Pancreatic cancer incidence by year. Highlighted bars indicate late-stage diagnosis in patients 55 years or younger at time of diagnosis.

Is there increased risk of late-stage diagnosis among cases 55 years of age or younger at time of diagnosis? To address this question we calculate a relative risk and confidence interval of late stage diagnosis in the 55 or younger age group as

$$RR = \frac{a/(a+b)}{c/(c+d)} \qquad CI(RR) = e^{\log_e RR \pm 1.96\sqrt{(\frac{b}{a(a+b)}+\frac{d}{c(c+d)})}}.$$

Here the quantities a,b,c and d are defined:

Age at diagnosis	**Stage at diagnosis**	
	Advanced	**Early**
55 and younger	a=6495	b=850
56 and older	c=7654	d=8626

The values shown are for the incident cases from 1985 through 2005. For these data the RR of being 55 or younger and late stage at diagnosis is 1.023, with 95% CI of 1.0033 to 1.0434. We thus find a small but statistically significant relative risk of being 55 or younger and late stage at diagnosis when we consider years 1985-2005 combined.

Does this relative risk increase through time? When we repeat this analysis by year the relative risk is well within the 95% confidence intervals that contain RR=1 (Figure 15).

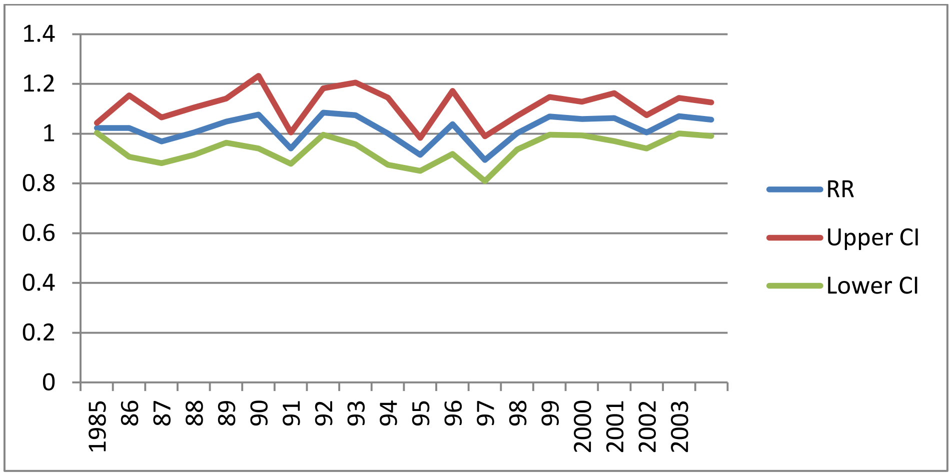


Figure 15. Relative risk of being less than 56 years of age and diagnosed with late stage pancreatic cancer through time.

Step 4: Interpret results. This analysis of pancreatic cancer in Michigan demonstrated several important findings. First, the burden of undiagnosed pancreatic cancers in this population is large, approximately 19 times the number of diagnosed pancreatic cancer cases. This indicates a screening test for detecting early stage pancreatic cancer, coupled with appropriate surgical and chemotherapeutic intervention, has the potential for dramatically reducing pancreatic cancer mortality in this population. Second, we estimate there are 1,822.6 undiagnosed advanced stage pancreatic cancer cases in this population. Some of these will be diagnosed prior to death, others will be diagnosed post-mortem. The demand on treatment resources in the last months of advanced pancreatic cancer are substantial and this estimate can be used to predict the demand for health care resources and to predict care expenses. Third, there is some evidence that pancreatic cancer risk in this population is increasing. The SEER results place pancreatic cancer incidence and mortality among the highest in all SEER registries, and the change in the annual incidence rate is about 0.9% per year. We found a small but statistically significant relative risk of being 55 or younger and late stage at diagnosis when we consider years 1985-2005 combined. This suggests the possible action of a risk factor for pancreatic cancer that is impacting younger members of this population. However, demographic factors such as differential migration cannot be excluded without further analysis, and in any event the relative risk is not large. Finally, the map of silent (yet to be diagnosed pancreatic cancer cases) directly supports targeting of diagnostic services, planning for upcoming in-home health care needs, and the geographic allocation of future screening programs to local populations with high demand.

Discussion

This research addresses several important topics in the modeling of space-time systems, cancer biology, and cancer surveillance. It has developed, to our knowledge, the first comprehensive modeling approach that estimates cancer latency, couples carcinogenesis and stage models, and that represents and links processes at the genomic level (e.g. mutation events, cascades of genetic changes that lead to cancer), cellular level (e.g. cell replication and death, DNA repair), organ level (e.g. carcinogenesis insitu and metastases to distant organs), individual level (e.g. cancer staging in the individual, progression of individuals through cancer stages), to the population level (e.g. geographic distributions of local populations in cancer stages, estimates of the predicted geographic distributions of undiagnosed cancers). Specific benefits of the approach are as follows.

1. It is process-based, capturing the known biological characteristics and mechanics of the cancer process at multiple scales (e.g. genomic to population).

2. It provides estimates of cancer latency, based on the known genetic and histologic characteristics of the cancer.

3. The latency estimates are integrated into spatio-temporal models of cancer incidence, mortality and future cancer burden.

4. The impacts of cancer screening and diagnosis may be represented in the model by diagnosis events through which individuals progress from undiagnosed (silent) to diagnosed stages. This provides a ready mechanism for modeling improvements in pancreatic cancer screening.

5. It predicts the burden of silent cancer (yet to be diagnosed), and geographically allocates these silent cancers by cancer stages into local geographic populations. This provides the quantitative support necessary for forecasting the future cancer burden.

6. The model is readily updatable. As knowledge of cancer genomics becomes more detailed it may be incorporated into the carcinogenesis model by updating the cascade of events that underpin the flows and stages.

7. It provides a quantitative basis for evaluating alternative treatments and for predicting treatment efficacy, provided by the equations and conditions for cancer progression, metastasis and remission.

Several caveats apply. Assumptions implicit in compartmental models include the homogeneity assumption, which states the particles being modeled behave in an identical fashion. This means the pancreatic cancer cells in each compartment of the carcinogenesis model, and the cases in each compartment of the stage model, are assumed to behave in identical fashions to other particles in the compartment under consideration. This assumption is typical of all modeling approaches (since all models involve simplification and abstraction), and can be relaxed when needed by adding additional compartments to capture important aspects of heterogeneity. A second assumption of the compartmental approach is that of instantaneous and complete mixing. This assures that the kinetics (e.g. necessary for calculation of transit and residence times) of each particle may be calculated without consideration of when they entered the compartment or the order in which they entered. A final assumption is that the particles in the compartments (e.g. cells or cases) are sub-dividable, such that a flow of 0.3 cells is possible. This clearly is incorrect for cells and people, but in practice is not a bad assumption when the number of particles in any given compartment is large.

The parameter estimates for cell replication, cell death, DNA mutation rates, repair rates, metastases initiation, and cancer promotion and so on where extracted from the literature by the author, who is not a trained oncologist or cell biologist. While the author believes the broad strokes are largely correct, the parameter estimates in this paper are initial ones only, and the specific results may need to be revised. The overall mathematical and systems biology approach at this juncture appears sound, and it is their exposition that is the main contribution of this paper (and not the initial parameter estimates).

There are several future directions for this research. First, incorporation of our knowledge of the exposome and its impacts on carcinogenesis may be incorporated by linking flows and coefficients related to specific exposures relevant carcinogenetic events such as mutation, cell proliferation, replication and other biological mechanisms through which environmental exposures impact cancer initiation and progression. For example, nonmutational mechanisms (i.e. epigenetic events that turn genes on or off through methylation) can be incorporated into the model through those model coefficients that impact tumor initiation and progression. This requires knowledge or hypotheses regarding how the epigenetic event under consideration impacts carcinogenesis.

Second, the diversity of different pathways to cancer may be represented by fitting models for each pathway. For pancreatic cancers, precursor lesions include the mucinous cycstic neoplasm (MCN), the intraductal papillary mucinous neoplasm (IPMN) and the pancreatic intraepithelial neoplasia (PanIN). In this paper we modeled the PanIN pathway, as it is the one responsible for the majority of pancreatic cancers. Pathway-specific models could be developed for cancers that are initiated by MCN and IPMN lesions.

Third, the carcinogensis model provides specific conditions for cancer progression, metastasis and remission. These could be used to predict treatment efficacy, and to evaluate alternative treatments by incorporating information on how specific treatments impact those model coefficients describing cancer cell proliferation, death, and progression to distant sites.

Information on how combinations of agents that differentially impact cancer cell proliferation, death and metastatic capacity could be used in the model to evaluate novel multi-chemothearaputic agent treatment regimes.

Finally, the technique is readily extensible to different cancers, and also to other chronic diseases.

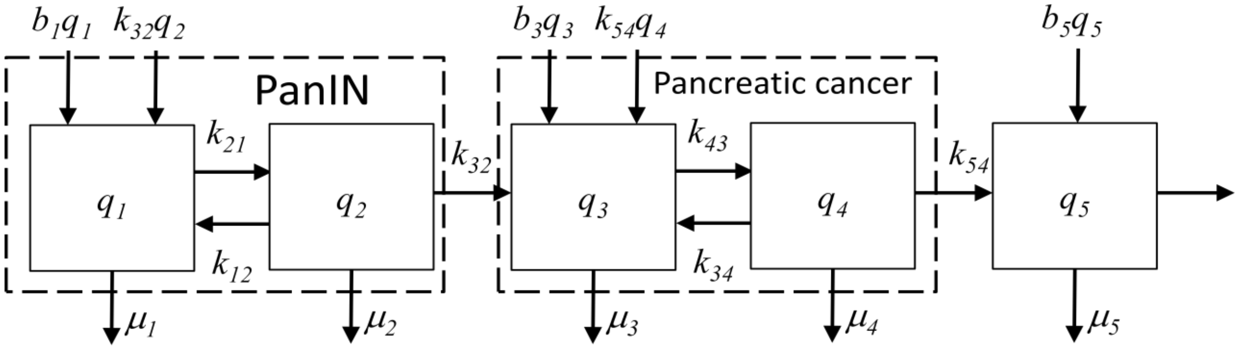
A note on latency modeling in geographic and dynamical systems is warranted. A frequently used approach available in most dynamical system modeling software is the incorporation of specific time lags, in which the model incorporates explicit delays, in the flow from one compartment to another. Hence one could simply represent cancer latency by explicitly delaying (e.g. holding back) the entry of particles in the model to a destination compartment once they have exited the source compartment. This has two disadvantages. First, apriori knowledge of the time lag is required, and second the use of explicit time lags implies the model is incomplete. When the compartmental system is properly specified a distribution of residence times is observed that is Erlang distributed and that is representative of the empirical latency times.

One of the original motivations for this research was to derive process-based approaches to estimate disease latencies suited for specification of the space-time lag needed to model dynamic geographical systems. Logical next steps are to apply these disease latency distributions in cluster analysis, surveillance, and space-time disease models.

Acknowledgements

This research benefited from discussions with Jaymie Meliker and Chantel Sloan. This research was funded in part by grants R44CA112743 and R44CA135818 from the National Cancer Institute; and grant R21LM011132 from the National Library of Medicine.

Appendix A. Model for meisosis.



Cell division results in the damaged strand of DNA going to 1 gamete and the normal strand to the other. This model would apply during gametogenesis, rarely if ever encountered for pancreatic cancers. However, it is useful when considering cancers that occur during childhood.

$$^{dq_1}/_{dt} = q_2(k_{12} + k_{32}) + q_1(b_1 - k_{21} - \mu_1) \qquad \text{(Eqn 4)}$$

$$^{dq_2}/_{dt} = q_1k_{21} - q_2(k_{12} + k_{32} - \mu_2)$$

$$^{dq_3}/_{dt} = q_2k_{32} + q_3(b_3 - k_{43} - \mu_3) + q_4(k_{34} + k_{54})$$

$$^{dq_4}/_{dt} = q_3k_{43} - q_4(k_{34} + k_{54} - \mu_4)$$

$$^{dq_5}/_{dt} = q_4k_{54} - q_5(b_5 - \mu_5)$$

References

Amikura, K., M. Kobari, et al. (1995). "The time of occurrence of liver metastasis in carcinoma of the pancreas." International Journal of Pancreatology **17**(2): 139-146.

Brownson, R. C. and F. S. Bright (2004). "Chronic disease control in public health practice: looking back and moving forward." Public Health Reports **119**(3): 230–238.

Campbell, P. J., S. Yachida, et al. (2010). "The patterns and dynamics of genomic instability in metastatic pancreatic cancer." Nature **467**(7319): 1109-1113.

Colditz, G. A. and A. L. Frazier (1995). "Models of breast cancer show that risk is set by events of early life: prevention efforts must shift focus." Cancer Epidemiology Biomarkers & Prevention **4**(5): 567-571.

Edge, S., D. Byrd, et al., Eds. (2010). Exocrine and endocrine pancreas. AJCC Cancer Staging Manual. New York, NY, Springer.

Jacquez, G. M. (2009). "Cluster Morphology Analysis." Spat Spattemporal Epidemiol **1**(1): 19-29.

Jacquez, J. and C. Simon (2002). "Qualitative theory of compartmental systems with lags." Mathematical Biosciences **180**(1): 329-362.

Jacquez, J. A. (1996). Compartmental analysis in biology and medicine. Ann Arbor, Biomedware Press.

Jacquez, J. A. (1999). Modeling With Compartments. Ann Arbor, BioMedware Press.

Jacquez, J. A. (2002). "Density functions of residence times for deterministic and stochastic compartmental systems." Math Biosci **180**: 127-139.

Juckett, D. (2009). "A 17-year oscillation in cancer mortality birth cohorts on three continents – synchrony to cosmic ray modulations one generation earlier." International Journal of Biometeorology **53**(6): 487-499.

Koopman, J. S., G. Jacquez, et al. (2001). "New data and tools for integrating discrete and continuous population modeling strategies." Ann N Y Acad Sci **954**: 268-294.

Kopp-Schneider, A., C. J. Portier, et al. (1991). "The application of a multistage model that incorporates DNA damage and repair to the analysis of initiation/promotion experiments." Mathematical Biosciences **105**: 139-166.

Maitra, A. and R. H. Hruban (2008). "Pancreatic Cancer." Annual Review of Pathology: Mechanisms of Disease **3**(1): 157-188.

MDCH. (2011). "Age-Adjusted Death Rates by Race and Sex Michigan and United States Residents, 1980- 2010."   Retrieved March 5 2013, 2013, from http://www.mdch.state.mi.us/pha/osr/deaths/dxrates.asp.

Miniño, A. M. and S. L. Murphy (2012). Death in the United States, 2010. NCHS Data Brief, National Center for Health Statistics. **99**.

Nachman, M. W. and S. L. Crowell (2000). "Estimate of the Mutation Rate per Nucleotide in Humans." Genetics **156**: 297–304.

Nikiforov, Y. and D. R. Gnepp (1994). "Pediatric thyroid cancer after the chernobyl disaster. Pathomorphologic study of 84 cases (1991–1992) from the republic of Belarus." Cancer **74**(2): 748-766.

Ries, L., D. Harkens, et al. (2007). "SEER Cancer Statistics Review, 1975-2004."   Retrieved March 2008, from http://seer.cancer.gov/csr/1975_2004/.

Rothman, K. J. (1981). "Induction and latent periods." Am J Epidemiol **114**(2): 253-259.

Seton-Rogers, S. (2012). "Tumorigenesis: Pushing pancreatic cancer to take off." Nat Rev Cancer **12**(11): 739-739.

Yachida, S., S. Jones, et al. (2010). "Distant metastasis occurs late during the genetic evolution of pancreatic cancer." Nature **467**(7319): 1114-1117.