

# Preliminary investigation of a mutation network model for predicting CRC prevalence based on information from colonoscopies

Eric Sherer<sup>1,2,3,4</sup>, Seza Orcun<sup>1</sup>, Ann Rundell<sup>5</sup>, and Doraiswami Ramkrishna<sup>6</sup>



1. e-Enterprise Center, Purdue University; 2. VA CoE on Implementing Evidence-Based Practice; 3. Indiana University Center for Health Services and Outcomes Research; 4. Regenstrief Institute, Inc.; 5. Biomedical Engineering, Purdue University; 6. Chemical Engineering, Purdue University

**Abstract:** A sequential somatic mutation network model is used to describe colorectal cancer (CRC) prevalence data. A hypothetical example is shown which demonstrates how model could be used to predict the likelihood and genetic characteristics of an individual patient's colorectal cancer based on results from colonoscopies. The model is currently being tuned to such information with the collection of Lagrangian polyp and CRC prevalence throughout patients' lifetimes. Future research will refine the genetic network by examining polyp and CRC biopsy characteristics collected throughout patients' lifetimes.

## Sequential network => CRC prevalence

### Asymmetric division

**Objective:** Develop a predictive model for the likely age-incidence and genetic characteristics of CRC.

**Methods:** A multi-state model is developed where single features are altered sequentially in a stochastic manner (Michor et al. 2005). These features could potentially include CIN associated genes, MSI genes, or hypermethylation states. As an example, a 3-state model is applied to the Indiana CRC incidence data from 2000-2004. The 12 rate parameters are varied in a 2-factorial manner. Each parameter set provides the initial guess for a local parameter optimization with the resulting fit for a "good" parameter set shown.

**Future Uses:** The state model will be coupled with models predicting growth, invasion, and response to adjuvant chemotherapy. The presence of polyps and their characteristics could be linked with the likelihood of CRC and its characteristics. These differences could be used to differentiate between the likely behaviors and responses of patients to various treatments.

$$R_i(t) = \exp(-\sum k_{ij} t)$$

$$R_{i,j}(t) = \int_0^t k_{ij} \exp(-\sum k_{ij} \tau) \int_0^{\tau} k_{ij} \exp(-\sum k_{ij} s) ds d\tau$$

$$P_i(t) = \int_0^t k_{ij} \exp(-\sum k_{ij} \tau) \int_0^{\tau} k_{ij} \exp(-\sum k_{ij} s) ds d\tau$$

CRC incidence, IN 2000-2004  
network model fits with most data

Cancer type for white females based on model fit to CRC incidence

Assumed to be CRC (red dot)  
Assumed to be polyp (green dot)  
Assumed to be asymptomatic (yellow dot)

**Advantages:**  
- Analytical formulation allows for rapid solution  
- Good fit to incidence data

**Disadvantages:**  
- Finite number of discrete patient classes  
- Preferred pathways and states unknown

## Patient Differentiation based on +/- "polyp" types

### Polyp prevalence => likelihood of behavior

Slight differences in behavior of stratification based on demographics implies that "environmental" conditions vary between the classifications. The variable conditions will produce different observable behaviors with time.

Parameter sets for the model were fit to the IN incidence data. Besides CRC incidence, these parameter also give the rates of formation for three (hypothetical) types of polyps. Since the parameter sets vary between the demographic groups, the likelihood of various polyps will also be different. So, the likelihood of an unknown patient - that is, of unknown race and gender - belongs to a given demographic group can be found by observing the polyp classification. The figures below are the probability that at least 1 polyp of a given type is present, but the resolution can be increased by including the odds of combinations of multiple polyps.

Probability that at least 1 polyp of a given type is present as a function of patient age. The resolution of the model could be increased by accounting for combination of multiple polyps.

Probability of having a certain race/gender based on the polyp presentation at age 50. It is assumed that the M:F ratio is 50:50 and the B:W ratio in Indiana is 10:90.

**Question of Interest:**  
Can the demographic of an unknown patient be identified from observations?

**Question of Interest:**  
Can the demographic probabilities - identified from observations - be used to predict the likelihood of developing CRC in the future?

## Patient Differentiation

### Alternative parameter sets

Multiple (>10) parameter sets have been identified which fit the CRC data nearly equally well and all could potentially be valid. While the predictions of CRC incidence are consistent between parameter sets, the pathways taken may differ. (A sensitivity analysis is currently underway.) Additional data on polyp incidence and classification will be necessary to resolve these differences and tune the model for CRC incidence following a polyp.

Probability that at least 1 polyp of a given type is present as a function of patient age. The resolution of the model could be increased by accounting for combination of multiple polyps.

Probability of having a certain race/gender based on the polyp presentation at age 50. It is assumed that the M:F ratio is 50:50 and the B:W ratio in Indiana is 10:90.

## Probabilities of patient classifications => likely CRC incidence

### Model (x2) predicted CRC based on a colonoscopy at age 50yrs

While the demographic to which the unknown patient belong cannot be determined exactly, the probability of belonging to each specific group can be determined from based on colon observations with time. An expected behavior can then be determined by accruing the contributions from each possible model. Such results could be used for designing screening protocols which are adaptive to the individual patient based on the results of their previous colonoscopies. An interesting feature is that a patient is less likely to develop CRC if they have a polyp of type II present at age 50 than if there were no polyp in Model #2. Polyps of type II are exclusively seen in white females, who have the lowest CRC incidence rates, while clean colons include contributions from white males and black males both with slightly higher CRC incidence.

Model predicted - for both the parameter set shown above and to the left - incidence of CRC based on whether a polyp of type III is present or not at age 50.

### Summary and Conclusions:

- Preliminary investigation using a hypothetical sequential mutation model fit to age / gender demographic CRC incidence
- A methodology is outlined for finding likelihood of race/gender demographic using polyp prevalence
- Due to the non-uniqueness of parameter sets there are several equally valid predictions
- From the likely demographics, it is shown how the overall likelihood of CRC calculated from polyp incidence at age 50. However, different conclusion can be drawn depending on the parameter set used.
- Additional data on polyp in incidence as well as state identification/mapping of polyps and CRC needed for model refinement and identification.

### References:

Biau G, Lasinski M, Orcun S, Hsu SH, Caruthers J, Delgass N, and Venkatasubramanian V. High fidelity mathematical model building with experimental data: A Bayesian approach. *Computers & Chemical Engineering*, 32: 971-980, 2008.

Luebeck EG and Moolgavkar SH. Multistage carcinogenesis and the incidence of colorectal cancer. *PNAS*, 99: 15095-15100, 2002.

McDougal SR, Anderson ARA, and Chaplain MAJ. Mathematical modelling of dynamic adaptive tumour-induced angiogenesis: Clinical applications and therapeutic targeting strategies. *Journal of Theoretical Biology*, 241: 564-589, 2006.

Michor F, Iwasa Y, Lengauer C, and Nowak M. Dynamics of colorectal cancer. *Seminars in Cancer Biology*, 15: 484-493, 2005.

Ramos-Conde I, Chaplain MAJ, and Anderson ARA. Mathematical modelling of cancer cell invasion of tissue. *Mathematical and Computer Modelling*, 47: 533-545, 2008.