University of Nevada, Reno

A Mathematical Model on the Influence of Diet on Colorectal Cancer

A thesis submitted in partial fulfillment
of the requirements for the degree of

BACHELOR OF SCIENCE IN MATHEMATICS SPECIALIZED IN STATISTICS

by

DANA PAIGE WINTERRINGER

Deena R. Schmidt, PhD, Thesis Advisor

May, 2021

UNIVERSITY OF NEVADA RENO

THE HONORS COLLEGE

We recommend that the thesis
prepared under our supervision by

DANA PAIGE WINTERRINGER

entitled

A Mathematical Model on the Influence of Diet on Colorectal Cancer

be accepted in partial fulfillment of the
requirements for the degree of

BACHELOR OF SCIENCE IN MATHEMATICS SPECIALIZED IN STATISTICS

_____

Deena R. Schmidt, PhD, Thesis Advisor

_____

Matthew Means, P.S., Dean, **Honors College**

May, 2021

Abstract

It is well known that diet has a large influence on our health and well-being. Diet also plays a big role in various diseases, such as colorectal cancer. The correlation between colorectal cancer and dietary factors has been studied widely in both the mathematical and biological fields, often with conflicting results. The aim of this study is to develop and analyze a stochastic model of tumor progression that focuses on how different dietary factors influence the risk of developing colorectal cancer.

Acknowledgments

I cannot express enough thanks to Dr. Schmidt for being the best mentor I could ask for. This project would have been impossible without your encouragement and guidance over the past two years. Thank you to Dr. Hurtado for advice and support in order to refine the project. I am looking forward to conducting more projects and working together in my graduate studies. I cannot thank you both enough for leading me to my true calling.

To my wonderful parents, thank you for always believing in my work and giving me the principles to succeed in my academic career. I would like to thank my lovely grandparents as well for supporting me in my time at the university.

Last but certainly not least, I would like to extend my gratitude to my friends. I am thankful I can share my work with such amazing peers.

Contents

List of Tables

List of Figures

1 Introduction

Colorectal cancer is cancer that affects the colon and/or the rectum. It is the second leading cause of cancer death in the United States and is the third most common type of cancer, excluding all skin cancers [7]. Understanding how diet influences one's risk for developing this disease is an important problem in today's society. Dietary factors that seem to increase risk are alcohol, red meat, and processed meat [5], [10] and one main factor that seems to decrease risk is high vegetable intake [19]. However, it's not clear that just avoiding red meat, for instance, will reduce one's cancer risk. With a mathematical model, we can study different combinations of dietary risk factors under different scenarios.

At the molecular scale, colorectal cancer occurs through one of two different biological pathways: chromosomal instability (CIN) and microsatellite instability (MSI). CIN accounts for 85% of all colorectal cancers and MSI accounts for 15% of all colorectal cancers [25]. In this project, we will focus on the CIN pathway. Using the mathematical model of this pathway developed in [25] as a starting point, we will employ a stochastic model in order to study the waiting time until the first cancerous tumor cell develops. Then we will analyze how changing the mutation rates for the tumor suppressor genes (TSGs). TSGs can act as a proxy for different dietary effects over an individual's lifetime due to direct correlations between diet and mutation rates. Using this information we will analyze how diet influences the timing of colorectal cancer development.

My research will focus on translating the process of carcinogenesis with diet as a risk factor into a mathematical model. I will use the statistical programming software R in order to create a simulation of colorectal tumor progression across an individual's lifetime. The model in [25] uses a stochastic process in order to model tumor progression, which I will utilize as the basis for this project. I will expand upon this model by incorporating several dietary risk factors and using the literature to refine those parameter values. This project

holds significance since I will be able to create a mathematical model of tumor progression which will not involve trials or lab work, but rather simulated data. Literature on this topic suggests that there has not been a unanimous conclusion reached in regard to the effect of such dietary factors on the progression of colorectal cancer. My research aims to use a unique statistical approach that has yet to be seen in previous works in order to draw a conclusion that indicates what kinds of dietary factors affect colorectal progression and what specific tumor suppressor gene they directly impact.

2 Literature Review

Across most studies of the effect of diet on colorectal cancer, there is no single, definite correlation that describes the relationship with a certain diet and colorectal cancer. Much of what is included in literature is reasonable speculation, such as if you generally eat in an unhealthy manner you may be more likely to contract colon cancer in your lifetime. Clearly, this must be broken down further.

Colorectal cancer itself has been well-studied and modeled. Most literature models colorectal cancer as a four or five stage model, sometimes condensed to even fewer stages for mathematical purposes. The five stage model presented in [25] is the foundation of our own original model. Modeling the progression of colon cancer mathematically has been an idea present since Armitage and Doll proposed a model in the 1950s after they related the age-specific incidence of cancer to ”rate-limiting steps” that lead to the formation of a malignant tumor [9]. At the time, this was a simple two stage model which was expanded upon once more research was conducted on the inactivation of tumor suppressor genes and activation of oncogenes in the progression of cancer. Thus, this is where the idea of creating states in a colorectal cancer model originates.

The early stages of colorectal cancer generally begin with the mutation of the APC

regulatory pathway [25], [8], [23]. This occurs in both the chromosomal instability pathway and the microsatellite instability pathway. The chromosomal instability pathway, however, is more well studied as it is the pathway for most colon cancers. Hence literature is more apt to track its progression via inactivation of tumor suppressor genes and relate it to what stage of growth in the colon it correlates to as well as the age of the patient. With microsatellite instability, this is not always the case. In [25], it is estimated that the remaining 15% of CRC takes on the MSI pathway as opposed to the CIN pathway, however this is somewhat disputed in [8] and [23] as there are other, more rare pathways proposed and specified. The standard 85:15 proportion is still assumed by modelers and researchers as most CRC tumors will either exhibit CIN or MSI, but not both. This means that one can conclude that other lifestyle factors such as diet can act as an accelerant of sorts to eventual malignancy depending on pathway. What also should be noted is that MSI is a more slow growth over time, characterized by a sequence of small-scale events that eventually lead to a carcinoma as opposed to CIN's larger scale, accelerated genetic changes [23].

It remains legitimate to model either pathway by states and transitions, which thus becomes an inherently stochastic process. In [13] a combined sequence is proposed for both pathways. For both CIN and MSI, they observed a normal stage transitioning to an early adenoma after many decades due to a mutation in APC and/or Wnt signaling. This leads to an immediate adenoma after two to five years as a result of KRAS activation, or possibly BRAF in the case of MSI. Following that comes a loss of heterozygosity for CIN with a potential mutation in the Smad4 gene or CDC4 gene, or a CDC4 mutation for MSI. This would occur after another two to five year period and transition into a late adenoma. Finally comes another loss of heterozygosity in the mutation of p53 for the CIN pathway that would lead to cancer, and a mutation for the TGFBR2, BAX, and/or IGF2R genes in the MSI pathway. In [25] the CIN pathway remains consistent, and in both articles the p53 mutation is regarded as a single transition due to its haplo-insufficiency. However

[25] assumes a singular BAX mutation following a TGF mutation for the MSI pathway. In general, most models for colorectal cancer that regard both pathways are seemingly stochastic with their basis in transitioning states by mutations and activations of genes. Note that although helpful in visualizing CRC progression, the moment in which certain mutations occur cannot be thought of as being completely linear. Much literature suggests that there are some tumors that give evidence of the absence of driver steps in pathways, such as the absence of an APC mutation in certain tumor incidences [16], [17]. Thus with such substantial modeling of progression and curious findings in real-life studies in tumors, one must question what factors could accelerate or dampen the progression and how they affect the proposed pathways.

In one example, a particular group of authors conducted a study on the influence of diet, activity, and lifestyle on colorectal cancer. Said study found that consumption of processed red meat was linked to mutations in tumor suppressor gene p53 [21]. However, eight years after this study the same group of authors looked to bolster their previous claims. In doing so, they found that their first claim was less significant than originally thought. Rather than finding a strong correlation with high consumption of processed red meat and p53 mutations, instead a correlation between eating a varied diet high in fiber and vegetable intake would reduce p53 mutations and risk of colorectal cancer [19]. These two types of diets are dubbed the ”Western” and the ”Prudent” diet, respectively.

Although claims of correlation of diet and p53 mutations are disputed, we observe a more concise narrative when reviewing studies on the APC gene. APC plays a key role in colorectal cancer as it is the first tumor suppressor gene to be inactivated in the progression of colorectal cancer [25], [8], [17]. The inactivation of this gene directly leads to dysplastic crypts in the colon. This is because the inactivation of both copies of APC leads to an increase in the birth to death ratio in the corresponding cell, hence leading to clonal expansion and subsequently, crypts. Such crypts can be directly affected by diets high in starch

and carbohydrates. [8].

Acrylamide, in particular, is a substance present in heat-treated carbohydrate-rich foods such as coffee, fried or baked potatoes, and various bakery goods. It has been labelled as a ”probable human carcinogen” by the International Agency for Research on Cancer. Acrylamide was positively associated with colorectal cancer risk, and more particularly it was found to have activated KRAS mutations among men, but strangely, not women. Although, similar to most current studies on diet and CRC, there is no true ”direct” association with any dietary factor and CRC, but rather observations on the molecular level lend to such logic. For acrylamide, alongside significant experimental data indicating a carcinogenic status, it is also observed that the substance is oxidized to epoxide glycidamide which then creates adducts with DNA bases and hence forms mutations. Acrylamide and glycidamide exposure is observed to influence the hormone levels in colorectal cells by increasing gene expression in sex hormones thereby leading to tumors circumventing surveilling apoptosis mechanisms [2]. However, these studies were conducted on small scales and thus led to authors being unable to state whether or not the findings were significant.

Such specific observations that focus more on a particular substance in diet rather than a food group appear to have stronger, more consistent correlations. For example, a simple specification of ”animal protein” and ”heme” versus ”red meat” produced stronger correlations and evidence that was more solid when evaluating what genes in the CRC process were directly affected by diet [2], [3]. Heme, the iron type that is found exclusively in animal proteins, generated particularly interesting findings. In [3] it was observed that there was a direct dose-response relation between heme iron consumption and colorectal cancers with the specific DNA transition from guanine to adenine in the KRAS and APC genes as well as the overexpression of the TP53 gene. Some contradictory results were presented in [2], but the same authors noted that studies in the relation of animal protein

to CRC were shown to have a positive correlation with cancer risk and high consumption. Specifically, high intake of animal protein per 17 grams was associated with tumors in the colon that possessed a mutation on codon 12 which is a KRAS mutation. Other notable dietary factors presented with strong correlations in [2] are fish and vitamin A, with the former having more variations by region, and the latter having a strong correlation with increased risk by low consumption.

The authors of [17] used a targeted gene sequencing analysis performed on 468 colorectal tumor samples across 1321 different genes in which the driving role of APC in colorectal cancer is extended to not only being the first key step in progression but also having a formally underestimated role in prognosis. In analyzing the mutational status of the gene, a compelling case was found for the second mutation of APC which supports the gene's bi-allelic nature. This idea is reinforced in both [25] and [8] that the inactivation of solely the first copy of the APC gene does not garner phenotypic changes. This is unlike the characteristics of genes such as p53. When the first copy of the APC gene is inactivated, then the other copy is inactivated by an additional point mutation [8]. Additionally, this correlates with the studies conducted in [17] in that tumors with exclusively zero or two APC mutations had a worse chance of survival than tumors that only presented a single APC mutation.

Perhaps the most intriguing finding of the multigene mutation classification is that tumors lacking any APC mutation carry a worse prognosis than single APC mutation tumors, however tumors with 2 APC mutations that also have mutant KRAS and TP53 confer the poorest survival among any subgroup of tumor. Additionally, they observed that almost 30% of all tumors harbored only one APC mutation without allelic loss, most of which were not of the microsatellite instability nature [17]. In our model, we may view this as the inactivation of both the APC gene copies in sequential order, since the inactivation of a tumor suppressor gene is likened to the first two mutations in the stochastic model from

[25]. Out of all the tumor suppressor 468 tumors studied, 199 had both APC and TP53 mutations, with only 1% exhibiting MSI behavior and thus this pairing would mostly correlate with CIN. In fact, tumors with wild-type APC, rather than mutant-type APC are more associated with the MSI pathway with a strong correlation to mutations of the BRAF gene. Tumors that exhibit MSI behavior also have a much better prognosis. From these observations of the APC gene's behavior in tumors, it was concluded that APC usually co-occurs with KRAS or TP53 mutations. This also suggests that APC needs to partner with one or more driver mutations, but can still be regarded as the necessary "first step" in the CIN pathway [17].

Furthermore, a survival analysis indicated that 2 APC truncating mutations in the presence of mutant KRAS and TP53 carry a substantially worse prognosis than single truncating mutations, but are equivalent to tumors lacking any APC mutations. APC mutations had a hazard ratio lower than that of KRAS or TP53 (obtained from a Cox Model), but the increased hazard for KRAS or TP53-mutated tumors may be largely borne by patients with both double APC mutations and APC/KRAS/TP53 mutations [17]. Since APC's first two copies are the first TSGs to be inactivated, and TP53 is considered to mutate alongside APC, then we can connect these two genes through the additional KRAS gene and diet to pose questions about how an individual's diet may affect their progression of colorectal cancer. Perhaps most crucial to consider would be the survivability of contracting the KRAS and/or p53 mutation after both copies of APC have been mutated. Since there is a confirmed discrepancy in prognosis outcome, we must investigate if diet improvement between the mutations of the three genes could prolong the progression or stop it prior to death of natural causes. Due to these relations amongst the genes, as well as dietary factors that connect them further, the question is raised on how one can demonstrate the changes in the APC/KRAS/TP53 dynamic as an individual's diet fluctuates over lifetime and what implications it may have for CRC risk.

Aberrant crypt foci, commonly abbreviated as ACF, are considered the earliest neoplastic lesions in the progression of CRC and are considered a chemopreventive response to diet [11]. They are one of the earliest premalignant lesions in the development of colorectal cancer and show loss of heterozygosity on the 5q chromosome, which is the position of the APC gene [9]. ACF are labeled as a biomarker for colorectal cancer risk, precede dysplasia and single ACF are prevalent in patients under the age of 40 [18]. This finding is rather interesting given that the general public assumption is that most instances of general cancer appear late in the life of an individual. Thus such findings give rise to the notion that diet quality at any given point in a person's lifetime couId be a key factor in colorectal cancer risk. Additionally, it is found that an increase in the prevalence of ACF is thought to be directly parallel to the step or "state" progression of colorectal cancer [18] which serves as yet another verification of the seemingly natural stochastic nature of CRC progression.

It is clear that CRC can effectively be modeled using a multistage stochastic process. Intuitively, one can think of employing continuous time with discretized states that correspond to the inactivation of tumor suppressor genes or proper cell functioning. There are experimental findings that conclude a link between diet and the progression of carcinogenesis, but it is unclear how this can occur on a molecular scale. It can also be concluded that each instance of carcinogenesis is not the same as the next and although multistage modeling is effective, other scenarios or alternate pathways should be considered as mutations ultimately do not occur in a linear fashion.

3 Mathematical Model of Colorectal Cancer

When modeling carcinogenesis, a stochastic approach is one that is well-documented and well-verified by the mathematical and biological communities. Going further, modeling carcinogenesis in colorectal cancer is best modeled as having multiple states with certain

cancer genes having been identified as being significant in its progression. In fact, it is theorized that a particular codon mutation of the KRAS gene may initiate colon cancer in humans as a whole. This naturally follows the three steps of carcinogenesis: initiation, promotion and progression, with the KRAS mutation being the initiation step [26]. This project shall focus primarily on tracking and analyzing the final stage of carcinogenesis, progression, as it behaves in relation to the two former stages.

Various accepted models of carcinogenesis will express a certain amount of stages. Such stages are defined when breaking down discrete events that lead to the eventual development of a carcinoma. The idea of mutli-stage carcinogenesis dates back to 1954 with the publishing of the Armitage-Doll model, a model that proposed a set of definite genetic events were the precursor to the onset of cancer [26]. For the most part, carcinogenesis models will at the very least consist of two stages. For some cases, one stage is deemed sufficient as the progression to a carcinoma. The mutation or transition at hand is said to be categorized as one singular, rare genetic event.

In this case however, we recognize that a five or six stage model is appropriate. Such models are well-accepted. They define a "normal cell" stage **N** which we can explain as a state having 0 tumor cells. Following that we have several intermediate stages given as $\boldsymbol{I_1, I_2, I_3}$ in the CIN pathway and $\boldsymbol{J_1, J_2, J_3, J_4}$ in the MSI pathway. In this case, each step between our intermediate stages represents the inactivation of each individual tumor suppressor gene, or copy of one of the tumor suppressor genes. Finally, we have the final stage **T** which we explain as being our tumor stage. In our model, we categorize this as being the appearance of the first tumor cell, which applies to the models in [25] and more generally [26].

The nature of colorectal cancer progression makes the process most logically explained by a stochastic model. When viewing the comprehensive two-pathway model of

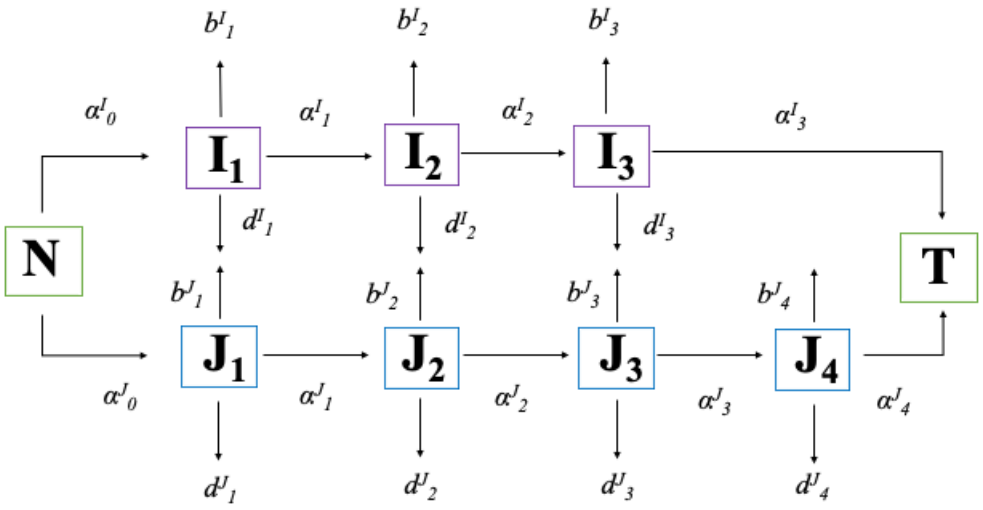

Figure 1: Two-Pathway Colorectal Cancer Model, Modified from Figure in [25]

colorectal cancer, it is mapped over time t with $T(t)$ being the number of tumors at each stage [25]. This process is assumed to be *essentially* Markov, as any future cell at time t will depend on the characteristics of the current cell at time t, with no dependance on on its past behavior. Also note that the mutation rates are extremely small and that the birth rates are greater than the death rates [26]. Thus, a stochastic model will fit the progression of colorectal cancer. Such rates are displayed in Figure 1, modeling the two pathways of colorectal cancer. Also note that biologically, stem cells follow a nonhomogeneous Feller-Arley birth-death process for their proliferation. Under the aforementioned assumptions, we then have a continuous time process.

We import the rates given in [25] for birth, death, mutation, and proliferation. Then we set a maximum number of simulations and compute the appearance of the first cell in the tumor cell. From here, we create our update matrix, which is formulated by the following: the number of rows is equivalent to the number of transition events, with the number of columns being our five possible states. For the CIN pathway, we account for four mutation

events, three birth events, and three death events, making for ten transition events in total. In the case of MSI, we assume six states, so this would equate to five mutation events, and four birth and death events respectively totaling thirteen rows. In both cases, this is followed by implementing the Gillespie Stochastic Simulation Algorithm.

Using the statistical programming software R, we proceed to accurately reflect the carcinogenesis model proposed in [25] while adding time-varying mutation rates. This model shall act as a reflection of age and diet. The model proposed here was a stochastic model developed using muti-level Gibbs sampling developed from SEER (Surveillance, Epidemology, End Results Program) data for cancer incidence in the United States. Thus, we use the SEER incidence data and its division by age group, as well as the mutation, birth and proliferation rates from [25]. In using this as our base model, we are able to make conclusions about the relations between cancer, age, and diet in the United States.

Building the simulation, it would be ideal to set the number of healthy cells at ten billion, defined as $\boldsymbol{N = 10000000000}$. The number of cells in the colon is very large as one may guess, with 1,000 to 4,000 cells per crypt and about ten million crypts in total [16]. However, such a large number is computationally infeasible with our model, thus we will focus on a smaller scale and discuss how to interpret the results in relation to the full size of the colon. The simulation uses the Gillespie Stochastic Simulation Algorithm and employs the rates given in [25]. Mutation and birth rates are given outright along with death rates, however, death rates are not directly employed. Rather, proliferation rates defined as $\boldsymbol{p_i}$ are utilized and are given from the death rate subtracted from the birth rate for each stage. Taking all these rates directly, we define this as being our Constant Rate Model for both the CIN and MSI pathway.

3.1 Chromosomal Instability Pathway

When studying the chromosomal instability pathway, we account for three tumor suppressor genes along the progression. Such genes are the APC gene, KRAS gene, and the p53 gene, in that order. In Figure 1, these are defined as our α_i^Is - however it should be noted that KRAS and p53 are combined into a singular mutation rate while APC is left as two separate mutation rates for two different copies of the gene - α_0^I and α_1^I respectively due to haploinsufficiency. This occurs when the first copy of the gene is inactivated, then the second copy is not sufficient enough to produce the standard phenotype. In plain terms, if the first copy is taken out in the process then the second copy is irrelevant [25]. Thus, we use a single mutation rate to represent both Smad4 and p53, additionally this condenses the number of stages in our model from potentially seven stages to five.

APC is identified as one of the most crucial steps in colorectal cancer carcinogenesis as it is quite literally the first step in progression, it is found in eighty percent of all tumors in the colon, and heterozygous mutations construct an autosomal predisposition for humans for dominant colon cancer [17]. APC is also largely believed to pair with KRAS mutations as well as p53 mutations. KRAS mutations are found in 35-45% of colorectal cancers while p53 mutations are found in 35-55% [24]. These findings are relatively consistent across all studies and literature.

The pathway in [13] and the "Vogelgram" in [16] correspond to a general sequence of an APC mutation, KRAS activation, some other genetic, molecular, or environmental events, and finally a TP53 mutation. This is similar to the model in [25] with the only difference being a Smad4 mutation occurring prior to TP53 inactivation. We aim to bridge the "environmental factors/molecular changes" [24], [2], [13] gap in the understanding of the CIN pathway through our incorporation of diet as a risk factor. Although other genes have been mentioned in the pathway to colon cancer and cited as being present in tumors,

we can exclude them and focus on the key players APC, KRAS, and TP53. Vogelstein simplified the model of CRC in this fashion due to the fact that each of these genes is associated with a major cell signaling pathway [16], [4] - this logic was followed in [25]. Additionally the study in [17] indicated that these three genes remained the most frequent in their tumor study with Smad4 being the fourth most frequent. Since these genes are the most documented due to their frequency, they are also among those most studied with correlations to dietary factors [21], [2].

As one can assume, such frequencies of mutations imply that this pathway of colorectal cancer may be consistent and linear in cases that progress to full malignancy, but will vary otherwise. In [17], 468 tumors surveyed did not always have mutations in these three genes. Other studies support this, in one stating that in their cohort study that about 60% of tumors had APC and TP53 mutations not cooccurring [1]. This suggests that there are alternate orderings in the CIN pathway. There is a very limited KRAS-first potential to progress to a full carcinoma, with the APC-first route having the highest potential to reach tumor status first [6]. Thus, we want to study if dietary factors that have been linked to these tumor suppressor genes and KRAS will change the order of mutations and hence accelerate or decelerate the progression of the CIN pathway.

Due to the large amount of literature on KRAS mutations, we regard the $\boldsymbol{I_2}$ state as a generalized "intermediate state," deviating slightly from the logic used in [25]. This is reasonable given the ambiguity in research of intermediary steps in the CIN pathway. The rates given in Tan and Yan's study, after all, were taken from Luebeck and Moolgavkar's 2002 SEER data study and even used exact rates from that study. The difference between the two studies is that Luebeck and Moolgavkar left the two events following the APC mutations as ambiguous, potentially epigenetic events [9]. In other models as well, there appears to be an agreement on APC mutations being the first stage and p53 mutations being the final stage, so we can use this model to explore different scenarios based on what

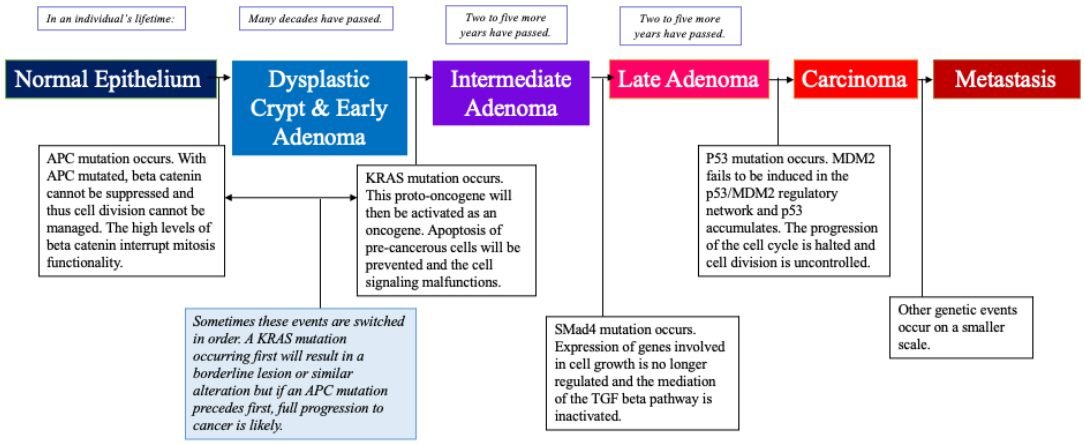

Figure 2: Biological Progression of Chromosomal Instability Pathway, Supported by [13], [6]

occurs in etween. From here, we can draw up different scenarios and compare against the CIN Constant Rate Model.

3.2 Microsatellite Instability Pathway

The microsatellite instability pathway of colorectal cancer is a sequence of more rare, sporadic events. For this portion of our model, we will craft our simulation somewhat differently as the way that MSI appears in tumors in real life is different to that of CIN occurrence. MSI in itself is caused by mutations in DNA mismatch repair genes, which in turn causes the microsatellites to change their own length at an unprecedented higher rate [23]. As we know, the MSI pathway occurs less frequently than the CIN pathway. The MSI pathway is can also be broken down further by other attributes such as HNPCC and the degree of instability, but this categorization is part of ongoing studies and most mathematical models such as that in [25] simplify CRC down to CIN or MSI. This makes sense in that

tumors can either have CIN or MSI but not both [23].

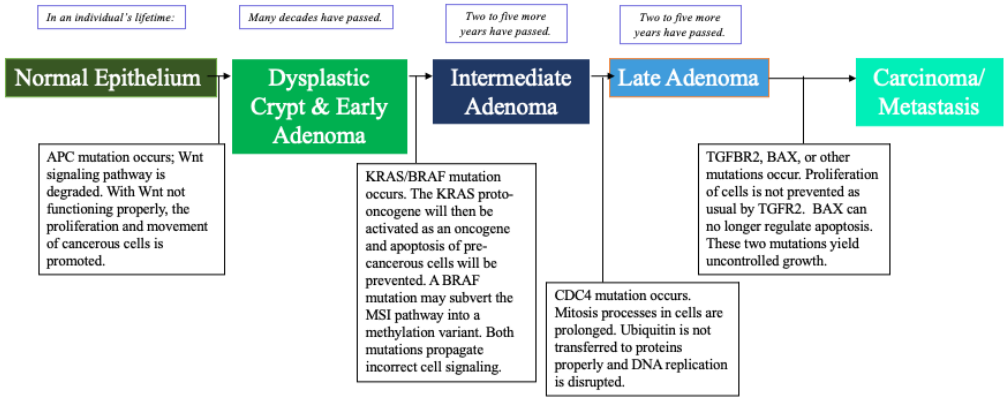

Figure 3: Biological Progression of Microsatellite Instability Pathway, Supported by [13], [27]

Similarly to the CIN pathway, many sources try to linearize the progression of cancer via the MSI pathway. While CIN has a handful of identifiable, singular mutations at each step, MSI is regarded as having more mutations and more possibilities of signaling being disrupted. Some sources differ slightly on the types of mutations occurring at each stage, but not at which instance they occur such as in the case of the TGFBR2 mutation [25], [27], [13]. Additionally, hereditary CRC is characterized by MSI and the pathway altogether is associated with post-replicative DNA mismatch repair deficiency [13]. This is potential reasoning as to why the MSI pathway in literature is explained in a few different ways with varied mutation checkpoints. These changes to normal-functioning proteins and signaling pathways are smaller in scale than the changes of the CIN pathway (see Figures 2, 3). Therefore in building our model, it is also important to note how MSI may behave differently over time. For one, its ties to DNA process disruption and familial inheritance could mean an earlier detection in a patient, as MSI has better prognosis and survival in

patients [15]. To maintain consistency with the previous SEER data, we will accept the number of stages in the MSI pathway as six, as stated in [25], but we still notice that the events that occur at each stage may be more ambiguous than that of the CIN pathway and account for this in our dietary modeling (see Figure 1). We will account for six official discrete states, with five mutation events $\boldsymbol{\alpha_j}$, four birth events $\boldsymbol{b_j}$ and four death events $\boldsymbol{d_j}$.

Clearly, the CIN pathway becomes more of a focus when incorporating dietary factors or studying the speed of carcinogenesis in CRC because of the different origins and effects of the pathways. It is also known that prognosis for MSI is better than that of CIN [17]. Other than this general knowledge and the link to mismatch repair, causes of MSI are unclear [20]. However, this does not mean that lifestyle factors are completely unrelated to the progression of the MSI pathway. In a large population, a case control study found that long term alcohol use was associated with patients having more MSI tumors [20]. Note that this study is one of few piece of literature that statistically confirms such a correlation, and also reinforces the mysterious nature of MSI and its causes. We also know that the KRAS mutation is said to occur in the MSI pathway as well [13], [25]. We will use this information to explore an accelerated model based on the KRAS "step" in the MSI pathway.

3.3 Dietary Factors

When we discuss diet and cancer, we must take in to account the fact that most individuals do not have a completely consistent diet over the course of their lifetime. For example, one individual may have a rather unhealthy diet in their adolescence and make strides to eat healthier as they age. Since mutations within the APC gene as well as biological predecessors such as ACF are strongly associated with diet, we are motivated to create a model that reflects changes in diet over time. Therefore we proceed by creating additional chromoso-

mal instability pathway simulations that incorporates reasonable changes from a good diet to a bad diet at a given points in a person's lifetime. We also will want to reflect consistent diets that may be more extreme cases in order to test the effect it may have on CRC carcinogenesis. In order to do this, dietary factors with substantial ties to certain points in the multistage process will be the most effective way to model this connection.

Table 1: Description of Dietary Factors in the Model.

Dietary Factor	Description	References
Red meat or heme	Positive correlation with APC, KRAS, and p53 mutations, with a focus on KRAS and p53 due to consistency in literature. Can toggle the first mutation between APC or KRAS in the CIN pathway and increase proliferation rates of the intermediate states.	[21] [2] [3] [22]
Vegetables	High vegetable intake has negative correlation with KRAS mutations. Decreases proliferation rate of $\boldsymbol{I_2}$ and toggles the first mutation in the CIN pathway.	[19] [2]
Alcohol	Long term intake associated with higher amount of MSI tumors, positive correlation with KRAS mutations in CIN pathway. Accelerates mutation rates in MSI pathway and increases proliferation rate of $\boldsymbol{I_2}$	[20]

Processed meat consumption showed an increased risk of APC mutations in a cohort study [22]. Red meat consumption showed an increased risk of KRAS mutations [3], [2] as well as TP53 mutations [21]. Associations with APC and TP53 are still disputed, thus through our simulation we plan to hypothesize if red meat could accelerate the progression

of CRC through the CIN pathway. Consumption of vegetables showed a negative association with KRAS mutations [2], [19] due to their richness in fiber and flavonols which inhibit nitroso compounds from forming (these will cause a base transition normally). Another study found that low consumption of vegetables was found to have a positive association with KRAS mutations [2]. Thus we can use KRAS as a point of reference for the Western vs. Prudent diet discussed in [19].

Ethanol is a widely confirmed mutagen and its metabolites within the human liver are considered carcinogenic. Thus, alcohol becomes a factor in any human cancer. In one case the results of a case control study suggested that long-term intake of alcohol is associated with a 60% increased risk of having a CRC tumor with MSI, particularly in the case of hard liquor consumption [20]. Note that this study did not focus on alcohol, but rather assessed several factors, and alcohol had the strongest correlation with MSI. For CIN, a cohort study suggested high lifetime alcohol intake made one more likely to have a tumor that possessed a KRAS mutation [12].

In regards to literature on a direct correlation between an individual's diet and their mutation, birth, death, or proliferation rates of the cells involved in carcinogenesis pathways, there have been no exploratory studies on the matter. However there have been suggestions in research on how this may occur. This includes the idea that an "accelerated" mutation may be the necessary step to aggressive colorectal cancer [23]. Thus, we can ponder conservative and extreme cases of dietary fluctuations in an individual and how that may affect various rates in their carcinogenesis progression. It would then follow to set an acceleration, or in the case of vegetables, a deceleration factor using the non-quantitative associations found in literature as a basis.

3.4 Assumptions

In terms of building the general pathway of carcinogenesis for our model, we assume N normal cells beginning at time t_0, with N large in regards to the epithelium. Consider the time interval $[t, t + \delta t]$. Then the probability that a normal cell at time t produces one normal and one mutated cell at time $t + \delta t$ is $\alpha_i(t)\delta t + o(\delta t)$ with $\lim_{\delta t \to 0} \frac{o(\delta t)}{\delta t} = 0$, for any α_i or α_j with $i, j > 0$.

Both the normal stem cells and mutated cells follow a nonhomogeneous general birth and death process with rates $b_i(t)$ and $d_i(t)$ for the chromosomal instability pathway, and rates $b_j(t)$ and $d_j(t)$ for the microsatellite instability pathway as shown in Figure 1. This notion of a birth death process preserves the Markovian nature of the state transitions. The birth, death, and mutation processes are all independent of one another. Additionally, each cell will behave independently of another cell. This logic was adopted and edited from Tan's stochastic modeling in [26]. Note that a Feller-Arley or Glompertz birth death process can be applied instead of a generalized one that inherently fits stochastic processes, but for the sake of simplicity and retaining the focus on diet we use the notion of a generalized birth death process.

Our model is a continuous time Markov chain. For example, in the CIN pathway, we know that we are accounting for the transitions between states as the APC, KRAS, and p53 mutations, respectively. Thus this yields a Markov chain matrix (triangular) that represents any possible case at time t of either pathway. This and the birth-death process to represent cells replicating and dying are the two key pillars of our model. We generate the rates of the different types of events (mutation, birth, death) by multiplying them to the update matrix. We calculate the total rate of an event happening, r, with the time to the next event being an exponential random variable with rate r. Then we generate a probability vector for each type of event and sample what kind of event occurred. Using this information we are able

to store the updated time and state variables in a new row of the update matrix. Finally we generate the time in which the first cell reaches the final state $\boldsymbol{T}$. From here we are able to plot the data, analyze the movement of cells through the stages and perform statistical testing for interpretation.

For the CIN pathway, accepted state-transition models will almost always include the TSGs APC and p53 and the oncogene KRAS. Note that in some models such as that in [25] the protein Smad4 is considered as an intermediate step, and in [4] the gene PIK3CA is considered. In many instances, other genes that are not as consistently mentioned are grouped together as "other genetic events" such as that in [16]. Recall that APC, KRAS, and p53 are the most commonly mutated in CRC [17] and are the most tested for dietary influences in literature [2], [21], [19], [20], [3]. For these reasons, we only consider these three genes as the transitional steps in our model of the CIN pathway.

In terms of diet, we use the mutation rates for each gene as a way to represent an acceleration or deceleration of carcinogenesis in colorectal cancer. As we have covered, the explicit relation between a mutation rate of a cell and dietary habits is not known. However, the variation of rates and range of mutation rates has been discussed and we will apply such knowledge to our model. A mutation rate in the context of tumor suppressor genes and oncogenes is said to vary by $\mathbf{10^2}$ [4]. In a study of mutations and selection in sporadic cancers, results in [14] suggest that there is a "set" mutation rate and once two MMR (CIN) mutations occur, then the original set rate increases by some factor. In these results, the rate of increase was given by being in the range of $\mathbf{10}$ or $\mathbf{10^4}$. Vogelstein's findings of rate variation fall into this range as well. We can equate this statement to having the two APC mutations occur and then have the KRAS and p53 mutations affected thereafter.

Thus we make the following assumptions about mutation rate variation and diet in

our CIN model: we consider an extreme case of dietary habits and increase or decrease the mutation rate for KRAS of $\mathbf{10^4}$, and the rate for p53 by $\mathbf{10^2}$ from the Constant Rate Model. We will also create two time-varying models, one with random oscillation added to $\boldsymbol{\alpha_2}$ and $\boldsymbol{\alpha_3}$, and another with random oscillation added solely to $\boldsymbol{\alpha_3}$. The former represents a conservative version of varying levels of red meat, vegetable, and alcohol intake, and the latter focuses on red meat intake. To define these scenarios, consider an individual with frequent bouts of high red meat and alcohol intake over their lifetime (what some may call a "poor" diet) and a lack of vegetables, with realistic recovery periods in between that consist of a more balanced diet.

With KRAS-related dietary factors identified, we consider the alternate pathway of CIN progression considered in [6]. Recall that most CIN cancers have mutations occur in APC, KRAS, and p53 in that order. The results in [6] proposed that for a third of the occurrences of CIN carcinogenesis, KRAS preceded APC. We want to test using the proposed variation factors in the above paragraph to test if diet would affect the possibility of a KRAS preceding an APC mutation and if this would yield a faster progression to a carcinoma or prevent a carcinoma from developing in an individual's lifetime, as suggested in [17]. The probability of one CIN progression versus another would also be represented in the Markovian matrix built in to our simulation. Note that it is suggested that there are other sequences for CIN carcinogenesis besides having mutations in APC first, KRAS second, and p53 last, and the case with APC and KRAS switched. These include progressions where p53 precedes APC and so forth. However, these are very rare (less than 3% likely) [6].Thus, we exclude the rest of the cases besides the sequences of APC, KRAS, p53 and KRAS, APC, p53. From here, the first two gene mutations will act as a toggle switch based on the probability matrix and dietary factors.

To explore the factor of alcohol on the MSI pathway, we will accelerate $\boldsymbol{\alpha_2}$ by a factor of 20% and the subsequent mutation rates $\boldsymbol{\alpha_3}$ and $\boldsymbol{\alpha_4}$ by 10%, mimicking the accelerations

proposed for the CIN pathway by Vogelstein, but in smaller shifts. We are using smaller factors of acceleration in this case because there are no exploratory studies regarding the direct affect that diet may have on the mutation rates of the MSI pathway. Since such options have not been accounted for, the experimental data we generate from this pathway will serve as a clue to confirm or deny a potential link between the MSI pathway and diet despite its origins being less "sporadic" than that of the CIN pathway.

Assuming 1000 cells per crypt at $\mathbf{10^7}$ crypts in the colon, we have an estimate of $\mathbf{10^10}$ cells in the colon [16]. For both CIN and MSI, we use a base model that is a direct simulation of the model given in [25]. From here, we expand on this and create a simulation of the toggle-switch alternate CIN pathway, the KRAS-first pathway, time varying models, and accelerated models in CIN. We also create an exploratory accelerated model for MSI. This in turn will make for a comprehensive study in simulating a human colon and several different scenarios given by an individual's diet. From here we use the multistage model via the Gillespie Stochastic Simulation Algorithm and keep track of years it takes to reach the first cell in the final tumor stage in any case.

4 Results

All simulations were run for $\boldsymbol{N = 1,000,000}$ normal cells and $\boldsymbol{k = 100}$ simulations each per scenario. The number of normal cells set is an estimate of the amount of cells in a fraction of the real human colon. As we know, the real number of cells in the colon is ten billion. Since this would lengthen computation time of the simulation, we conducted a few benchmarking simulations with ten billion cells to verify the orders of magnitude in order to create a real time scale for the scenarios. The chosen amount of simulations is to generate enough data for proper statistical analyses. We built the model to use real units of time, in this case years. Thus in a square centimeter of a colon, which is about what one

million cells represents, we are able to scale and create a real time frame of how long it would take the individual to develop cancer.

4.1 Chromosomal Instability Pathway

4.1.1 Constant Rate Model

For the constant rate model for the CIN pathway, we lifted the mutation, birth, and death rates from Tan & Yan's 2010 study [25]. No additional adjustments were made to them. We classified three birth events, three death events, and five mutation events for this pathway - making the pathway the standard five stage, APC-first common occurrence of colorectal cancer. This will be used as our base control model for comparison to the other CIN models listed below.

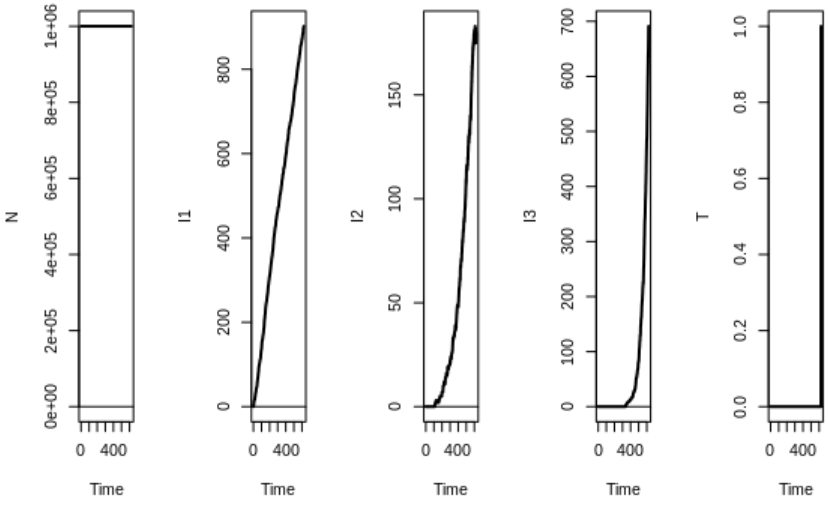

Figure 4: Movement of cells through states for CIN Constant Rate Model, n = 1000000 cells, k = 100 simulations

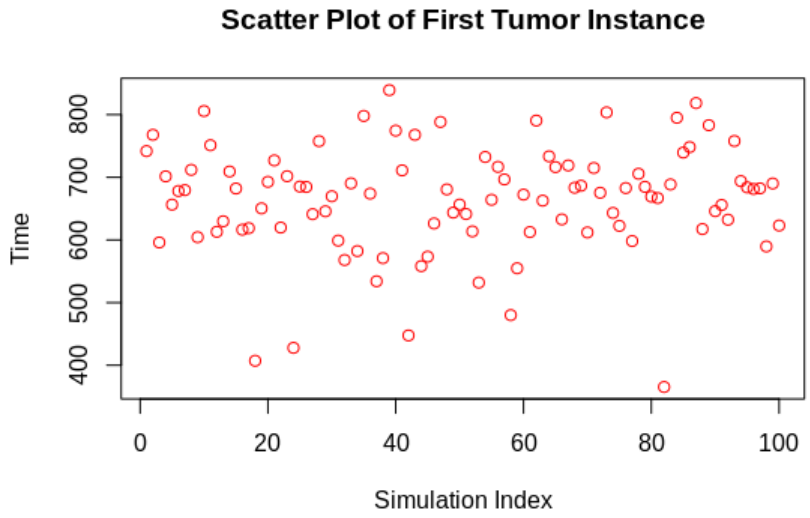

Figure 5: Scatter plot of first tumor instance for CIN Constant Rate Model, n = 1000000 cells, k = 100 simulations

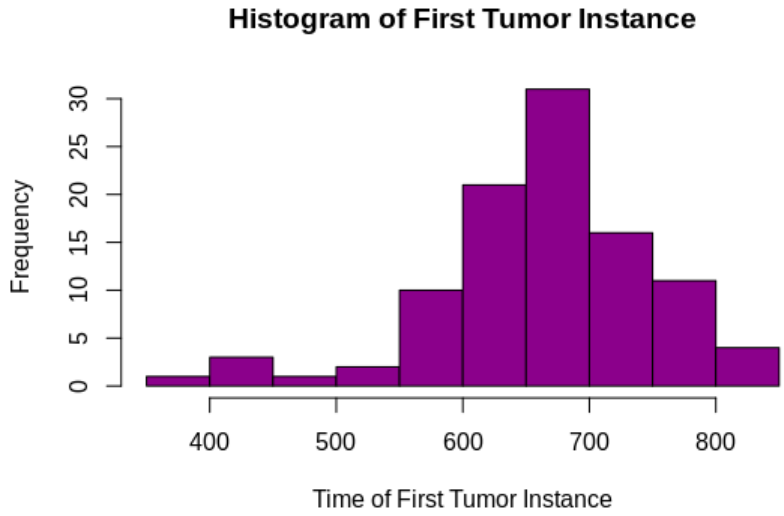

Figure 6: Histogram of First Tumor Instance for CIN Constant Rate Model, n = 1000000 cells, k = 100 simulations

Our mean time to reach the first tumor cell was $\mathbf{664.681}$. Thus, with one million cells, it takes on average this long in years to reach the tumor state. If you scale this down by orders of magnitude, in this case $\mathbf{10^4}$ as this would place us near the number of cells in

the real human colon, then this mean would actually be close to around $\mathbf{100}$ years, meaning that the individual would most likely not get cancer in their lifetime. This correlates with real life occurrence. It is also important to note the distribution here, which appears to be an almost gamma-type distribution.

4.1.2 Toggle Switch Model (KRAS First)

To simulate the KRAS-first version of the CIN pathway, we rearranged the order of the pathway to use $\boldsymbol{I_3}$ as the first state, such that the pathway in sequential order is as follows: $\boldsymbol{N}$, $\boldsymbol{I_3}$, $\boldsymbol{I_1}$, $\boldsymbol{I_2}$, $\boldsymbol{T}$. This is so that the first mutation rate occurring is $\boldsymbol{\alpha_2}$ which corresponds with the KRAS mutation rate. Additionally, mutation rates were increased by a factor of $\mathbf{10^2}$ for each so that the KRAS-first pathway would be equalized in comparison with the APC-first pathway.

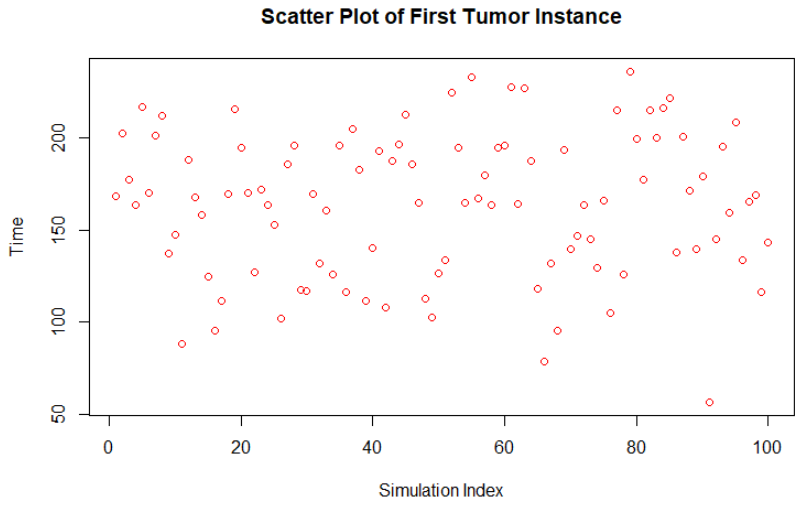

Figure 7: Scatter plot of first tumor instance for CIN Constant Rate Model, n = 1000000 cells, k = 100 simulations

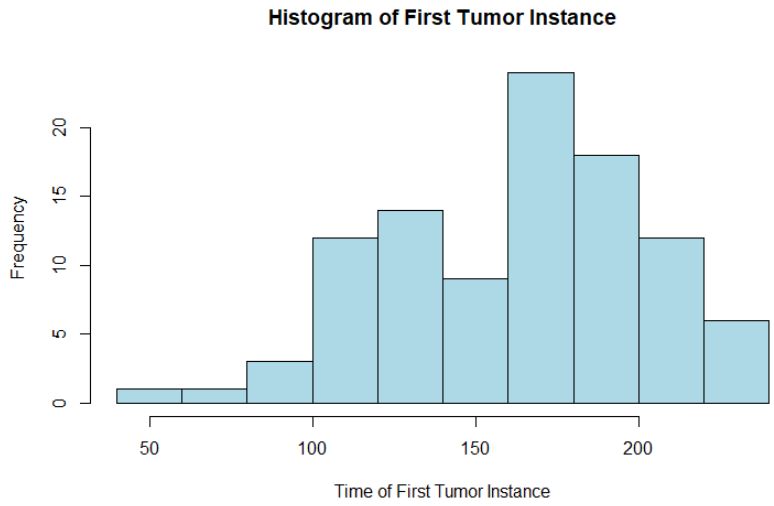

Figure 8: Histogram of First Tumor Instance for CIN Constant Rate Model, n = 1000000 cells, k = 100 simulations

For the KRAS-first CIN pathway, our mean time to reach the first tumor cell was $\mathbf{163.36}$. We use this scenario to explain our dietary factors that have direct connections with KRAS. Say for instance, an individual has extremely low vegetable intake paired with extremely high red meat/heme intake throughout their lifetime. This result suggests that this could speed up the CIN-type carcinogenesis and change the order of the pathway as a whole.

4.1.3 Time Varying Model

Our time varying model added random oscillation to mutation rates to $\boldsymbol{\alpha_2}$ and $\boldsymbol{\alpha_3}$. This represented a mild fluctuation in red meat and vegetable intake over time, since these mutation rates are for KRAS and p53. To give a more realistic sense of this scenario, assume an individual frequently eats high amounts of red meat and lacks vegetables, however, they have off periods where they opt for a better diet in between, one with higher vegetable and less red meat intake. We consider this to be a mild case of poor dietary habits. The rates

were varied by creating a low-end constant rate by multiplying the rate by $\mathbf{0.1}$, creating a high-end rate by multiplying the rate by $\mathbf{1.9}$, and then drawing a uniform random variable over time from this range for both $\boldsymbol{\alpha_2}$ and $\boldsymbol{\alpha_3}$.

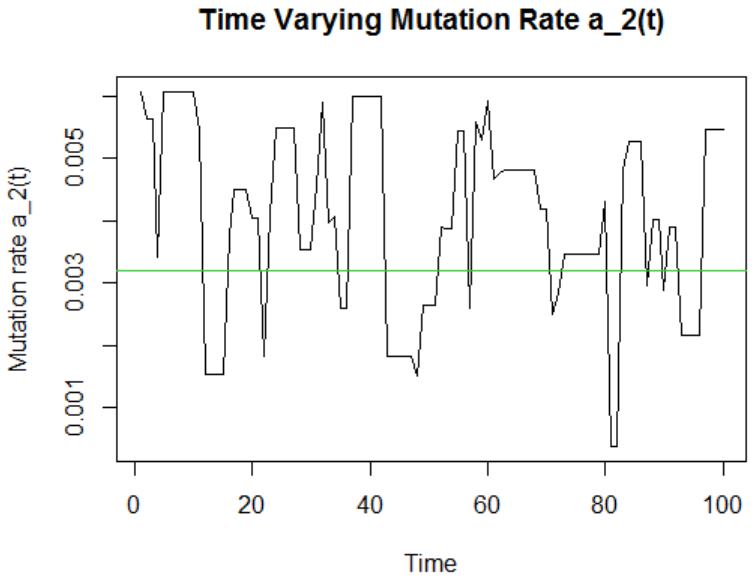

Figure 9: Time Varying Rate for $\boldsymbol{\alpha_2}$

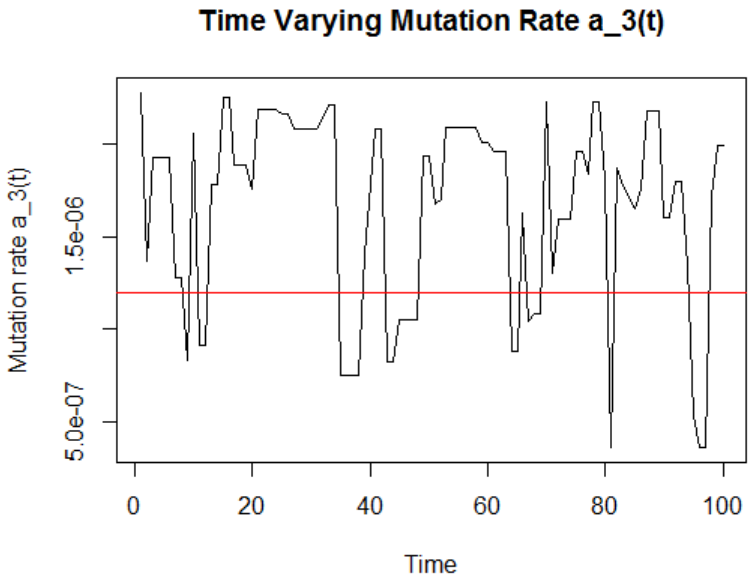

Figure 10: Time Varying Rate for $\boldsymbol{\alpha_3}$

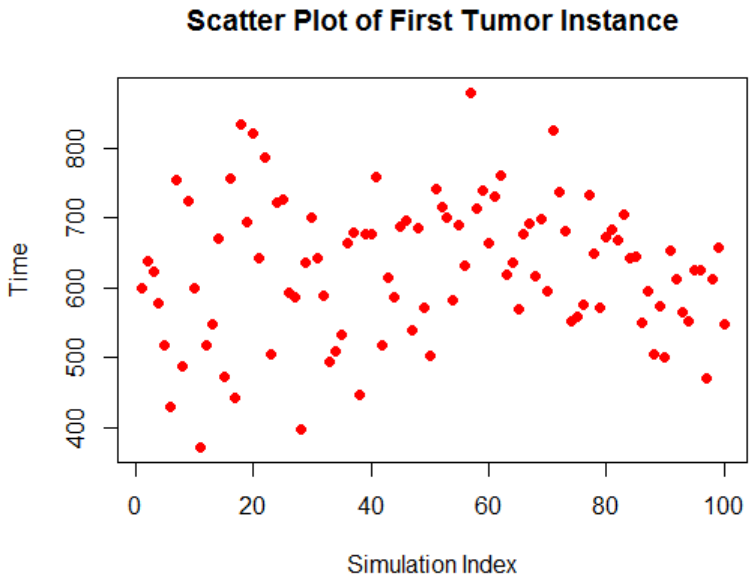

Figure 11: Scatter plot of first tumor instance for CIN Time Varying Model, n = 1000000 cells, k = 100 simulations

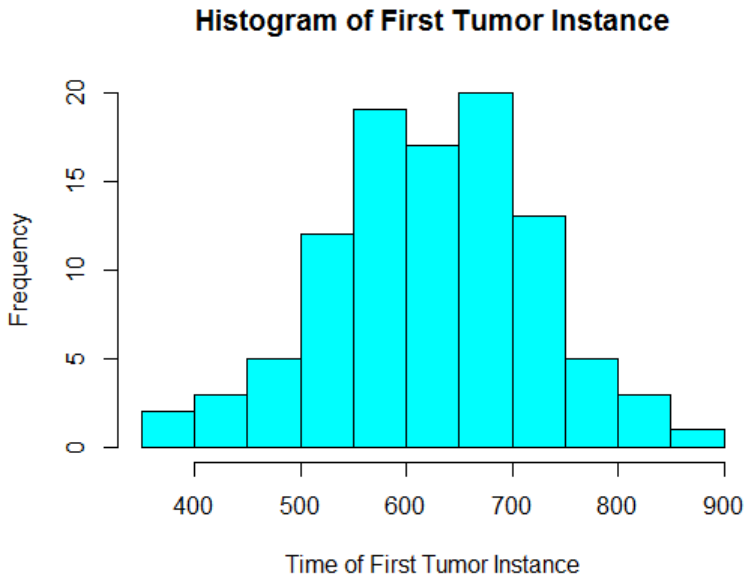

Figure 12: Histogram of first tumor instance for CIN Time Varying Model, n = 1000000 cells, k = 100 simulations

As we can see, the distribution is a normal one, with a mean time to the first tumor

state of $\mathbf{626.06}$. Clearly this is not too far off from the Constant Rate Model. To focus on a fluctuating diet in terms of solely red meat, we use the same methods but only apply a time-varying rate to p53's mutation rate. We obtain the following results:

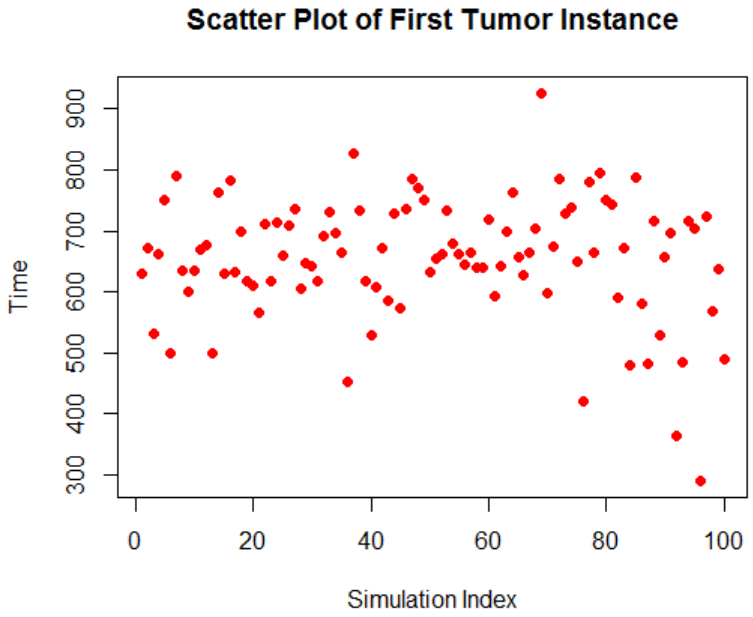

Figure 13: Scatter plot of first tumor instance for CIN Time Varying Model for $\boldsymbol{\alpha_3}$, n = 1000000 cells, k = 100 simulations

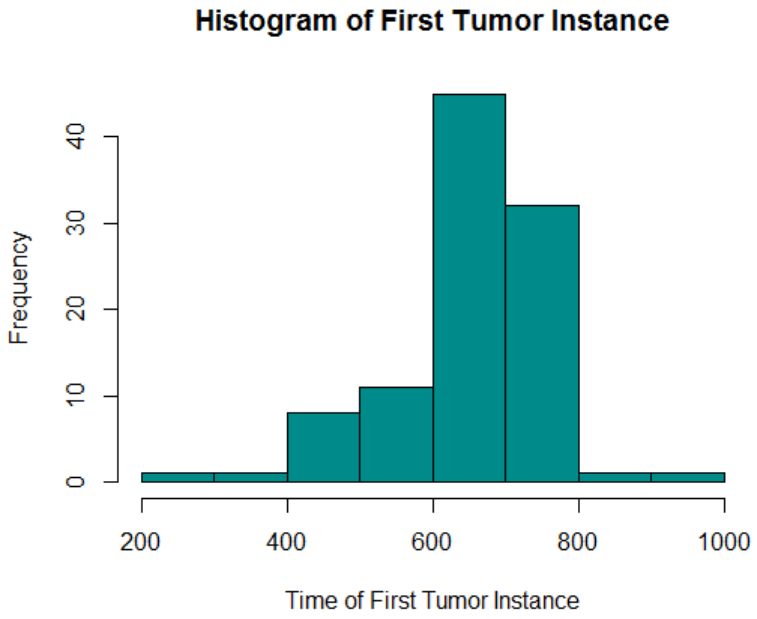

Figure 14: Histogram of first tumor instance for CIN Time Varying Model for $\boldsymbol{\alpha_3}$, n = 1000000 cells, k = 100 simulations

We obtain a mean time of $\mathbf{654.777}$, even closer to our Constant Rate model.

4.1.4 Accelerated & Rate Model

We created an accelerated version of the CIN constant rate model by increasing the mutation rate of KRAS/the intermediate state ($\boldsymbol{\alpha_2}$) by $\mathbf{10^3}$ and the mutation rate for p53 ($\boldsymbol{\alpha_3}$) by a factor of $\mathbf{10}$. This ensures that we are creating a realistic, cascade-like scenario of an extremely poor diet represented by such rates. In this case, consider an individual with longtime and frequent use of red meat and alcohol, while having a diet almost completely devoid of vegetables.

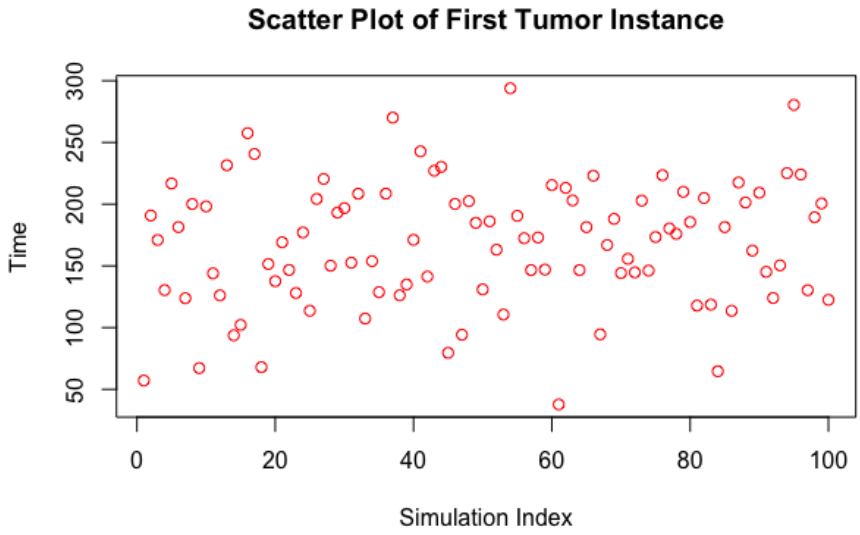

Figure 15: Scatter plot of first tumor instance for CIN Accelerated KRAS/p53 Model, n = 1000000 cells, k = 100 simulations

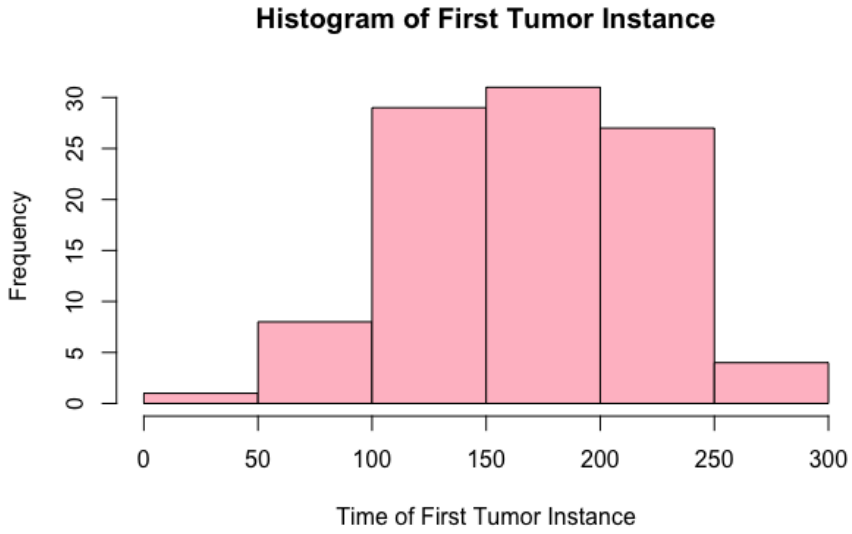

Figure 16: Histrogram of first tumor instance for CIN Accelerated KRAS/p53 Model, n = 1000000 cells, k = 100 simulations

As expected, this scenario yields a much quicker mean time to the first tumor cell. During this simulation, we obtained an average time of $\mathbf{167.649}$. This in turn shows

that more rapid mutating at a molecular level will certainly increase the progression to carcinogenesis, even if the mutation rates remain rather small.

4.2 Microsatellite Instability Pathway

4.2.1 Constant Rate Model

Similarly to our CIN model, we again adopt the rates given in Tan & Yan's 2010 study and create a six stage model for the MSI pathway [25]. In this case we have four birth and death events respectively and five mutation events. This is the standard, accepted version of the MSI pathway with no adjustments made to the rates and will act as a control group just like in the case of CIN.

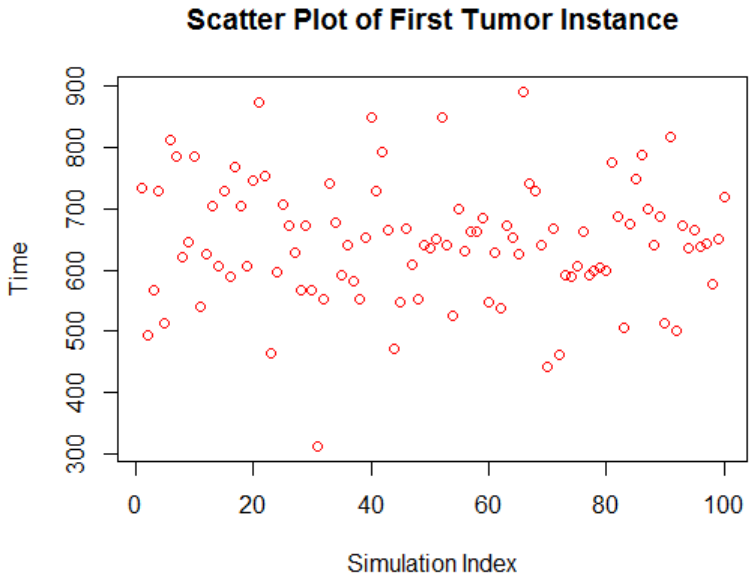

Figure 17: Scatter plot of first tumor instance for MSI Constant Rate Model, n = 1000000 cells, k = 100 simulations

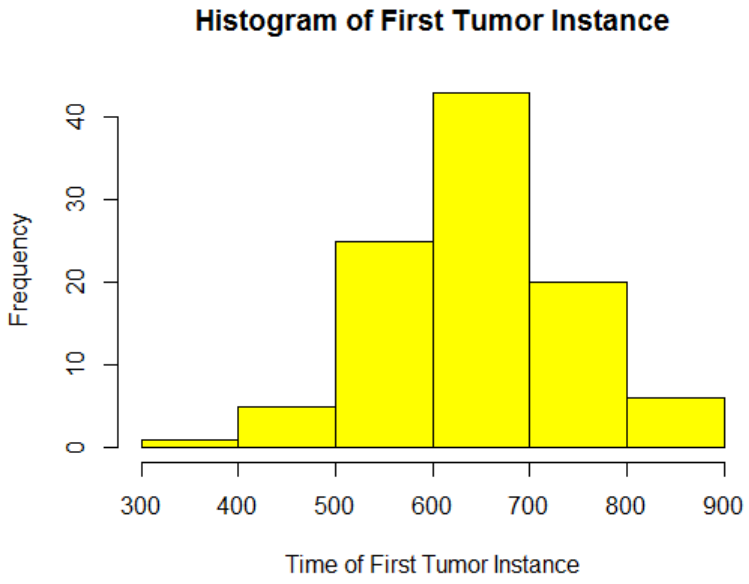

Figure 18: Histrogram of first tumor instance for MSI Constant Rate Model, n = 1000000 cells, k = 100 simulations

Note the clear normal distribution. Additionally, the mean time to the first tumor cell was $\mathbf{646.232}$ in this case which is close to our CIN model. Thus this confirms that both control groups behave so that they have a similar end result and thus can be used as a means of comparison.

4.2.2 Accelerated Rate Model

In the case of an accelerated MSI pathway, we consider a diet consisting of excessive and frequent alcohol consumption. We count this as another ”extreme” scenario in which an individual has an undoubtedly poor diet. In this case we increased $\boldsymbol{\alpha_2}$ by a factor of $\mathbf{10^3}$ and the two following mutation rates by a factor of $\mathbf{10}$, similar to the CIN accelerated model. Again these rates were chosen in order to exhibit real life extreme fluctuations of mutation rates. $\boldsymbol{\alpha_2}$ in particular was chosen as the base rate to be increased because this is the mutation rate for KRAS in the MSI pathway in accordance with the base model.

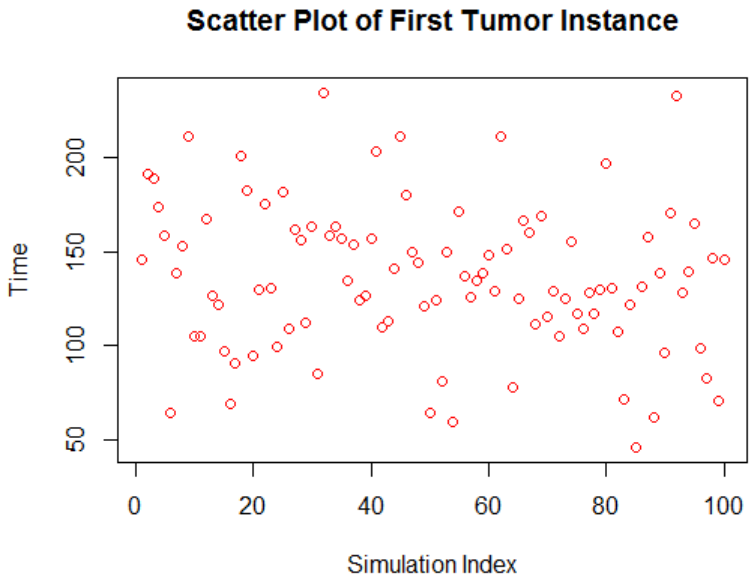

Figure 19: Scatter plot of first tumor instance for Accelerated KRAS Model, n = 1000000 cells, k = 100 simulations

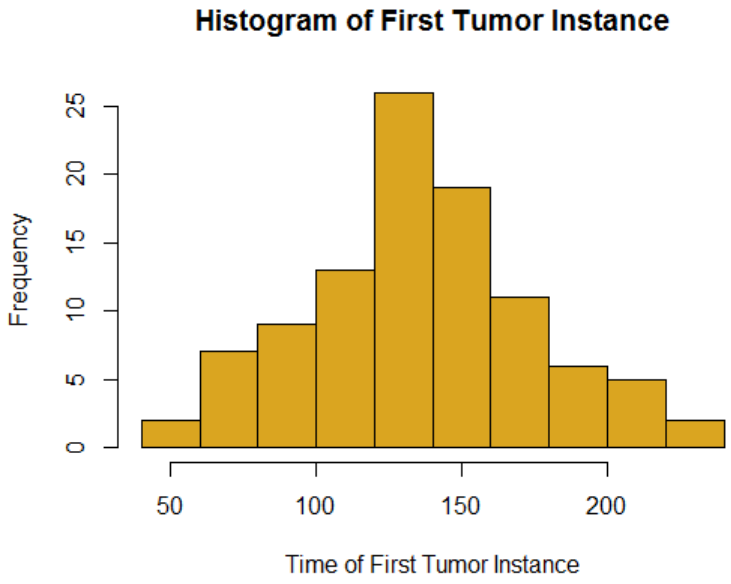

Figure 20: Histrogram of first tumor instance for MSI Accelerated KRAS Model, n = 1000000 cells, k = 100 simulations

In this case we observed a mean time of $\mathbf{135.87}$. The shorter time length corresponds

with real life MSI progression and confirms the findings of Slattery et. al - KRAS-linked mutations brought on by alcohol could speed up the process of carcinogenesis [20].

5 Discussion

5.1 Summary

Table 2: Quartile Summary of Results.

Case	Min	1st Quartile	Median	3rd Quartile	Max
CIN Constant	365.278	619.523	676.696	712.592	839.105
CIN KRAS	56.918	133.136	166.36	194.774	235.605
CIN TV (both)	371.059	563.266	633.291	692.425	879.283
CIN TV (a3)	289.938	615.22	662.0199	719.259	926.305
CIN Accl	37.754	130.827	171.757	202.910	293.862
MSI Constant	312.016	591.168	643.166	703.770	892.084
MSI Accl	45.91	111.027	133.158	159.025	234.751

5.2 Conclusions & Statistical Analyses

From the results we have obtained, we can conclude that a mildly poor diet, or some fluctuations towards a bad diet, will not affect one's progression to the final state of carcinogenesis, assuming the CIN pathway. There is a 6% increase in the time it takes to the first tumor cell in the Time Varying Model as compared to the Constant Rate Model. However, the time to the first tumor cell is 75% slower in the Constant Rate Model than in the accelerated model for the CIN pathway. In comparing to the KRAS-first model, the Constant Rate Model was again about 75% slower. Thus, more extreme changes in diet will truly accelerate the CIN pathway.

As for the MSI pathway simulations, we can say the same - extreme excess of alcohol consumption will accelerate the progression to carcinogenesis. More specifically, the time

to the first tumor cell is 80% slower in the Constant Rate Model than in that of the Accelerated Model. At this point, we can infer that more extreme poor dietary habits, specifically those involving *consistent* and high consumption of red meat and alcohol, will certainly accelerate the progression of CIN and MSI-type colorectal carcinogenesis as a whole on a molecular scale. This is specifically via the corresponding mutation rates.

Additionally, performing an unpaired t-test comparing the Constant Rate CIN Model to the Time Varying Model incorporating both $\boldsymbol{\alpha_2}$ and $\boldsymbol{\alpha_3}$ yields a t-value of $\mathbf{2.9452}$. Assuming $\boldsymbol{p = 0.05}$, we reject the null hypothesis and still conclude that there is significant difference between the two models. In comparing the Constant Rate Model against the Accelerated Model, we have a t-value of $\boldsymbol{t = 49.9277}$, obviously allowing us to reject the null. Against our other extreme scenario, the KRAS-first model, the test statistic was also a very high value of $\boldsymbol{t = 53.0732}$ which of course gives the same conclusion. Interesting, though, is the result for the Time Varying Model for strictly $\boldsymbol{\alpha_3}$, which yielded a test statistic of $\boldsymbol{t = 0.7541}$. With $\boldsymbol{k = 100}$ simulations, we fail to reject the null. In terms of diet, this could mean that one dietary factor with mildly increased or fluctuating intake is not enough to completely change the outcome of colorectal carcinogenesis. A t-test was performed to compare the MSI Constant Rate and Accelerated Rate Models as well. This had similar results to our CIN pathway with a test statistic of $\boldsymbol{t = 47.7837}$. The test statistics of our extreme diet scenarios indicate highly statistically significant results. This further validates our other findings in regards to poor and consistent dietary choices.

As we can see, there is a slight shift in what should be a normal distribution for the Constant Rate CIN Model. Performing a Kolmogrov-Smirnov Test yielded a KS test statistic of $\boldsymbol{D = 0.089965}$ and a p-value of $\boldsymbol{p = 0.3932}$, which confirmed that the Constant Rate model did not differ from a normal distribution. Performing this test on the KRAS-first pathway yielded a similar result of $\boldsymbol{D = 0.089676}$ and $\boldsymbol{p = 0.3972}$. As for the Time Varying Models, we obtained $\boldsymbol{D = 0.096665}$ and $\boldsymbol{p = 0.3075}$ for the model

accounting for two mutation rates, but $\boldsymbol{D = 0.13823}$ and $\boldsymbol{p = 0.04379}$ for the model that varied $\boldsymbol{\alpha_3}$ only. As with the t-tests for this model, this result is variant of the rest within the CIN models. This is not normally distributed, therefore the KS test was used again to test for a gamma distribution, in which we obtained $\boldsymbol{D = 0.12014}$ and $\boldsymbol{p = 0.1115}$. This is our sole model that exhibits a gamma distribution. The CIN Accelerated Model exhibited results of $\boldsymbol{D = 0.051401}$ and $\boldsymbol{p = 0.9543}$. The MSI Constant Rate Model gave a test statistic of $\boldsymbol{D = 0.066258}$ with $\boldsymbol{p = 0.7723}$ and the accelerated version gave $\boldsymbol{D = 0.046474}$ and $\boldsymbol{p = 0.9822}$. Thus, we can assume that the rest of the models are normally distributed, verified as such by the KS test, with fairly good fits for each as the test statistics are values fairly close to zero.

Most of our scenarios exhibited similar results, with our more extreme dietary scenarios having the highest statistical significance and consistent normal distributions when testing the time it took to get to the first tumor cell. Specifically, the pathways can be significantly accelerated in its progression by high and consistent intake of red meat and alcohol. Colorectal carcinogenesis can also be accelerated by a significantly low vegetable intake over one's lifetime, as verified by our KRAS-first results. Due to the outcome of our Time Varying Models, we can say that even at a milder scale, a combination of poor dietary factors that can directly affect two or more mutation rates can still accelerate the progression, but by a very small factor. However, because of our results in the Time Varying $\boldsymbol{\alpha_3}$ scenario, we can reasonably conclude that mild fluctuations in red meat alone will not accelerate carcinogenesis - a finding that explains previous results in literature about the questionable significance of red meat intake and CRC.

5.3 Future Work

Since we observed some close results indicating that a combination of dietary factors may continue to change the results, in the future we may want to incorporate even more dietary factors with more fluctuations in mutation and possibly proliferation rates along these pathways. In this project, based on the literature we expected to not have significant results for the MSI pathway in regards to diet. However, given the similarity between the various models in both pathways, and the confirmed effect with accelerated mutation rates, perhaps more exploratory studies on MSI and diet will be useful to create a more comprehensive view on CRC, despite the fact that some literature believes MSI to be solely epigenetic in behavior.

In general, the next step in this study would be refining simulations to account for the entire colon. Using external packages in R software, such as those that speed up the Gillespie Stochastic Simulation Algorithm, would be ideal for this matter. The simulations can be further adjusted to account for the dietary ”combination” scenarios. Additionally, the established factors such as acceleration and random oscillation can be applied to more rates than the ones that were studied for these models.

References

[1] A. Conlin, "The prognostic significance of K-ras, p53, and APC mutations in colorectal carcinoma," Gut, vol. 54, no. 9, pp. 1283–1286, 2005.

[2] A. El Asri, B. Zarrouq, K. El Kinany, L. Bouguenouch, K. Ouldim, and K. El Rhazi, "Associations between nutritional factors and KRAS mutations in colorectal cancer: a systematic review," BMC Cancer, vol. 20, no. 1, 2020.

[3] A. M. Gilsing, F. Fransen, T. M. de Kok, A. R. Goldbohm, L. J. Schouten, A. P. de Bruine, M. van Engeland, P. A. van den Brandt, A. F. de Goeij, and M. P. Weijenberg, "Dietary heme iron and the risk of colorectal cancer with specific mutations in KRAS and APC," Carcinogenesis, vol. 34, no. 12, pp. 2757–2766, 2013.

[4] B. Vogelstein, N. Papadopoulos, V. E. Velculescu, S. Zhou, L. A. Diaz, and K. W. Kinzler, "Cancer Genome Landscapes," Science, vol. 339, no. 6127, pp. 1546–1558, 2013.

[5] B. W. Stewart and F. T. Boseman, "Chapter 5.5: Colorectal Cancer," in World Cancer Report 2014, IARC Press, 2014, pp. 392–402.

[6] C. Paterson, H. Clevers, and I. Bozic, "Mathematical Model of Colorectal Cancer Initiation," Proceedings of the National Academy of Sciences of the United States of America, vol. 117, no. 34, 2020.

[7] "Colorectal Cancer Statistics: How Common Is Colorectal Cancer?," American Cancer Society, 2021. [Online]. Available: https://www.cancer.org/cancer/colon-rectal-cancer/about/key-statistics.html.

[8] D. Wodarz and N. L. Komarova, Dynamics of cancer: mathematical foundations of oncology. Singapore: World Scientific Pub. Co, 2014.

[9] E. G. Luebeck and S. H. Moolgavkar, “Multistage carcinogenesis and the incidence of colorectal cancer,” Proceedings of the National Academy of Sciences, vol. 99, no. 23, pp. 15095–15100, 2002.

[10] E. Theodoratou, M. Timofeeva, X. Li, X. Meng, and J. P. A. Ioannidis, “Nature, Nurture, and Cancer Risks: Genetic and Nutritional Contributions to Cancer,” Annual Review of Nutrition, vol. 37, no. 1, pp. 293–320, 2017.

[11] F. A. Orlando, D. Tan, J. D. Baltodano, T. Khoury, J. F. Gibbs, V. J. Hassid, B. H. Ahmed, and S. J. Alrawi, “Aberrant Crypt Foci as Precursors in Colorectal Cancer Progression,” Journal of Surgical Oncology, vol. 98, no. 3, pp. 207–213, Jul. 2008.

[12] H. Jayaskara, E. Williamson, A. Hodge, and R. Macinnis, “Lifetime Alcohol Intake Is Associated with an Increased Risk of KRAS+ and BRAF-/KRAS- but Not BRAF+ Colorectal Cancer,” International Journal of Cancer, vol. 140, no. 7, Dec. 2016.

[13] H. Nguyen and H. Q. Duong, “The molecular characteristics of colorectal cancer: Implications for diagnosis and therapy (Review),” Oncology Letters, 2018.

[14] I. P. Tomlinson, M. R. Novelli, and W. F. Bodmer, “The mutation rate and cancer,” Proceedings of the National Academy of Sciences, vol. 93, no. 25, pp. 14800–14803, 1996.

[15] L. Lawrence, “What Impact Does Microsatellite Instability Have on Survival in Stage III Colon Cancer?,” Cancer Network. [Online]. Available: https://www.cancernetwork.com/view/what-impact-does-microsatellite-instability-have-survival-stage-iii-colon-cancer.

[16] M. A. Nowak, Evolutionary dynamics: exploring the equations of life. Cambridge, MA: Belknap Press of Harvard University Press, 2006.

[17] M. J. Schell, M. Yang, J. K. Teer, F. Y. Lo, A. Madan, D. Coppola, A. N. Monteiro, M. V. Nebozhyn, B. Yue, A. Loboda, G. A. Bien-Willner, D. M. Greenawalt, and T. J. Yeatman, "A multigene mutation classification of 468 colorectal cancers reveals a prognostic role for APC," Nature Communications, vol. 7, no. 1, 2016.

[18] M. L. Clapper, W.-C. L. Chang, and H. S. Cooper, "Dysplastic Aberrant Crypt Foci: Biomarkers of Early Colorectal Neoplasia and Response to Preventive Intervention," Cancer Prevention Research, vol. 13, no. 3, pp. 229–240, 2020.

[19] M. L. Slattery, J. S. Herrick, A. Lundgreen, F. A. Fitzpatrick, K. Curtin, and R. K. Wolff, "Genetic variation in a metabolic signaling pathway and colon and rectal cancer risk: mTOR , PTEN , STK11 , RPKAA1 , PRKAG2 , TSC1 , TSC2 , PI3K and Akt1," Carcinogenesis, vol. 31, no. 9, pp. 1604–1611, 2010.

[20] M. L. Slattery, K. Anderson, K. Curtin, K.-N. Ma, D. Schaffer, and W. Samowitz, "Dietary intake and microsatellite instability in colon tumors," International Journal of Cancer, vol. 93, no. 4, pp. 601–607, 2001.

[21] M. L. Slattery, K. Curtin, K. Ma, S. Edwards, D. Schaffer, K. Anderson, and W. Samowitz, "Diet, Activity, and Lifestyle Associations with p53 Mutations in Colon Tumors," Cancer Epidemiology, Biomarkers & Prevention, vol. 11, no. 6, pp. 541–548, Jun. 2002.

[22] M. Lüchtenborg, M. P. Weijenberg, A. F. de Goeij, P. A. Wark, M. Brink, G. M. Roemen, M. H. Lentjes, A. P. de Bruïne, R. A. Goldbohm, P. van 't Veer, and P. A. van den Brandt, "Meat and Fish Consumption, APCGene Mutations and hMLH1 Expression in Colon and Rectal Cancer: a Prospective Cohort Study (The Netherlands)," Cancer Causes & Control, vol. 16, no. 9, pp. 1041–1054, 2005.

[23] S. A. Frank, Dynamics of cancer: incidence, inheritance, and evolution. Princeton, NJ: Princeton University Press, 2007.

[24] S. D. Markowitz and M. M. Bertagnolli, "Molecular Basis of Colorectal Cancer," New England Journal of Medicine, vol. 361, no. 25, pp. 2449–2460, 2009.

[25] W. Y. Tan and X. W. Yan, "A new stochastic and state space model of human colon cancer incorporating multiple pathways," Biology Direct, vol. 5, no. 1, p. 26, 2010.

[26] W. Y. Tan, Stochastic models of carcinogenesis. Boca Raton, FL: CRC Press, 2020.

[27] Z. Kemp, A. Rowan, W. Chambers, N. Wortham, S. Halford, O. Sieber, N. Mortensen, A. von Herbay, T. Gunther, M. Ilyas, and I. Tomlinson, "CDC4 Mutations Occur in a Subset of Colorectal Cancers but Are Not Predicted to Cause Loss of Function and Are Not Associated with Chromosomal Instability," Cancer Research, vol. 65, no. 24, pp. 11361–11366, 2005.