

Adenomatous polyps and it's growth models



**ASSIGN
BUSTER**

Introduction

Adenomatous polyps, often called simply adenomas, are lesions that arise in the large intestine from epithelial cells. Though initially benign, it is possible for an adenoma to undergo carcinogenesis and develop into colorectal cancer, or CRC, which is a leading cause of morbidity and mortality among cancers in the United States and Europe. Several methods of screening exist to detect adenomas and early-stage cancer, including fecal occult blood testing, sigmoidoscopy, and colonoscopy. Of particular interest, previous studies have shown that screening programs that involve colonoscopy can lead to decreases in CRC risk while also being cost-effective.

Many questions remain, however, regarding the best ways to utilize colonoscopy screening programs, such as whether screening should begin at a younger age for certain groups that may be more at risk. Part of this uncertainty stems from gaps in knowledge concerning the growth and development of adenomas from their initial appearance to malignant tumors. Only a small percentage of adenomas will ever become malignant, but it can be difficult to identify exactly which of these lesions are candidates for carcinogenesis in an individual. Additionally, endoscopic detection of adenomas is limited, since only lesions that have grown to a certain size can be found. Smaller adenomas may therefore remain unidentified and could give rise to cancer. Addressing these issues may help improve screening programs for CRC.

Models are therefore of interest, as they can be used to provide insight to the growth and development of adenomas in a population. In Bavaria,

Germany, there exists population-wide data on outpatient colonoscopies performed between 2006 and 2009, which can be utilized in the construction of such models. Using these records, Kollings et al. have previously developed logistic regression models to determine the effects of age and sex on the risk of having advanced colorectal neoplasia at the time of colonoscopy. The results suggest that at any given age, men have a higher chance of having such a neoplasia at colonoscopy than women. The approach undertaken by Kollings et al. treats the data as prevalence data, providing information about colorectal neoplasia at a specific point in time. This paper presents other approaches to growth models using the same Bavarian colonoscopy data.

The data can be reinterpreted as current status data in order to create a different set of models. Current status censoring is used to reinterpret the data in this manner. At the time of colonoscopy, it is known whether or not an adenoma has reached a given threshold in its growth, such as reaching 10 mm in size. This feature is treated as an outcome variable and is given a value of 0 if the adenoma has not reached this point at the time of colonoscopy; if it has, the variable is given a value of 1. Although the exact time an adenoma transitioned to a specific stage in its growth is not known, censoring the data in this manner allows for hazard functions to nevertheless be used in constructing growth models. This provides information on the risk of an adenoma transitioning to another stage of growth during a period of time, as some time periods may have higher risks of specific transitions than others. Current status censoring is therefore applied to the colonoscopy data to investigate these risks. The censored data is modeled using generalized

additive models, or GAMs, which are useful for detecting non-linearities in the data.

While both logistic regression and current status analysis can provide information about adenoma development, they are based on general statistical methodology rather than on the biological mechanisms driving adenoma development. However, biological models exist that consider the organic initiation, growth, and transition to malignancy of adenomas. One such model is the multistage clonal expansion, or MSCE, model, which describes relevant stochastic processes that lead to the development of CRC. Jeon et al. have developed mathematical expressions for the detectable pre-malignant adenoma size distributions based on this model. Here, “detectable” refers to being identifiable endoscopically, such as with a colonoscopy. The equations presented by Jeon et al. are used to derive a model that utilizes two parameters of unknown value. Using a grid search, maximum likelihood estimation is performed to determine plausible values of the parameters. The model is then applied to the Bavarian dataset using the estimated parameter values.

Methods

CHI Dataset

The study is based on outpatient colonoscopies that were performed in Bavaria, Germany between January 2006 and December 2009. The database from which the data was drawn contains no individual-related data, and it meets standards for data protection as approved by the regulatory authorities in Bavaria.

There exists in Bavaria a compulsory health insurance system, or CHI system, which constitutes funding for health services. The Kassenärztliche Vereinigung Bayern (KVB) is an organization that manages financial transactions between health care providers and insurance companies that participate in the CHI system. In the past, the KVB has conducted quality programs in regards to certain medical services. A program concerning colonoscopies performed in Bavaria was set up, and as a result, outpatient colonoscopies were fully documented by CHI-affiliated physicians for a period of time. These records constitute the database with which our studies are performed.

These outpatient colonoscopy records of CHI patients are stored in an electronic database. The database contains 1, 096, 230 records of colonoscopies that were performed between January 2006 and December 2009. Of interest are records in which there is no indication of prior CRC at the time of the colonoscopy. CHI insured individuals are entitled to two screening colonoscopies at an interval of 10 years upon reaching 55 years of age. Records which indicate a colonoscopy was performed for a purpose other than screening were excluded, providing 258, 116 records. Among those, records which showed a negative diagnosis for adenomas were excluded, meaning the remaining 66, 232 records indicate the screened patient was positively diagnosed with adenomas. We further excluded records in which the largest polyp was measured to be smaller than 0. 5 cm, which is a reasonable threshold for endoscopic detection. A total of 36, 391 records meet this criterion and are thus used in our analyses.

The recorded findings are based on the largest lesion found during the colonoscopy. Biological specimens were collected during colonoscopy in cases where cancer was suspected and were analyzed by local pathologists. Photo documentation of cecal landmarks was also conducted, as such documentation is required for financial reimbursement for the performed colonoscopy. Endoscopists who perform colonoscopies on patients who are insured in the CHI system must have documented experience of a minimum of 200 colonoscopies and 50 polypectomies in the past two years. Additionally, they must also continuously encounter a high volume of both procedures.

Current Status Analysis

As stated previously, applying current status censoring to the data allows for the use of hazard functions. Current status data can be modeled by a GAM using an extension of the proportional hazards model, as follows. If Y is a current status variable, t a monitored point in time, and X_i a set of covariates, then P_x can be defined as the probability that Y occurs after time t in a patient with X_i covariates; or written otherwise:

$$P_x = P(Y > t | X_i)$$

Following the model, P_x can further be rewritten:

$$\exp[-\Lambda_{X_i}(t)]$$

Here $\Lambda_{X_i}(t)$ is a hazard function for those with covariates X_i . The function can be expanded which allows P_x to be written as:

$$\exp[-\Lambda_0(t) * \exp(\beta_1 X_1 + \beta_i X_i)]$$

The function $\Lambda_0(t)$ is the hazard function for those in which X_i equals 0. An additional reiteration of P_x can be described as:

$$\exp\{-\exp[S_0(t) + \beta_1 X_1 + \beta_i X_i]\}$$

$S_0(t)$, which is equal to $\log(\Lambda_0(t))$, can be seen as an arbitrary increasing function of the covariate t . Written in this form, it becomes observable that a complementary log-log transformation of P_x would yield the linear predictor function:

$$\log(-\log(P_x)) = S_0(t) + \beta_1 X_1 + \beta_i X_i$$

Using R software, a generalized additive model can be constructed, and the coefficients β_i can be used to construct a type of survival curve using the c-log-log transformation. The curves illustrate the probability of remaining event free, i. e. not experiencing the outcome variable with the given covariates. When using the R function "gam," which is available with the package "mgcv," the c-log-log link must be specified, as well as the family "binomial." This stands to reason, as the outcome variable is coded as a binomial variable; 0 indicates the patient has not experienced the event, while 1 indicates the event has occurred. The next prerequisite to creation of a model is to choose a current status variable as an outcome, and to choose covariates that are of interest. The variables chosen for GAMs in this analysis are described below.

Previous studies have suggested that men are found to have advanced adenoma at a higher rate than women when adjusted for age, and although the risk of CRC is similar in both sexes over a lifetime, men tend to develop

CRC earlier than women. Thus, sex is an important predictor variable to include in the current status analysis. An additional important characteristic of adenomas is localization, i. e., where the adenoma is positioned in the colon. In the dataset, a patient's polyps may be described as proximal (right-sided colon), distal (left-sided colon), or both. Though evidence supports the notion that CRC outcomes can be improved through colonoscopy screening, several studies the U. S. also suggest that proximal and distal adenomas are affected differently by such programs. For these reasons, localization is included as a predictor variable in the current status analysis. However, previous analyses show that adenoma coded as " both" behave almost identically to those labeled " distal," and so the variable has been recoded so that " distal" also includes records that were previously classified as " both," making the variable binomial. A third covariate of interest is polyp shape. There are three shapes a polyp can be classified as in the data: flat, sessile, and peduncular. Previous calculations using the data set suggest that polyp shape may have a significant effect on adenoma growth, and thus it is included in this analysis.

The outcome variable that is used in the current status analysis is advanced adenoma. The classification of an adenoma as " advanced" was given by physicians involved with the performed colonoscopy using common criteria. Common criteria stipulate that an adenoma is advanced when it is larger than 10 mm in diameter, has high grade dysplasia, or a combination of these features. The variable for advanced adenoma is coded as a binomial, with 0 indicating that there is no advanced adenoma, and 1 indicating presence of advanced adenoma. Thus, no further action is required to reinterpret the

outcome variable as current status data. Since screening colonoscopy in the CHI system is performed only in individuals who are at least 55 years of age, the variable that is used for time t in the model is patient age minus 54. This takes into account the explicit exclusion of records of those younger than 55.

The Multistage Clonal Expansion Model

Model Overview

As precursors to CRC, adenomas undergo their own growth process before becoming an adenocarcinoma. The MSCE model describes the steps colonic stem cells undertake in the formation and subsequent growth of adenoma. A core tenet of the model is that adenomas form from colonic stem cells that have suffered at least two rare epi-genetic rate-limiting events. Although the model framework allows for more than two events to be considered, evidence suggests that only a small number of rate-limiting steps are needed for the formation of adenoma. The events that occur before the last step needed in a model are called pre-initiation events. For example, if three rate-limiting events are needed at the cellular level for adenoma formation and expansion, then the first two events are considered pre-initiation events.

If $K (\geq 1)$ is the number of pre-initiation events in the model, then rate-limiting events occur at a stage $k (k= 1, \dots, K)$. A cell at the k th stage has therefore suffered k events but is not yet capable of undergoing clonal expansion. However, it can undergo asymmetric division to yield two daughter cells, one of which acquires a rate-limiting mutation. In the MSCE model, there are two variations in which this can occur. In the first path, one daughter cell remains at stage k , and one daughter cell is at stage $(k + 1)$. This path is called Poisson process division because it can be mathematically

modeled as a Poisson process. In the second variation, a k -stage cell also yields a daughter cell at stage $(k + 1)$; however, the other cell undergoes differentiation and death. This type of asymmetric cell division in the model is referred to as Armitage-Doll division. In both cases, a stem cell may undergo clonal expansion if it has acquired $(K + 1)$ rate-limiting events. At this stage, it is considered to be initiated.

Dewanji et al. present several assumptions associated with the MSCE model in the context of CRC. At the cellular level, an adenoma is defined as the collection of all initiated cells, i. e. cells at stage $(K + 1)$, that are the progeny of one single cell that is at stage K . This also means that in the model, adenomas are allowed to be multi-focal; a stage K progenitor may experience transient amplification, resulting in sub-clones that together form the polyp known as adenoma. The rate-limiting events responsible for the development of the adenoma are assumed to be the biallelic inactivation of the APC tumor suppressor gene, though other tumor suppressor genes can also experience such inactivation. Though our model is not the same as that presented in Dewanji et al., it does assume the biallelic inactivation of a tumor suppressor gene to constitute the rate limiting-events. The microbiological processes presented here form the basis of another adenoma growth model.

Statistical Analysis

Based on the MSCE model and its stochastic processes in the development of CRC, Jeon et al. have derived equations that express the distribution of adenoma number and sizes for detectable adenomas. The equations they present take into account both the number of pre-initiation stages as well as

the prior presence of CRC. Of particular interest is equation 3.34 given by Jeon et al.. From this, we can derive an expression which gives the probability that at time t an adenoma with more than y_0 initiated cells has even more than $y_1 > y_0$ initiated cells by summing over all $n > y_1$ (or y_0), forming a quotient of these expressions for y_1 and y_0 , and taking the limit $\rho \rightarrow 0$, yielding:

$$P[Y(t) > y_1 | Y(t) > y_0] =$$

$$\omega^{(y_1 - y_0)} \cdot \frac{\Gamma(1 + y_1)}{\Gamma(1 + y_0)} \cdot \frac{\int_0^\infty t^{y_1} \cdot e^{-t} \cdot (1 - \omega \cdot e^{-t}) dt}{\int_0^\infty t^{y_0} \cdot e^{-t} \cdot (1 - \omega \cdot e^{-t}) dt}$$

$$\omega = (\alpha \cdot [\exp\{\gamma \cdot t\} - 1]) / (\alpha \cdot \exp\{\gamma \cdot t\} - (\alpha - \gamma))$$

This biomathematical model relies on two parameters: α , the cell division rate per year, and β , which is the cell inactivation rate per year. In the formal expression for the model, β is replaced by γ , which is equivalent to $\alpha - \beta$, and is the net clonal expansion per year. There are numerical restraints associated with the expression. When $\omega = 1$, the integrals in the equation are undefined. This is not unexpected, as it implies that $\gamma = 0$, meaning there is not null or negative clonal expansion of adenoma. Consequentially, since α and γ inherently influence the value of ω , their values should be restricted so that ω remains less than 1. Numerical calculations indicate that for γ larger than approximately 0.1, ω becomes too close to 1 to be numerically treatable.

The model expressions can be translated into R language, and the resulting code is as follows:

<https://assignbuster.com/adenomatous-polyps-and-its-growth-models/>

```
library(VGAM)
```

```
#
```

```
omega.rfc <-function(alpha, gamma, t)
```

```
{
```

```
return(alpha*(exp(gamma*t)-1)/
```

```
(alpha*exp(gamma*t)-(alpha-gamma)))
```

```
}
```

```
#
```

```
pAdenoLarger.alpha.gamma <-function(alpha, gamma, y.0, y.1, t)
```

```
{
```

```
omega <-omega.rfc(alpha, gamma, t)
```

```
return(omega^(-y.0+y.1)*
```

```
lerch(omega, 1, 1+y.1)/lerch(omega, 1, 1+y.0))
```

```
}
```

```
#
```

The package VGAM contains the necessary Lerch function. The algorithms require four variables: y_1 , y_0 , α and γ . These variables can be assigned values in R, and time t can be defined as age in years. When this is done, the

<https://assignbuster.com/adenomatous-polyps-and-its-growth-models/>

algorithms can be used to find $P[Y(t) > y_1 | Y(t) > y_0]$, the probability of an adenoma having more than y_1 initiated cells, given that it has y_0 initiated cells, for a range of values for t when the cell division rate is α and the net clonal expansion rate is γ . Maximum likelihood estimation can be used to determine the best values for the unknown parameters α and γ . A grid search will be used as the method of estimation. The resulting model with estimated parameters can then be applied to the Bavarian dataset. Based on results from the current status analysis, the effects of polyp shape and adenoma histology are of significant interest. Thus, the data is stratified by polyp shape and adenoma histology in the biologically-based adenoma growth model.

An exploration of the parameters y_1 , y_0 , α and γ of the MSCE-based growth model was conducted using R as described above. For each variable, a range of low values and a range of high values were selected, and the values were used to calculate $P[Y(t) > y_1 | Y(t) > y_0]$, the probability of an adenoma having more than y_1 initiated cells, given that it has y_0 initiated cells, for a range of values for t when the cell division rate is α and the net clonal expansion rate is γ . For each calculation, the range of values for t remained constant, 50 – 80 years of age.