Proceedings of the 2019 Winter Simulation Conference
N. Mustafee, K.-H.G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, eds.

THE EFFECT OF THE DISTRIBUTION OF THE INVERSE GROWTH RATE ON PANCREATIC CANCER PROGRESSION

Lena Abu-El-Haija

Department of Industrial Engineering
German Jordanian University
Amman Madaba Street, P.O. Box 35247
Amman, 11180, JORDAN

Julie S. Ivy
Osman Ozaltin

Department of Industrial and Systems Engineering
North Carolina State University
Daniels Hall, Lampe Drive
Raleigh, NC 27695, USA

Walter Park

Division of Gastroenterology and Hepatology
Stanford University Medical Center
291 Campus Drive
Stanford, CA 94305, USA

ABSTRACT

Pancreatic cancer is a low-incidence disease, where tumor progression studies using patient longitudinal data had limited sample sizes. Estimating the tumor inverse growth rate and its distribution are a challenge. Using a tumor progression model that incorporates the distribution of the inverse growth rate as the underlying assumption of the model, pancreatic cancer progression models were built assuming two distributions for the inverse growth rate: Uniform and Gamma. This study uses simulation to evaluate the effect of the tumor inverse growth rate distribution on the tumor progression models by examining tumor timelines. It was found that the tumor timeline is about nine months longer under the assumption that the inverse growth rate follows Gamma distribution. It was inconclusive whether tumor progression is faster or slower in older patients as the tumor progression models with the different underlying assumptions on the inverse growth rate yielded opposite results.

1 INTRODUCTION

Pancreatic Cancer (PC) is the fourth leading cancer cause of death in the United States, and it is projected to be the second by year 2030 (Rahib et al. 2014). Pancreatic cancer tends to be asymptomatic leading to late symptomatic detection often after the cancer has metastasized to other organs (Xu et al. 2014). As a result, metastasis is common in PC, accounting for up to 90% of its mortality and the dismal 5-year survival rate of 8% (Yachida and Iacobuzio-Donahue 2013). Initially, it was thought that PC is aggressive, fast growing in nature, and hence unlikely to be caught at an early stage. However, it was found that there is an average of 6.8 ± 3.4 years from the first parental clone to metastasis in the progression of PC (Yachida et al. 2010). The resulting time window allows for the discussion of pancreatic cancer screening to increase early detection before patients reach metastasis.

In order to produce effective screening policies, a pancreatic cancer progression model needs to be developed. Progression models exist with the general underlying assumption that growth is exponential

with a growth rate per unit time to describe the speed at which the tumor progresses. Plevritis et al. (2007) derived probability models of symptomatic detection and the transition between stages of cancer with an underlying assumption that the inverse growth rate is Gamma distributed (Plevritis et al. 2007).

There are studies that measure the doubling time of pancreatic cancer which can be linearly transformed to the inverse growth rate of the tumor (Furukawa et al. 2001; Sun et al. 2014). Furukawa et al. (2001) studied the pancreatic cancer tumor growth rate in nine patients before metastasis. Their sample yielded a doubling time of the tumor that ranged between 64 and 255 days (Furukawa et al. 2001). Sun et al. (2014) developed a mathematical model to replicate pancreatic cell mutations from the first lesion to cancer (Sun et al. 2014). Sun et al. (2014) used Yachida et al. (2010) patient summary data with their model and estimated the tumor doubling time to range between 84 and 105 days (Yachida et al. 2010; Sun et al. 2014). However, a distribution of the result could not be inferred due to the small sample of patients used to estimate tumor doubling time. This distribution can significantly influence the overall progression of the tumor.

There are studies that reveal the importance of understanding the growth rate of tumors. Mehrara et al. (2007) performed a Monte Carlo simulation to evaluate the current use of the tumor doubling time as the measure of tumor growth (Mehrara et al. 2007). The result of Mehrara et al. (2007) simulations indicate that there are advantages in using the reciprocal of the inverse growth rate instead of the tumor doubling time when describing the growth rate of a tumor (Mehrara et al. 2007). Weedon-Fekjær et al. (2008) estimated the growth rate of breast cancer using almost 400,000 women and evaluated that the growth rate follows a log-normal distribution (Weedon-Fekjær et al. 2008). The study compared their resulting tumor progression model with existing Markov models in the literature and found that their model has a higher predictive value of breast cancer progression (Weedon-Fekjær et al. 2008). This highlights the importance of accurately modeling the growth rate of a tumor. In breast cancer, there is a higher number of patients and screenings in effect where both lead to large databases that were used to estimate the growth rate of the tumor. However, pancreatic cancer is less prevalent with a shorter lifespan after detection due to the late symptoms. Therefore, the challenge of accurately estimating pancreatic cancer growth rate and its distribution is of higher importance until the distribution can be represented with patient data.

In previous work, we utilized the progression model presented by Plevritis et al. (2007) and adopted the findings of Furukawa et al. (2001) and Plevritis et al. (2007). In this paper we utilize our progression model to characterize the effect of the underlying assumption made about the inverse growth rate on tumor progression. We employ discrete-event simulation to quantify the difference between the models as a function of various underlying assumptions of the inverse growth rate.

2 METHODS

For this study, we utilized the progression model developed by Plevritis et al. (2007) (Plevritis et al. 2007). We modified the progression model by changing the underlying assumption of the distribution that inverse growth rate follows. We parameterized the models using patient level data from the Surveillance, Epidemiology, and End Results (SEER) databases for pancreatic cancer (National Cancer Institute 2015). We use the cancer stage definitions provided by SEER, where *local* stage cancer is one that is confined within the pancreas; *regional* stage cancer is one that has grown out of the pancreas into neighboring organs; and *distant* stage cancer is one that has metastasized and reached distant organs. We define *symptomatic detection* as the event of a cancer patient naturally showing symptoms due to the cancer. We created discrete-event simulation for each progression model created. The simulation models were implemented using Python 3.6. The remainder of this section summarizes the use of SEER data within the simulation models and the progression model by Plevritis et al. (2007) with our modifications.

2.1 Progression Model

The simulated progression models are discussed in detail in (Abu-El-Haija 2019). The framework of each progression model is characterized by three main events: (i) tumor volume at which symptomatic detection occurs; (ii) tumor volume at which the transition from the local stage to the regional stage occurs; and (iii) tumor volume at which the transition from the regional stage to the distant stage occurs. The base assumption of the models is that tumor volume grows exponentially according to:

$$V(t) = c_0 e^{t/R}, \tag{1}$$

where c_0 is the spherical initial tumor volume with a diameter of 2mm. The inverse growth rate of the tumor, R, was originally assumed to follow a Gamma distribution (Plevritis et al. 2007; Atkinson et al. 1983):

$$\Gamma_R(\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} r^{\alpha-1} e^{-\beta r},$$

where the parameters α and β are solved for using the method presented in Plevritis et al. (2007). The tumor volume doubling time is proportional to the inverse growth rate R, where doubling time (DT) is:

$$DT = \ln(2) \times R.$$

The inverse growth rate, R, defines the underlying behavior of the tumor progression. It is the condition on which the three main events distributions depend and defined as follows:

1. Distribution of tumor volume at symptomatic detection, D:

$$F_{D|R}(d|r) = 1 - e^{-\gamma r(d-c_0)} \tag{2}$$

2. Distribution of tumor volume at transition from local stage to regional stage of cancer, N:

$$F_{N|R}(n|r) = 1 - e^{-\eta r(n-c_0)} \tag{3}$$

3. Distribution of tumor volume at transition from regional stage to distant stage of cancer, M:

$$F_{M|N,R}(m|n,r) = 1 - e^{-\omega r(m-n)} \tag{4}$$

The parameters γ, η, and ω are estimated using SEER data, where the unconditioned distributions are used instead. To personalize our model to pancreatic cancer, we used the findings of Furukawa et al. (2001). As the study was based on nine patients only, we could not infer about the distribution of the tumor doubling time and hence the distribution of the inverse growth rate. Therefore, we first assumed that the inverse growth rate follows a Uniform distribution, where:

$$R \sim \text{Uniform}(a,b), \text{ where } a = \frac{64}{365 \cdot \ln(2)} \text{ and } b = \frac{255}{365 \cdot \ln(2)}.$$

We then assumed that the inverse growth rate follows a Gamma distribution. We used the limits of the uniform distribution a and b to be the 10th and 90th percentiles, respectively. We also assume the same mean for both distributions. We iteratively solved for the corresponding parameters of the distribution, where:

$$f_R(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \text{ where } \alpha = 5.1590 \text{ and } \beta = 8.1832.$$

We study the effect of the distribution of the inverse growth rate on the tumor progression model output. Using the findings of Furukawa et al. (2001), we perform two analyses where we assume the inverse growth rate follows (i) a Uniform distribution, and (ii) a Gamma distribution. We incorporated the range of the tumor doubling time and maintained the mean to reduce the variability of the two assumed distributions. The progression model was assumed to be dependent on race, gender, and age. Considering gender, two races (Black and White), and four age groups by percentile, we developed sixteen growth models that were parameterized individually under the two different distribution assumptions of R using SEER data. We used the Kullback-Leibler divergence measure to evaluate the performance of the two models compared to the SEER data used, and we found that the model with $R \sim$ Gamma to be a closer fit to the data than the model with $R \sim$ Uniform.

2.2 Simulation Model

We used the SEER data to build distributions of cancer initiation age. The SEER patient records are diagnosis records where the given tumor size is at diagnosis. We utilized the inverse of Equation 1, expressed as:

$$T(V) = R \ln\left(\frac{V}{c_0}\right) \tag{5}$$

where the $T(V)$ is the time it takes for the tumor to grow from volume c_0 to volume V. Using the age at diagnosis, tumor diameter, race, and gender, for every patient we:

1. Sampled an inverse growth rate, R.
2. Calculated the time from tumor volume c_0 until the indicated volume in SEER using Equation 5.
3. Calculated the age at diagnosis minus the time calculated previously, and record the age of cancer initiation.
4. Generated a total of ten cancer initiation ages per patient and record the average.

By race and gender, we fit the distributions of age of cancer initiation under both assumptions of R using Johnson SB. We simulated experiments to study the impact of the inverse growth rate distribution on the resulting tumor progression timeline. We estimated (i) the population's average tumor progression; (ii) a subpopulation of race and gender average tumor progression; and (iii) a subpopulation of race, gender and age average tumor progression. Patients for every experiment were randomly generated, and the tumor was created simultaneously using the two distinct progression models.

2.2.1 Patient Creation

To create a cancer patient in the general pancreatic cancer population, we estimated the ratio of White male to White female to Black male to Black female and found it to be 446:419:65:70. Using the random uniform number generator, we generated a probability p and compared:

1. if $p < 0.446$, we set the race and gender to White male, or
2. else, if $p < 0.865$, we set the race and gender to White female, or
3. else, if $p < 0.930$, we set the race and gender to Black male, or
4. else, we set the race and gender to Black female.

Based on the generated race and gender, we generate a cancer initiation age from the race/gender specific distributions created. For the second set of experiments, we wanted to evaluate the average tumor progression by race and gender. We created patients in a certain race and gender by setting the race and gender rather than random generation, and we generate a cancer initiation age. For the last set of experiments, we test the tumor progression by race, gender, and age. For every patient, we used the race, gender, and age specific

tumor growth parameters that we solved for in Abu-El-Haija (2019) when simulating their tumor growth (Abu-El-Haija 2019).

2.2.2 Tumor Creation

To simulate tumor progression for a patient, we first simulate an inverse growth rate, either following a Uniform or Gamma distribution. We utilize the built-in random generation functions in Python. Given an inverse growth rate, we create the tumor timeline by simulating the tumor volume at the transition from the local stage to the regional stage, the transition from the regional stage to the distant stage, and symptomatic detection. To simulate the transition events, we use the inverse of Equations 2, 3, and 4 expressed as:

$$G_{X|R,Z}(x|r,z) = z - \ln\left(\frac{1-x}{y \cdot r}\right) \tag{6}$$

where R is the inverse growth rate, X is a randomly generated probability from Uniform(0,1), Z is the starting volume c_0 or n, and Y is the distribution parameter γ, η, or ω.

- **Step 1** Create a Patient:
  - For a heterogeneous population: Using the distribution of race/gender in the SEER data, sample a race and gender.
    - Using the race/gender sampled, identify the appropriate distribution of cancer age initiation.
    - Sample an age.
  - For a homogeneous population of race and gender: Specify a race and gender.
    - Using the race/gender sampled, identify the appropriate distribution of cancer age initiation.
    - Sample an age.
  - For a homogeneous population of race, gender, and age: Specify a race, gender, and age group.
- **Step 2** Using the patient's race, gender, and age from Step 1, retrieve the appropriate progression model parameters γ, η, and ω.
- **Step 3** From the distribution of the inverse growth rate, R, generate a sample r.
- **Step 4** Create Tumor:
  - Generate a random probability p_D to generate a tumor volume at symptomatic detection, d, using Equation 6.
  - Generate a random probability p_N to generate a tumor volume at transition from local to regional, n, using Equation 6.
  - Generate a random probability p_M to generate a tumor volume at transition from regional to distant, m, using Equation 6.
- **Step 5** Translate tumor volumes generated to time in years using the Equation 5 and record the time stamps of the events generated.

The model assumes that cancer stages progress from local to regional to distant. We assume that it is not possible to move from local to distant without passing through the regional stage. The model also assumes that the event of symptomatic detection is independent from the tumor stage transitions from local to regional to distant. Figure 1 shows the flow of a simulation run where cancer patients are created, and the tumor grows based on the patient's race, gender, and age.

3 RESULTS

For every experiment, we generated 100,000 patients and estimated the average and the 95% confidence interval around the mean of the time between tumor initiation and transitions in the tumor progression. We estimated the average effect of the underlying distribution of R by creating a heterogeneous population of pancreatic cancer patients and simulated their tumor timelines. Figure 2 shows the times from cancer

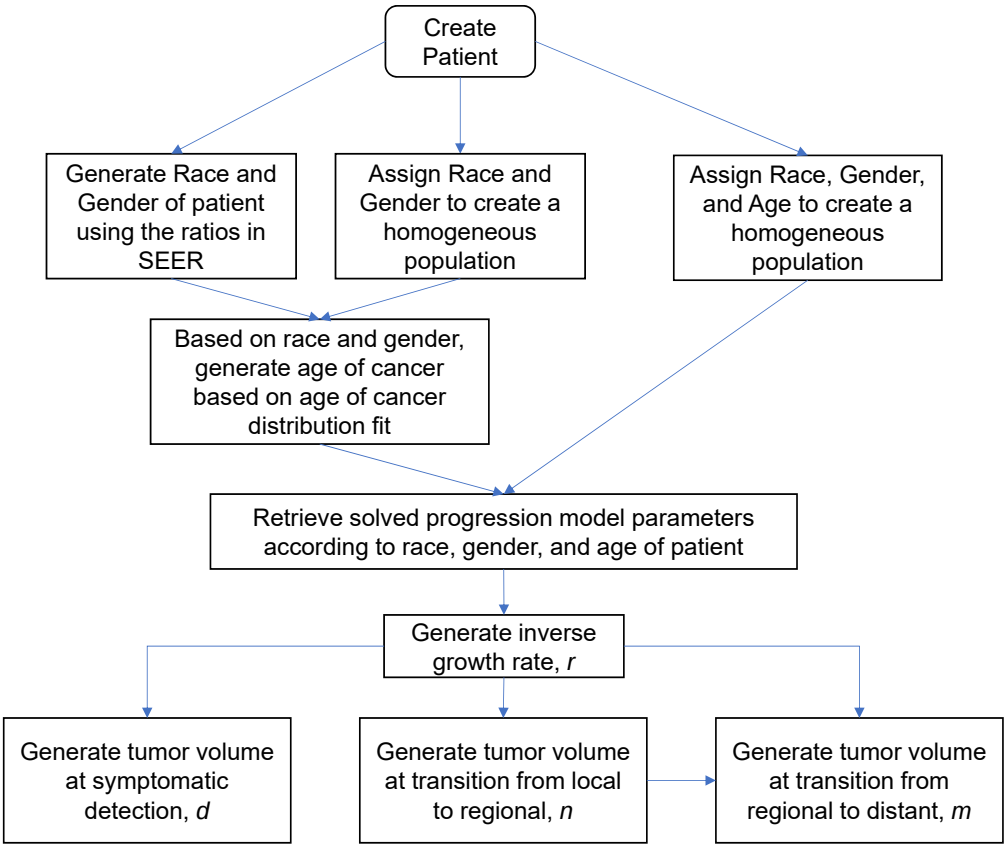

Figure 1: Flow diagram of a simulation run of tumor progression.

initiation until transition to regional, transition to metastasis, and transition to symptomatic detection. The tumor progression was found to be generally slower under the assumption $R \sim$ Gamma when the average time until regional, time until metastasis, and time until symptomatic detection of a heterogeneous population. However, the time span of the regional stage is shorter when $R \sim$ Gamma, and this trend continues as we investigate further. Consistently, the average time until symptomatic detection is longer than the time until metastasis. This indicates that the tumor is often symptomatic at the metastatic stage under both assumptions of R, which is in agreement with the literature that pancreatic cancer is a late symptom disease.

We found that there is a trend in the timelines, where the time until transition to regional is longer, the time in the regional stage is shorter, and the total time until transition to metastasis is longer under $R \sim$ Gamma. This trend remains in both the progression models by race and gender and the progression models by race, gender, and age. Figure 3 shows the time until metastasis of pancreatic cancer by race and gender with $R \sim$ Uniform and $R \sim$ Gamma. Black males have the longest timelines under both distributions of R, and White females have the shortest timelines, with a difference between of one and a half to three months between their timelines. There is also a difference of approximately nine months between the progression timelines of the models under the different assumptions of R when comparing the time until metastasis within each race and gender group. $R \sim$ Uniform causes the tumor progression to be more conservative with the shorter timelines, which consecutively would call for a more aggressive screening policy to detect the tumor before metastasis.

We further investigated the effect of the distribution of R by generating the tumor timelines by race, gender, and age group. Figure 4 shows the resulting time until metastasis of the sixteen different subgroups. The tumor timelines of Black males still remain to be the longest under $R \sim$ Gamma, but they are not under

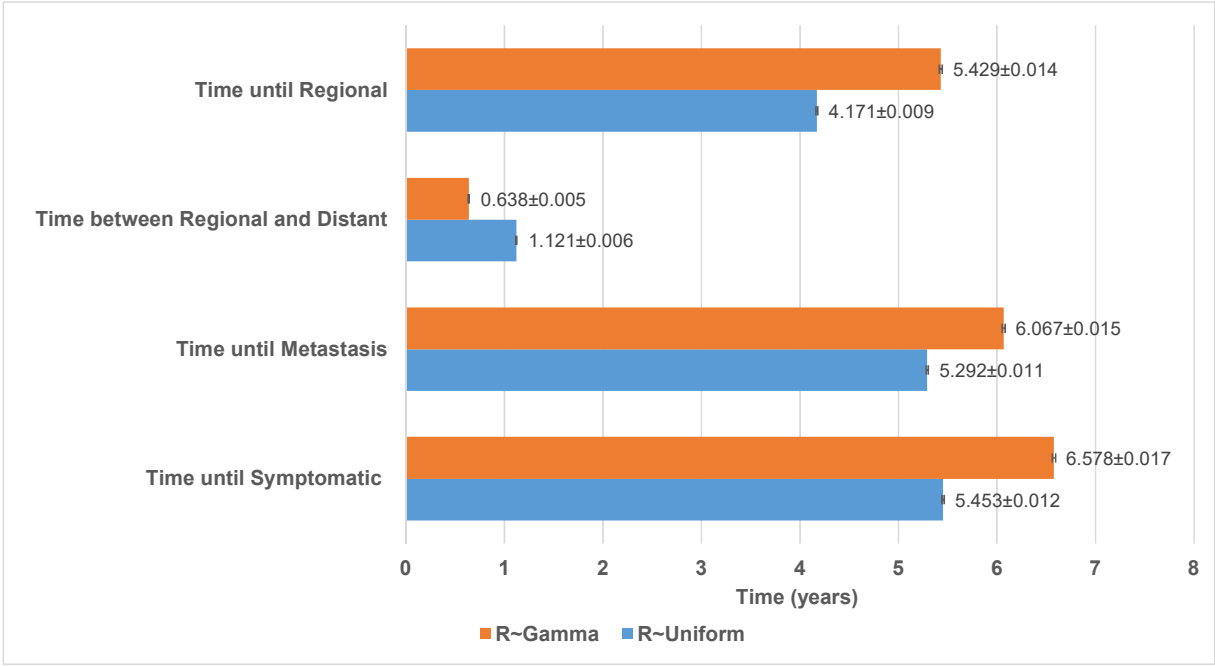

Figure 2: Heterogeneous population tumor average timeline.

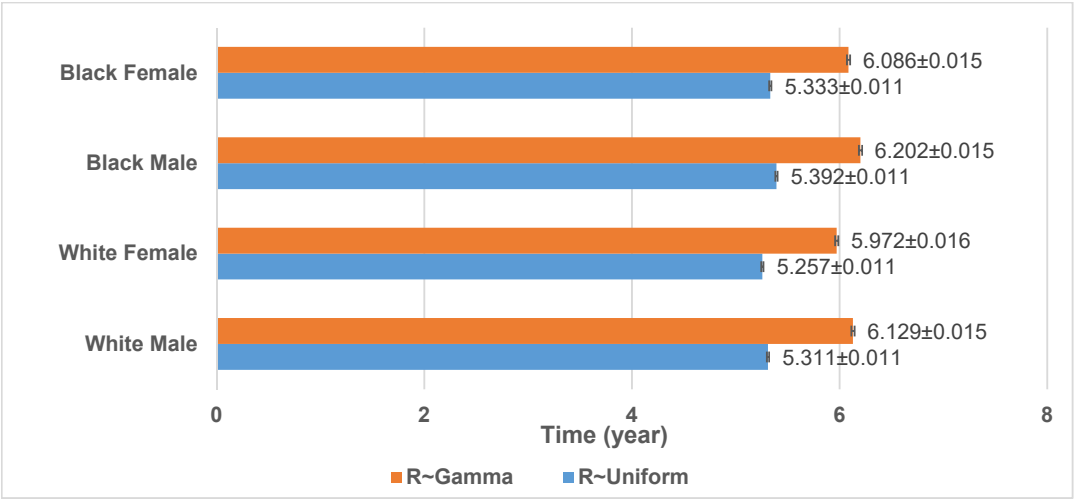

Figure 3: Time from cancer initiation until metastasis by race and gender.

$R \sim$ Uniform. The time until metastasis decreased (making the tumor faster in growth) as the patients got older under $R \sim$ Gamma for every race and gender group. However, the time until metastasis increased (making the tumor slower in growth) as the patients got older under $R \sim$ Uniform for White male, White female, and Black female. Essentially, as the patients got older, the time until metastasis for each of the different models became closer to each other. Other than the time until metastasis consistently being shorter under $R \sim$ Uniform, there did not seem to be other differences in the progression.

Time until other generated transitions in the tumor progression among the race, gender, and age groups were also considered. We analyzed the transitions from local to regional to distant in both models. Figure 5 details the times until transition from local stage to regional stage and transition from regional stage to

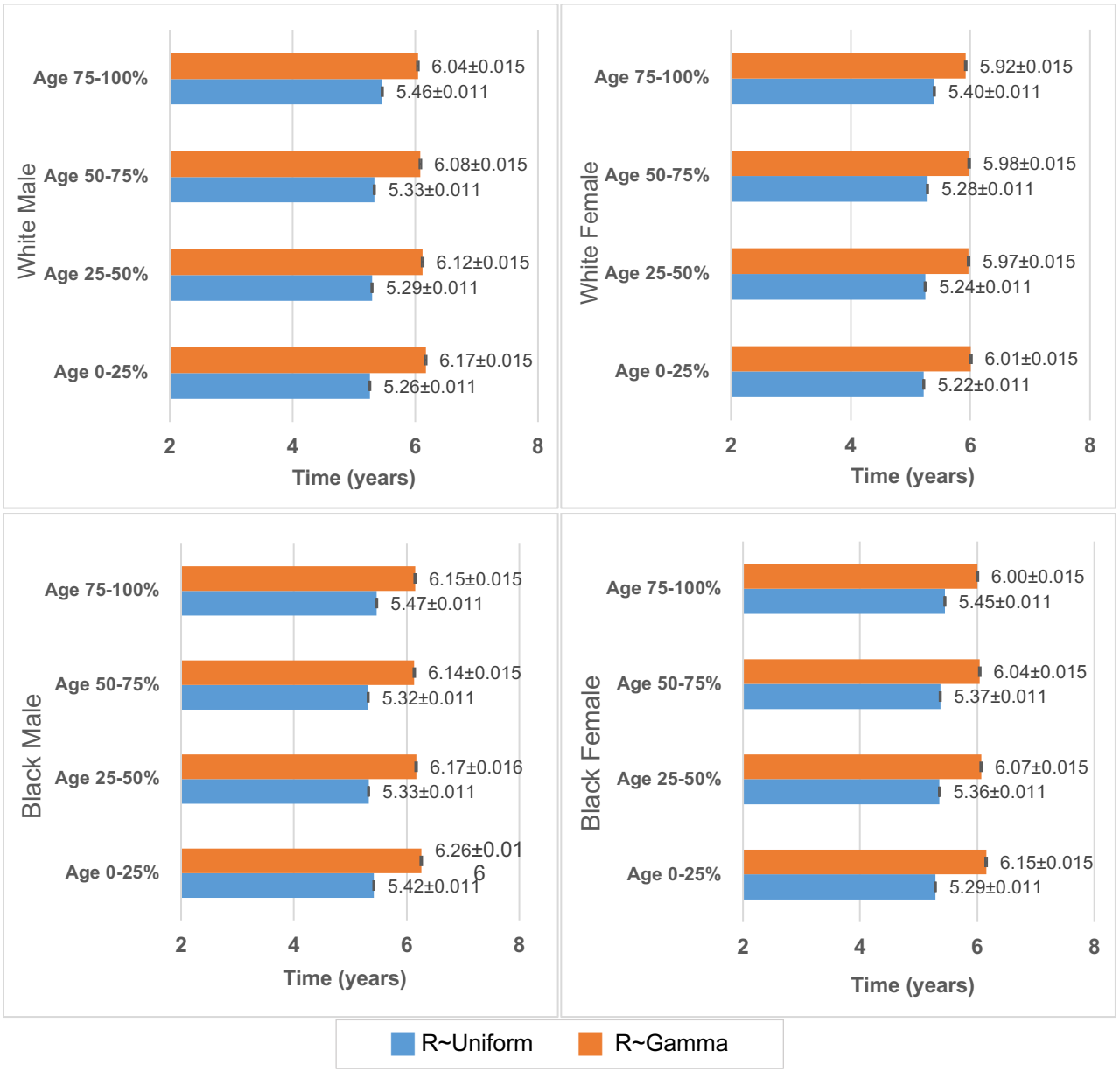

Figure 4: Mean time from cancer initiation until metastasis by race, gender, age group, and under different inverse growth rate distribution assumptions, labeled in years with 95% confidence interval.

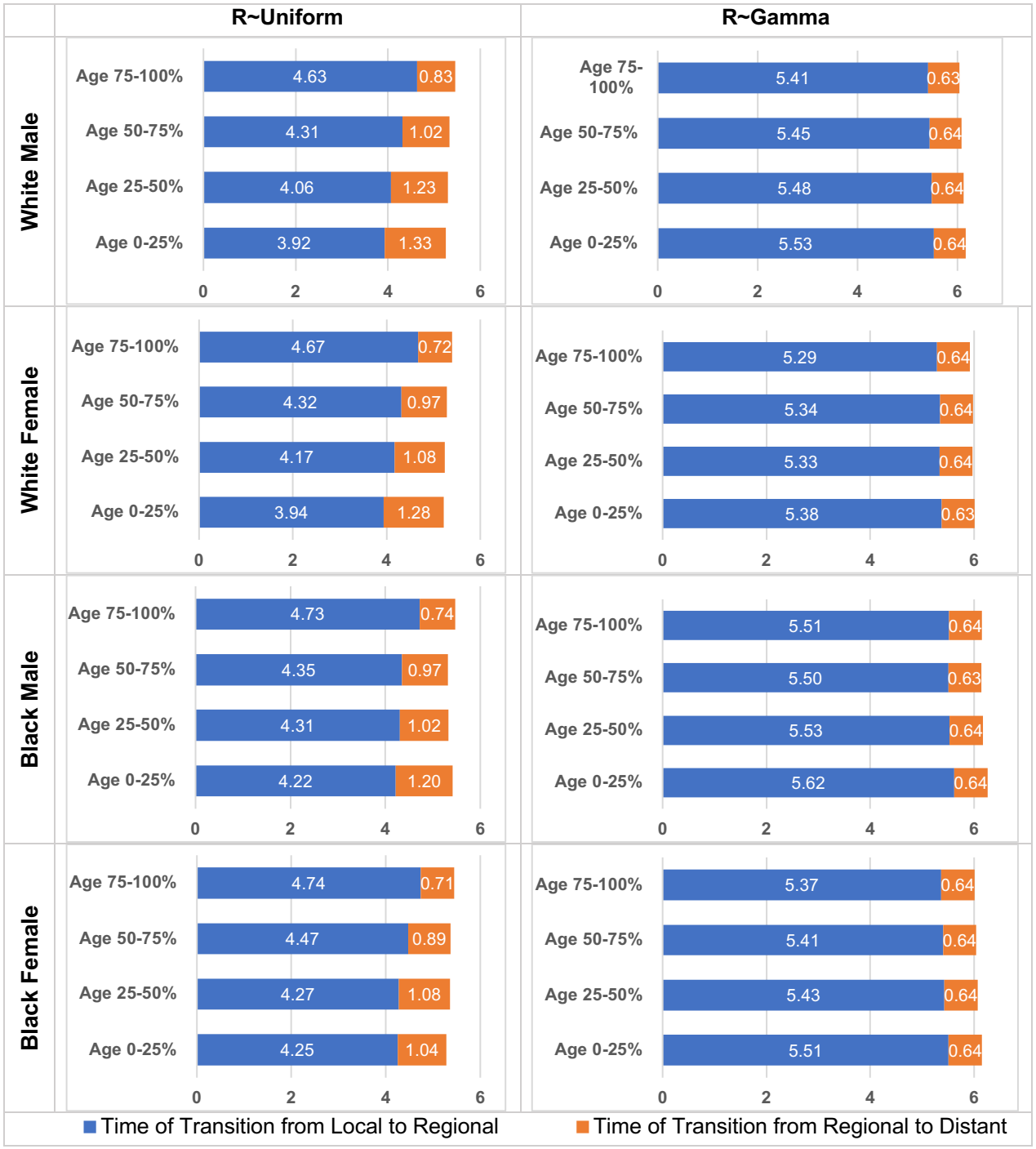

Figure 5: Mean time in years until tumor transitions from local to regional and from regional to distant by race, gender, age group, and under different inverse growth rate distribution assumptions.

distant stage. The time spent in the regional stage across all models with $R \sim$ Gamma is about 0.64 years (7.68 months). Therefore, only the time spent in the local stage reduces as patients get older leading to the shorter time until metastasis in the $R \sim$ Gamma based models. On the other hand, the length of the regional stage reduces from over a year to less than a year as patients get older in all models with $R \sim$ Uniform, as patients get older the length of the local stage increases by up to nine months.

4 DISCUSSION AND CONCLUSION

We developed a pancreatic cancer progression discrete-event simulation model to estimate the effect of the distribution of the inverse growth rate on the output of the progression model. The progression model consisted of probability distributions of tumor volume at symptomatic detection, transition from local to regional, and transition from regional to distant. The probability distributions are conditioned on the inverse growth rate, R. In some literature, the inverse growth rate was assumed to follow a Gamma distribution. However, studies that estimated the tumor doubling time of pancreatic cancer only used a small sample of patients. Therefore, when using their result, we could not conclude a distribution that R follows.

We conducted simulation experiments of two growth models where they only differed by the distribution that R followed: Uniform or Gamma. Generally, we found that the tumor progression models with $R \sim$ Gamma yielded longer tumor timelines than the progression models with $R \sim$ Uniform. Without knowledge of the true distribution of R, we cannot conclude on the tumor progression behavior with older versus younger patients. The progression models with $R \sim$ Gamma indicated that the tumor grows faster in older patients, while the progression models with $R \sim$ Uniform indicated that the tumor grows faster in younger patients. This type of information can be extracted from literature on other cancers, yet it remains unclear in the pancreatic cancer realm.

There is value in accurately modeling the distribution of the inverse growth rate since it is the underlying distribution of the tumor growth model. There are measures of accuracy, such as the Kullback-Leibler divergence, that evaluate the performance of the model compared to existing data. In our previous work, we found that the model with $R \sim$ Gamma as the underlying assumption of the model to be a closer fit to the data that was used to fit the parameters of the model. However, there can be another distribution of R, or a Gamma distribution with more appropriate parameters, such that the model would be an even closer fit to the data. We can conclude that the inverse growth rate drives the tumor progression model, and without a higher accuracy in estimating R, our models would not well represent pancreatic cancer patients.

We have introduced the problem of scarcity of data in pancreatic cancer and how we used simulation techniques to show the effects it has on tumor progression modeling. Given the limitation of the doubling times of pancreatic cancer in the literature, simulation can be used to populate data points to increase the sample size. In future work, we can extend this model to evaluate the value of information in estimating tumor doubling time or growth rate. We hope to drive the conversation to recruiting pancreatic cancer patients and estimating the tumor growth rate such that we can build higher accuracy models.

REFERENCES

Abu-El-Haija, L. 2019. *PRECiSE: Pancreatic Cancer Prioritization and Screening Evaluation Tool*. PhD dissertation, Department of Industrial and Systems Engineering, North Carolina State University, Raleigh, North Carolina. http://www.lib.ncsu.edu/resolver/1840.20/36866, accessed 1st May 2019.

Atkinson, E. N., R. Bartoszynski, B. W. Brown, and J. R. Thompson. 1983. "On Estimating the Growth Function of Tumors". *Mathematical Biosciences* 67(2):145–166.

Furukawa, H., R. Iwata, and N. Moriyama. 2001. "Growth Rate of Pancreatic Adenocarcinoma: Initial Clinical Experience". *Pancreas* 22(4):366–9.

Mehrara, E., E. Forssell-Aronsson, H. Ahlman, and P. Bernhardt. 2007. "Specific Growth Rate Versus Doubling Time for Quantitative Characterization of Tumor Growth Rate". *Cancer Research* 67(8):3970–3975.

National Cancer Institute 2015. "Surveillance, Epidemiology, and End Results (SEER) Program, Research Data (2004-2014)".

Plevritis, S. K., P. Salzman, B. M. Sigal, and P. W. Glynn. 2007. "A Natural History Model of Stage Progression Applied to Breast Cancer". *Statistics in Medicine* 26(3):581–595.

Rahib, L., B. D. Smith, R. Aizenberg, A. B. Rosenzweig, J. M. Fleshman, and L. M. Matrisian. 2014. "Projecting Cancer Incidence and Deaths to 2030: The Unexpected Burden of Thyroid, Liver, and Pancreas Cancers in the United States". *Cancer Research* 74(11):2913–2921.

Sun, S., F. Klebaner, and T. Tian. 2014. "A New Approach for Estimating the Progression of Pancreatic Cancer". In *Proceedings of the 2014 8th International Conference on Systems Biology (ISB)*, edited by L. Chen, X.-S. Zhang, L.-Y. Wu, and Y. Wang, 118–123. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Weedon-Fekjær, H., B. H. Lindqvist, L. J. Vatten, O. O. Aalen, and S. Tretli. 2008. "Breast Cancer Tumor Growth Estimated Through Mammography Screening Data". *Breast Cancer Research* 10(3):R41.

Xu, Z., S. P. Pothula, J. S. Wilson, and M. V. Apte. 2014. "Pancreatic Cancer and Its Stroma: A Conspiracy Theory". *World Journal of Gastroenterology* 20(32):11216–11229.

Yachida, S., and C. A. Iacobuzio-Donahue. 2013. "Evolution and Dynamics of Pancreatic Cancer Progression". *Oncogene* 32(45):5253–5260.

Yachida, S., S. Jones, I. Bozic, T. Antal, R. Leary, B. Fu, M. Kamiyama, R. H. Hruban, J. R. Eshleman, M. A. Nowak, V. E. Velculescu, K. W. Kinzler, B. Vogelstein, and C. A. Iacobuzio-Donahue. 2010. "Distant Metastasis Occurs Late During the Genetic Evolution of Pancreatic Cancer". *Nature* 467(7319):1114–1117.

AUTHOR BIOGRAPHIES

LENA ABU-EL-HAIJA is a recent PhD graduate from the department of Industrial and Systems Engineering at North Carolina State University. She is an Assistant Professor in Industrial Engineering at the German Jordanian University. Her research interests include cancer screening models using simulation techniques. Her email address is lena@haija.org.

JULIE IVY is a Professor in Industrial and Systems Engineering at North Carolina State University and the previous advisor of L. Abu-El-Haija. Her research interests include the mathematical modeling of stochastic dynamic systems applied to health care, manufacturing, and service environments. Her email address is jsivy@ncsu.edu.

OSMAN OZALTIN is an Assistant Professor in Industrial and Systems Engineering at North Carolina State University. His research interests include theoretical, computational, and applied aspects of mathematical programming, focusing on problems arising in public health policy making, personalized medical decision-making, and healthcare delivery. His email address is oyozalti@ncsu.edu.

WALTER PARK is an Assistant Professor of medicine alt Stanford University Medical Center, specialized in Gastroenterology and Hepatology. His research interests include diagnosis and management of pancreatic cysts, acute and chronic pancreatitis, and pancreatic cancer. His email address is wgpark@stanford.edu.