Simulation

Tissue scale agent-based simulation of premalignant progressions in Barrett's esophagus

Ozgur Ozmen[1] (iD)**, James Nutaro**[1] (iD)**, Sassan Ostvar**[2]**, Chin Hur**[2]
and Chung Yin Kong[3,4]

Abstract
Barrett's esophagus (BE) is a benign condition of the distal esophagus that initiates a multistage pathway to esophageal adenocarcinoma (EAC). Short of frequent intrusive (and costly) surveillance, effective screening for neoplasia in BE populations is yet to be established since progressors are rare and virtually undetectable without routine biopsies, which often sample only a small portion of the BE tissue. As a result, reliable estimation of the true prevalence of dysplasia in a BE population and evidence-based optimization of screening for at-risk individuals is challenging. Data-driven microsimulations, i.e., model-generated instances of disease history in a predefined virtual population, have found utility in the EAC screening literature as low-overhead alternatives to real-world hypothesis testing of optimal interventions for dysplasia. Despite the successes, computational limitations, paucity of knowledge and data on Barrett's dysplasia, and the complexities of disease progression as a multiscale multiphysics process have hindered the treatment of disease progression in BE as a spatial process. Agent-based modeling of nucleation and proliferation processes in dysplasia warrants exploration in this context as an approximation that operates at a trade-off between computational tractability and precise representation of the composition and physics of the substrate (tissue). In this study, we describe spatially resolved simulations of premalignant progression toward EAC in a coarse-grained model of Barrett's tissue that resolves the metaplastic tissue at a length scale of 0.42 mm (~3300 crypts/mm^2). The model is calibrated to reproduce historical high-grade dysplasia prevalence when model-generated patients are screened using the Seattle protocol.

1. Introduction

The puzzling growth in incidences of esophageal adenocarcinoma (EAC) in western populations is a significant public health challenge that can be mitigated in part with better neoplasia screening strategies for at-risk individuals.[1,2] However, it is impractical to perform statistically significant comparisons of screening strategies, and this has seriously hindered improvements in the rate of early detection. To overcome this challenge, a simulation model that captures the process by which EAC emerges can be used to create large numbers of synthetic tissues (patients), and the effectiveness of competing detection strategies can be compared in reference to this virtual population. The primary aim of such a model is to accurately reproduce the physical distribution over time of cell types that are detected via biopsies.

A prominent pathway to EAC is believed to be the emergence of neoplasia in Barrett's esophagus (BE), a precursory metaplastic alteration of the esophageal lining where epithelial squamous cells are replaced with mucus-secreting columnar cells,[3] possibly due to chronic reflux damage.[4] Premalignant growth in the metaplastic tissue presents in the form of low-grade dysplasia or high-grade dysplasia (HGD). The latter is considered a significant

[1]Computational Sciences and Engineering Division, Oak Ridge National Laboratory, USA
[2]Columbia University Irving Medical Center, USA
[3]Division of General Internal Medicine, Icahn School of Medicine at Mount Sinai, USA
[4]Institute for Technology Assessment, Massachusetts General Hospital, USA

Corresponding author:
Ozgur Ozmen, Computational Sciences and Engineering Division, Oak Ridge National Laboratory, 1 Bethel Valley Road, Oak Ridge, TN 37831, USA.
Email: ozmeno@ornl.gov

marker for progression toward EAC, hence a primary target for screening.

Preventive screening for EAC has proved challenging, in part, because, within the interval that the disease is believed to be still curable, increasingly advanced premalignant stages require increasingly invasive procedures to detect. BE—estimated to affect 5.6% of the adult US population alone[5]—is symptomatically silent and requires endoscopies for detection. Individuals diagnosed with BE are regularly surveilled via repeated biopsies of the metaplastic tissue to enable early detection of neoplasia, the yield of which is subject to the efficiency of the biopsy protocols. Therefore, our posterior knowledge on the incidence of advanced disease stages, particularly HGD, is subject to uncertainties stemming from who is selected to receive an endoscopy and how efficiently can we detect signs of malignant mutations in metaplastic tissue. These uncertainties occur on distinct levels of abstraction within the problem, highlighting its multiscale nature.

A data-driven state-transition model, EAC Model (EACMo), has been used in epidemiology to amend missing or scarce population-level observational data.[1,5,6] EACMo has been used to address the HGD screening problem. However, the model is a state-transition model, which does not provide tissue-level details of premalignant growth. A potential remedy is to complement state-transition population models with tissue-level simulations of premalignant growth. Grid-bound individual-based models are particularly well-suited for this due to their balancing of simplicity with diverse dynamical output. The primary focus of this study is to show how such a model can be constructed and deployed at scale to address the screening problem outlined above.

Specifically, we show how a spatially resolved model for the BE to EAC pathway can be constructed using the discrete event system specification (DEVS,[7] see Wainer[8] for other grid models). This formulation of the model allows for fast simulations to produce sizable populations of virtual tissues. We use this synthetic population to evaluate the Seattle biopsy protocol,[9] the current standard (recommended) protocol for detection of neoplasia in Barrett's patients, and calibrate the model parameters toward historical HGD prevalence and missed malignancy rates.

2. Methods

2.1. Three-dimensional agent-based model

Agent-based models typically describe the rule-based evolution of objects (i.e., agents) through a set of predefined states over time. In our implementation, objects are groups of adjacent cells that constitute a segment of crypt-structured tissue. The states of the model correspond to different health states of individual tissue segments, and these segments are assembled to construct the esophageal tissue. The esophagus is

modeled as an annulus with inner diameter r_0 and thickness ℓ, and is discretized into a regular grid of cubic tissue segments with side length δ and crypt density η.

The model follows each tissue segment through the natural history as:

$$\text{normal} \longrightarrow \text{BE} \xrightarrow{\mu_1} \text{dysplasia} \xrightarrow{\mu_2} \text{cancer} \qquad (1)$$

where μ_1 and μ_2 are the mutation rates associated with each irreversible transition. The instant of mutation for a tissue segment is determined by sampling from an exponential distribution with the appropriate mean. The mutation rates are calculated as follows:

$$\mu_i = \frac{1}{k_i \eta \delta^2} \qquad (2)$$

where k_i is a parameter to be calibrated.

In addition to these random mutations, a tissue segment can change state in response to invasions from adjacent segments of cancerous or dysplastic tissue. The process of cell duplication and spreading is characterized by a proliferation rate, μ_*. This value is calculated as:

$$\mu_* = k_* \delta^2 \qquad (3)$$

where k_* is to be calibrated. Cancer can spread in three dimensions, but the spread of dysplasia is restricted to two dimensions on the epithelium. Each proliferation event affects a single neighbor selected at random.

A historical study shows that metaplastic tissue segments occur at some point between the ages 20 (minimum age) and 50 in most patients.[10] For our study, we begin the simulation of a tissue with onset of BE at age τ_{BE} that is drawn for each patient from a real dataset. At time τ_{BE}, a transition from normal to BE occurs. The initial length of the BE tissue, L_{BE}, is drawn from a bimodal Gaussian distribution to reflect a distinction between short- and long-segment BE. Long-segment BE is assigned to 25% of the population and short segment BE to the remainder. Similar to BE onset age, we incorporated death from natural causes in the model. For each tissue (patient), death age is drawn from real mortality rates for the 1960 cohort. A summary of the model parameters is presented in Table 1.

Individual tissue segments are described by event-driven state transition systems called atomic models in the parlance of the DEVS modeling formalism. Atomic models interact by exchanging events. A complete description of the DEVS modeling technique and its simulation procedures can be found in Zeigler et al.[11] and Nutaro.[12] Our model is built using ADEVS (a discrete event system simulator), which is an implementation in C++ of the DEVS modeling constructs and simulation algorithm.[12] In this implementation, the tissue segments are maintained in a future event list. At each iteration of the simulator, ADEVS selects the imminent tissue segments

Table 1. Calibration targets, calibration algorithm values, and the associated model rates.

Parameter	Value	Reference
$D^{0.05}$	0.008	O'Conner et al.[25]
$D^{0.50}$	0.033	Sharma et al.[24]
$D^{0.95}$	0.075	Weston et al.[23]
$M^{0.05}$	0.57	Peters et al.[26]
$M^{0.50}$	0.35	Falk et al.[27]
$M^{0.95}$	0.12	Cameron and Carpenter[28]
δ	0.42 mm	Assumed
η	3300 cells/mm^2	Curtius et al.[15]
r_0	24 mm (quarter diameter)	Assumed
ℓ	4 mm (width of three pennies)	Assumed
k_1	1.2405×10^{-8} year/cell	Calibrated
k_2	4.0060×10^{-8} year/cell	Calibrated
$k*$	0.2103 year	Calibrated
L_{BE} long	Mean 6.4 cm, SD 3.1 cm	Assumed
L_{BE} short	Mean 1.4 cm, SD 0.7 cm	Assumed
Age of biopsy	60	Assumed
Samples (biopsy)	100	Algorithm
Simulation replications	500	Algorithm
Generations	500	Algorithm

from this list, executes its output function by which events are sent to neighboring segments, and then executes the segment's state transition function to update state variables in the imminent segment (Source code for the simulation model is available in the article's supplementary materials).

If the event at an imminent tissue segment is a mutation, then the state of the segment is updated appropriately, the new time of the next event for the segment determined, and then it is reinserted into the future event list. Otherwise, the event is a proliferation. In this case, a destination segment is selected and the state of that segment updated accordingly. Then new times of next event are determined for the pair of segments and these are reinserted into the future event list.

This approach has two benefits over a typical agent-based model that steps through time in fixed increments, examining each agent at each step. First, time is continuous in the discrete event model as next event times are selected from the real numbers. This avoids truncation errors that frequently occur in time-stepped simulations of agent-based models[13,14] when continuous distributions are used (such as exponentially distributed mutation occurrences in our model). The second benefit is that computational effort at each iteration of the simulation algorithm is restricted to the imminent tissue segments and their neighbors. This allows for the multi-year growth of malignant tissues to be simulated very rapidly.

2.2. Interpretation of μ_i and μ_*

Note that the mutation rates μ_i do not correspond to rates of somatic mutations of individual epithelial stem cells,

but a group of cells constituting a BE crypt. We created this coarse-grained model because it is computationally impractical to simulate individual cells of the esophagus. For instance, μ_1 characterizes the transition from Barrett's esophagus to the onset of dysplasia, i.e., the emergence of proliferative dysplastic foci. This process can be further subdivided into (1) (potentially multiple) rate-limiting initiation steps on the scale of individual stem cells, (2) the turnover of a BE crypt into a dysplastic one, and (3) the turnover of a tissue segment due to the transformation of its constituent crypts. Similarly, μ_* agglomerates proliferation mechanisms (e.g., crypt fission) below the length scale of an individual tissue segment. In this sense, these rates represent the aggregation of smaller-scale mutation and proliferation mechanisms on a length scale equal to δ.

2.3. Screening via simulated biopsies

We use the tissues generated by the simulation runs to mimic the process of screening BE tissues for HGD. The current recommended procedure for HGD screening is the Seattle biopsy protocol[9] where 5 mm $\times$ 3 mm jumbo biopsies are extracted from the surface of the metaplastic tissue every 1–2 cm, making up four-quadrant biopsies (also called a band) in total. The samples are then surveilled by a pathologist for signs of neoplasia. To account for the potential inefficiency in this step, we adopt the biopsy sensitivity measure from a previous study.[15] A detection threshold, ϕ_{det} is defined in terms of the fractional area in a single jumbo biopsy covered by neoplastic tissue (n_f) as follows:

$$\phi_{det} = (1 - n_f) \times 100 \tag{4}$$

Note that ϕ_{det} encompasses all contributing sources of uncertainty during the biopsy process including the human factors, the number of samples that are actually tested, and so forth. In practice, adherence to the standard sampling protocol itself is imperfect.[16] To address this, we repeated the simulated biopsies, slightly varying different aspects of the process as follows.

Sampling is started at a randomly chosen location between 1 and 2 cm above the base of the metaplastic tissue. To mimic imperfect placement of jumbo samples, we uniformly perturb the gap between four-quadrant jumbo biopsies on the same band by $U(-6.72, 6.72)$ mm that corresponds to ± 16 tissue segments. On the length of esophagus, we move randomly in 1–2 cm increments and sample another four-quadrant jumbo biopsies. Similarly, the location of the jumbo biopsy on the y-axis is also uniformly perturbed by $U(-3.36, 3.36)$ mm that corresponds to ± 8 tissue segments. To account for first-order uncertainty, the process is repeated 100 times, each representing an independent surveillance of the Barrett's tissue, from which mean detection and false negative rates are calculated.

The outputs of this procedure are the mean prevalence of HGD as a function of the sensitivity of the biopsy protocol within the test population and the mean malignancy rate that is missed by the biopsy protocol. Cross-validating these curves with historically reported values for prevalence and missed malignancy forms the basis for an inverse problem to determine the model parameters k_1, k_2, and k_* as described in the following section.

2.4. Model calibration

Model calibration is a process to find the optimal parameters that realize the desired outputs from a simulation model[17–19] (for calibration with adaptive optimization techniques). Our goal here is to find the model parameters at which the simulation model will output historical HGD prevalence and missed malignancy rates. We use the particle swarm optimization (PSO) algorithm[20,21] to calibrate our model. PSO is a population-based meta-heuristic that was developed based on analogies drawn from collective behavior of animals and sociopsychology.[22] It conceptualizes a search in the space of possible solutions as the motion of a population of particles that each instantiate a unique solution (i.e., position). Each particle explores the space over time guided by local information, as well as a global drift toward where the swarm has found the most promising solution up to that point in time.

The process is formalized using a motion operator that updates a given particle's position at each iteration (eqv. generation). Explicitly, the future position, $x_i(t)$, of the ith particle is a function of its current position, $x_i(t-1)$, its current velocity, $v_i(t-1)$, its best preceding position in terms of fitness, p_i, and the best preceding position, p_g, globally within the swarm up to that generation. The future positions and velocities are obtained as:

$$v_{i,t} = v_{i,t-1} + \varphi_1(p_i - x_{i,t-1}) + \varphi_2(p_g - x_{i,t-1}) \qquad (5)$$

$$x_i = x_{i,t-1} + v_{i,t-1} \qquad (6)$$

where φ_1 and φ_2 are scalar drift terms. In our case, the solution space is a subset of $\mathbb{R}^3$ containing plausible values for k_1, k_2, and k_*.

Fitness, which is distance from an optimal parameter choice, is the difference between the simulation outputs and historical data. For the first measure of fitness, we adopt an upper bound,[23] middle point,[24] and lower bound[25] for HGD prevalence from different studies as HGD prevalence; see Equations (7)–(9). These three equations are the mean Euclidean distances between the calculated prevalence values P^s and targets D^s for different sensitivity values s. The second measure of fitness is an upper bound,[26] a middle point,[27] and a lower bound[28] for missed malignancy; see Equations (10)–(12). These equations are the mean Euclidean distances between the calculated missed malignancy values Q^s and targets M^s for different sensitivity values s. Both measures of fitness have penalty values when the rates exceed ground truth rates at the lower and upper bounds as:

$$f_1^{0.05}(x) = \begin{cases} (|P^{0.05} - D^{0.05}|)^2, & \text{if } D^{0.05} \leqslant P^{0.05} \\ 100000, & \text{otherwise} \end{cases} \qquad (7)$$

$$f_1^{0.50}(x) = (|P^{0.50} - D^{0.50}|)^2 \qquad (8)$$

$$f_1^{0.95}(x) = \begin{cases} (|P^{0.95} - D^{0.95}|)^2, & \text{if } P^{0.95} \leqslant D^{0.95} \\ 100000, & \text{otherwise} \end{cases} \qquad (9)$$

$$f_2^{0.05}(x) = \begin{cases} (|Q^{0.05} - D^{0.05}|)^2, & \text{if } Q^{0.05} \leqslant D^{0.05} \\ 100000, & \text{otherwise} \end{cases} \qquad (10)$$

$$f_2^{0.50}(x) = (|Q^{0.50} - D^{0.50}|)^2 \qquad (11)$$

$$f_2^{0.95}(x) = \begin{cases} (|Q^{0.95} - D^{0.95}|)^2, & \text{if } D^{0.95} \leqslant Q^{0.95} \\ 100000, & \text{otherwise} \end{cases} \qquad (12)$$

The fitness function we calibrate toward is the mean of all aforementioned cost values, specifically

$$\min_x \frac{1}{6} \left(\sum_{n=1}^{2} \sum_{\substack{i=0.05 \\ \text{step} 0.45}}^{0.95} f_n^m(x) \right) \qquad (13)$$

Note that we fit only toward sensitivities of $s = 0.05$, 0.50, and 0.95, which were found sufficient to generate the desired curves. The algorithm does not guarantee a global optimum but improves over generations with the goal of reaching an acceptable solution. Starting with a population of solutions, the fitness value of each particle is calculated. Then, the particle with the best fitness value is recorded
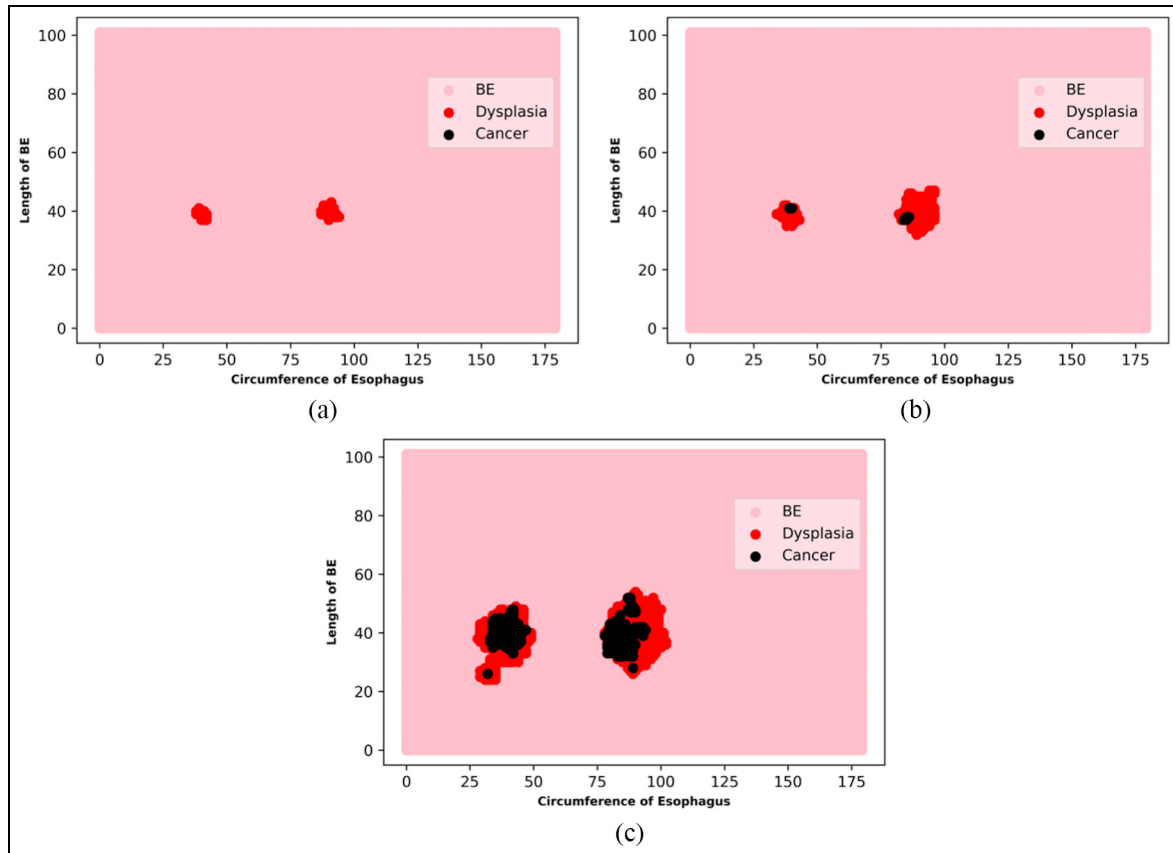
Figure 1. Sample model outputs. The progression from onset of (a) high-grade dysplasia to (b) onset and (c) growth of epithelial carcinoma. The surface of Barrett's tissue has been flattened into a rectangle.

and all particles move to their next position. The search is terminated once a certain number of generations or specified fitness value has been reached. Please see supporting materials for details of the optimization run and its performance. The fitted values and the parameters of the model and the calibration algorithm are presented in Table 1.

3. Results

Figure 1 illustrates the simulated evolution of the epithelial surface over time for one test case. These simulated tissues are leveraged to sample biopsies. Figure 2 illustrates the biopsy process via the spatially perturbed Seattle protocol that was used in the calibration and the analysis (one sample).

Although the mutation rates are uniform in parameter space, the invasion mechanism generates multifocal neoplastic growth with different sizes and shapes. Another strength of the model compared to other tumor growth models[29] is to mimic three-dimensional (3D) tissue structure and allow neoplastic growth deeper into the tissue in different shapes (not only circular). Figure 3 illustrates the depth of cancer invasion (tissue segments) that are not on the surface of esophagus. The invasion mechanism in the
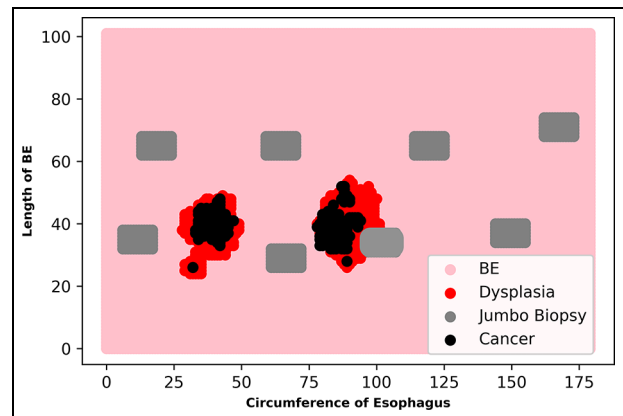
Figure 2. Biopsy sampling based on Seattle protocol on epithelial surface. The sampling shows missed jumbo biopsies and a biopsy sample that overlapped with HGD segments.

model allows more realistic representation of neoplastic growth such as invasion in the tissue and re-surfacing at non-adjacent segments.

The probability of HGD detection as a function of biopsy sensitivity is reproduced in Figure 4(a) at age 60 for 500 replications. The shape and boundaries of the
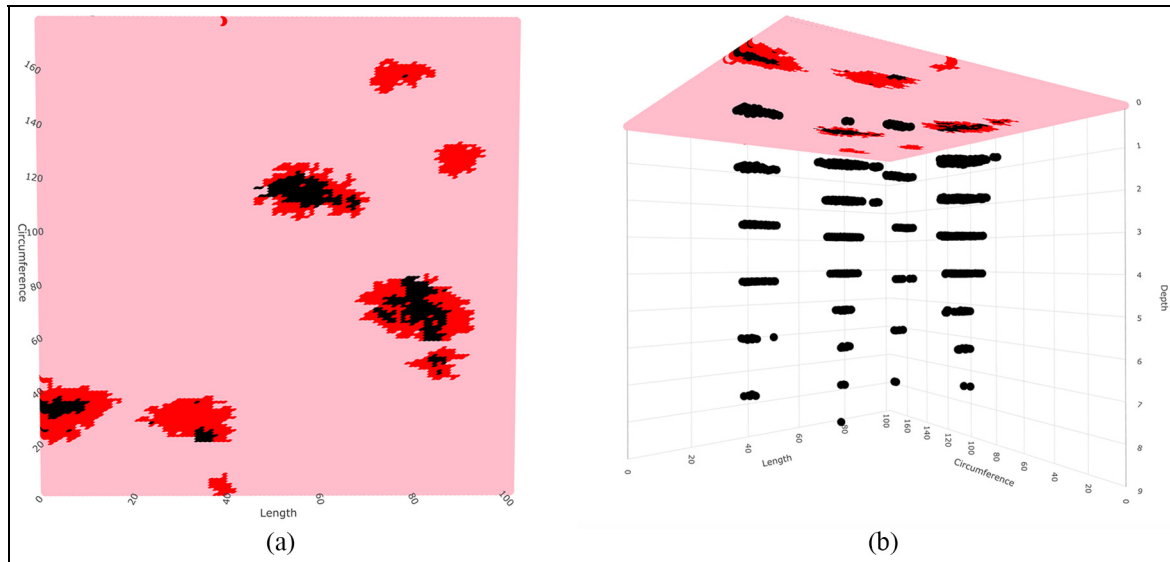
Figure 3. Sample model outputs representing cells deeper in the tissue. (a) Surface of a progressed case and (b) layers of tissues showing that neoplastic growth progressed deeper into the tissue.
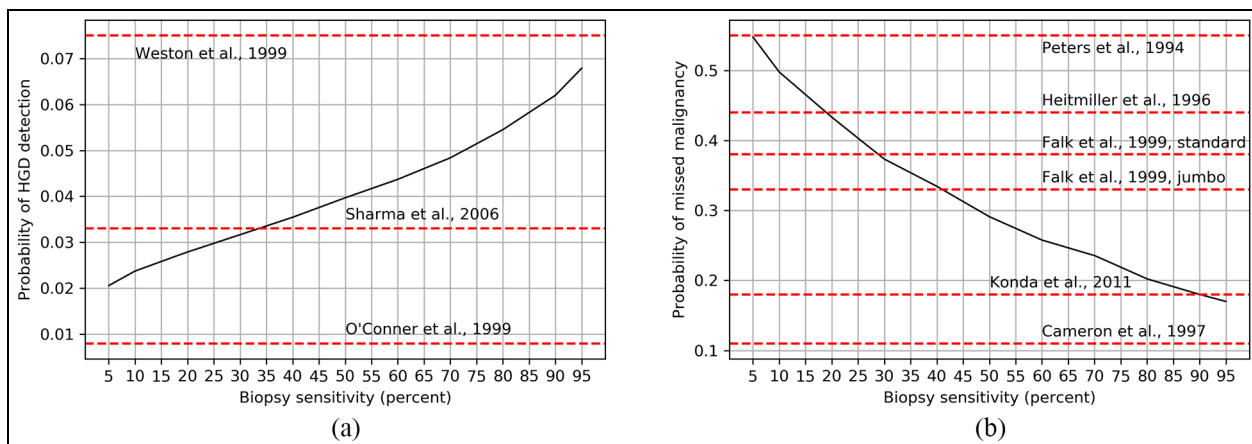

Figure 4. Modeled probabilities. (a) High-grade dysplasia (HGD) detection as a function of the biopsy sensitivity using the Seattle biopsy protocol and (b) modeled probability of missed malignancy when Seattle protocol is used.

curve are consistent with the results from the analysis in the study by Curtius et al.[15] These rates were estimated to evaluate the sensitivity of the biopsy protocol. Figure 4(b) presents the probability of missed malignancy. Missed malignancy is calculated by counting the number of cases that included cancer and remained undetected after the biopsy. It is also presented against the biopsy sensitivity. The predicted ranges are aligned with studies of HGD patients who remained undetected after biopsies.[26–31] The predicted probability of missed malignancy falls well within the range from literature for all biopsy sensitivity values and exhibits a better fit than the analysis in the previous study.[15]

To further validate the calibrated model, it is run for 100,000 replications to generate 100,000 synthetic tissues

until age 100, using the fitted parameters (see Table 1). Figure 5 represents onset age and death age distributions that are drawn from real datasets for the 1960 cohort. Onset age distributes similar to that by Rubenstein et al.[10] in which majority of the population exhibits BE formation by age 50. Death distribution exhibits infant mortalities and positive skewness as expected. Figure 6 represents time intervals from BE to dysplasia and from dysplasia to cancer for first occurrences.

Both dysplasia and cancer exhibit Gaussian distributions. This is due to the invasion mechanism of dysplasia and cancer and summations of multiple exponential time intervals of dysplasia cells to become cancer. Figure 7 presents distributions of ages when first dysplasia (left) and cancer (right) cells are formed. Distribution of dysplasia
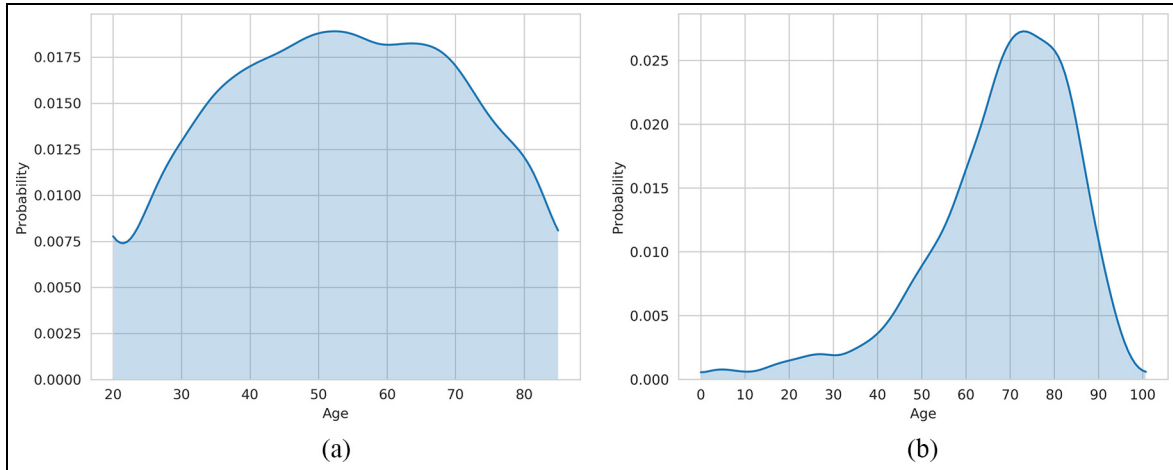
Figure 5. Onset age distribution and death age distribution. (a) Onset age distribution is presented for 81,852 tissues who had BE before age 84 and (b) death age distribution is presented for 99,982 tissues who died before age 100.
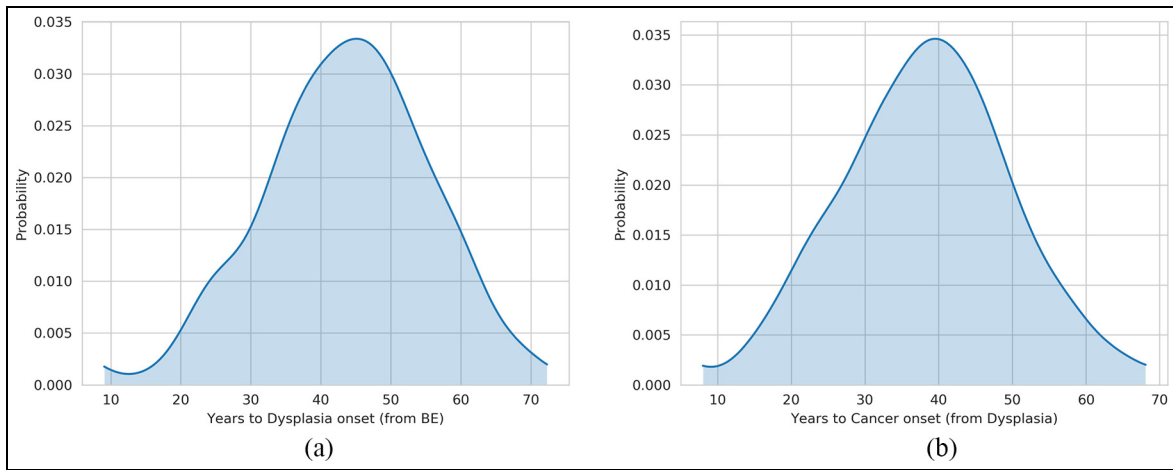

Figure 6. Distributions of time intervals between disease progressions. (a) Distribution of the time between BE and dysplasia occurrences for 38,130 tissues, who had at least one dysplasia segment. (b) Distribution of the time between dysplasia and cancer occurrences for 7027 patients, who had cancer.
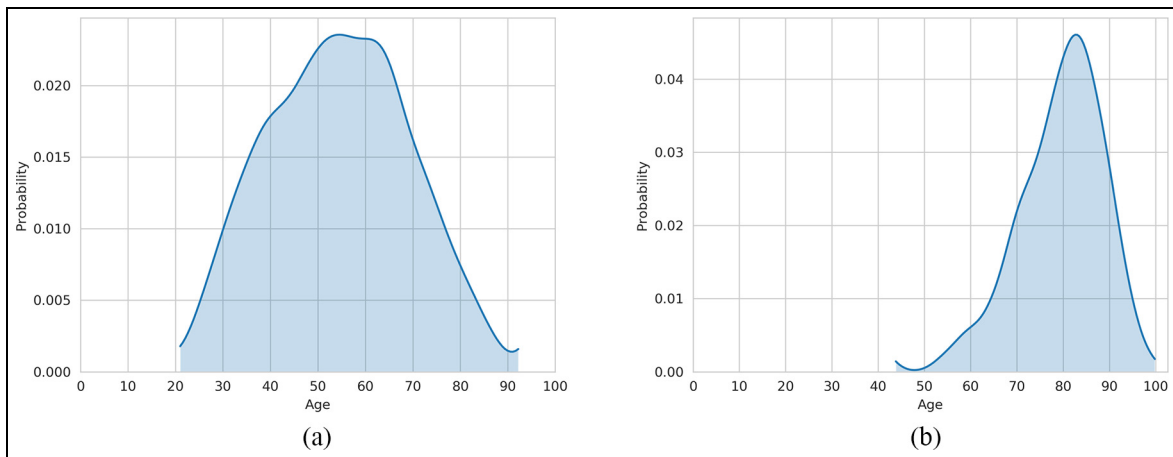
Figure 7. Age distributions of (a) dysplasia and (b) cancer.

age peaks around 55–60 and cancer peaks in the interval of age 80–84 (as expected by Mathieu et al.[32]), which further validates the capability of the model generating realistic emerging patterns that are neither inputted nor calibrated. Also, it means that the rates that are calibrated are realistic and generated historical cancer onset age distribution.

4. Conclusion

We have leveraged prior work on a two-dimensional (2D), time-stepped, multi-scale model[15] to create a 3D discrete event model of the growth of esophogeal cancer. The proposed approach has two benefits over typical agent-based models: (1) time is continuous in the discrete event model, thereby avoiding truncation errors that frequently occur in time-stepped simulations of agent-based models[13] using continuous distributions, and (2) computational effort at each iteration of the simulation algorithm is restricted to active tissue segments, thereby allowing the multi-year growth of malignant tissues to be simulated very rapidly. The model also provides information concerning the 3D spatial progression of the disease that was not considered before. Consequently, it can simulate the invasion of cancer into deep tissues, which creates more realistic segments of EAC when compared to the previous work.[15]

Probability distributions of onset age and death age are taken from data available in the extant literature that are set at the initialization stage. We calibrated the model's other parameters so as to reproduce historical rates of HGD prevalence and missed malignancy rates by incorporating screening of individual patients directly in the simulations. The calibrated scenario is validated against the HGD prevalence versus biopsy sensitivity and missed malignancy rate versus biopsy sensitivity curves that are previously presented in literature.[15] The predicted probability of missed malignancy falls well within the range from literature for all biopsy sensitivity values and exhibits a better fit than the analysis in the previous study.[15] Furthermore, time intervals between progressions in the population are realistic and validated by the findings in the literature. Specifically, cancer age peaks around the age 80–84 years,[32] which is an emerging pattern that is neither inputted nor calibrated.

In addition, the optimization routine is designed to work on parallel computing platforms and benefits from multi-core parallelization to facilitate model calibration in a feasible amount of time. The results of the speed test are promising from the point of view of expanding the existing model toward more realistic simulations (see supplemental materials). Initially, the scalability analysis showed almost linear decrease in runtime with the number of cores utilized. However, due to significant input/output (I/O) operations required to read and write simulation outputs, there is a diminishing return of the number of compute cores leveraged. Most of the parallelization benefits are gained from parallel execution of simulation experiments.

Although the model exhibits good validation against the historical dynamics, one possible extension to the model would be to study the variation of detection for different BE lengths, grid sizes, and tissue areas. The dynamics in the results rely on the assumed data in Table 1 that are based on expert opinions. Those assumptions can be relaxed by exploring different distributions. Another addition to strengthen the model is to adopt different biopsy sensitivity definitions other than the one derived from the study by Curtius et al.[15] Nonetheless, the model demonstrated that the simulated tissues can be used to evaluate the efficiency of existing biopsy protocols. Moreover, the model is not limited to the Seattle protocol and alternative biopsy protocols can potentially be introduced. A simple revision of the PSO implementation would enable us to evaluate alternative biopsy protocols that maximize the detection rate and minimize the false negative rate while using fewer jumbo biopsies than Seattle protocol does. In conclusion, this study provides a tool that can serve as a test-bed to design alternative biopsy protocols and evaluate their efficiency.

ORCID iDs

Ozgur Ozmen (iD) https://orcid.org/0000-0002-5806-0116
James Nutaro (iD) https://orcid.org/0000-0001-7360-2836

Supplementary material

Supplemental material for this article is available online.

References

1. Hur C, Miller M, Kong CY, et al. Trends in esophageal adenocarcinoma incidence and mortality. *Cancer* 2013; 119(6): 1149–1158.
2. Blot WJ, Devesa SS, Kneller RW, et al. Rising incidence of adenocarcinoma of the esophagus and ngastric cardia. *JAMA* 1991; 265(10): 1287–1289.
3. Spechler SJ and Souza RF. Barrett's esophagus. *N Engl J Med* 2014; 371(9): 836–845.
4. Lagergren J, Bergström R, Lindgren A, et al. Symptomatic gastroesophageal reflux as a risk factor for esophageal adenocarcinoma. *N Engl J Med* 1999; 340(11): 825–831.
5. Hayeck TJ, Kong CY, Spechler SJ, et al. The prevalence of Barrett's esophagus in the US: Estimates from a simulation model confirmed by seer data. *Dis Esophagus* 2010; 23(6): 451–457.
6. Hur C, Hayeck TJ, Yeh JM, et al. Development, calibration, and validation of a US white male population-based simulation model of esophageal adenocarcinoma. *PLoS ONE* 2010; 5(3): e9483.
7. Zeigler BP, Praehofer H and Kim TG. *Theory of modeling and simulation: integrating discrete event and continuous complex dynamic systems*. San Diego, CA: Academic Press, 2000.
8. Wainer GA. Advanced cell-devs modeling applications: a legacy of Norbert Giambiasi. *Simulation*. Epub ahead of print 30 April 2018. DOI: 10.1177/0037549718761596.
9. Abela JE, Going JJ, Mackenzie JF, et al. Systematic four-quadrant biopsy detects Barrett's dysplasia in more patients than nonsystematic biopsy. *Am J Gastroenterol* 2008; 103(4): 850–855.
10. Rubenstein JH, Mattek N and Eisen G. Age- and sex-specific yield of Barrett's esophagus by endoscopy indication. *Gastrointest Endosc* 2010; 71(1): 21–27.
11. Zeigler B, Muzy A and Kofman E. *Theory of modeling and simulation*. 3rd ed. San Diego, CA: Academic Press, 2018.
12. Nutaro JJ. *Building software for simulation: theory and algorithms, with applications in C++* . Hoboken, NJ: John Wiley & Sons, 2011.
13. Özmen Ö, Nutaro JJ, Pullum LL, et al. Analyzing the impact of modeling choices and assumptions in compartmental epidemiological models. *Simulation* 2016; 92(5): 459–472.
14. Nutaro J, Ozmen O and Schryver J. Disaggregation and refinement of system dynamics models via agent-based modeling. In: *Proceedings of the 2014 summer simulation multiconference*. Society for Computer Simulation International, p. 11, https://dl.acm.org/doi/10.5555/2685617.2685628
15. Curtius K, Hazelton WD, Jeon J, et al. A multiscale model evaluates screening for neoplasia in Barrett's esophagus. *PLoS Comput Biol* 2015; 11(5): e1004272.
16. Spechler SJ, Sharma P, Souza RF, et al. American gastroenterological association technical review on the management of Barrett's esophagus. *Gastroenterology* 2011; 140(3): e18–e52.
17. Özmen Ö, Pullum LL, Ramanathan A, et al. Augmenting epidemiological models with point-of-care diagnostics data. *PLoS ONE* 2016; 11(4): e0153769.
18. Nguyen HK, Chiong R, Chica M, et al. Understanding the dynamics of inter-provincial migration in the Mekong Delta, Vietnam: an agent-based modeling study. *Simulation* 2021; 97(4): 267–285.
19. Lidbe AD, Hainen AM and Jones SL. Comparative study of simulated annealing, tabu search, and the genetic algorithm for calibration of the microsimulation model. *Simulation* 2017; 93(1): 21–33.
20. Eberhart R and Kennedy J. A new optimizer using particle swarm theory. In: *Proceedings of the sixth international symposium on micro machine and human science, 1995 (MHS'95)*. IEEE, pp. 39–43, https://ieeexplore.ieee.org/document/494215
21. Eberhart RC, Shi Y and Kennedy J. *Swarm intelligence*. London: Elsevier, 2001.
22. Dro J, Ptrowski A, Siarry P, et al. *Metaheuristics for hard optimization: methods and case studies*. New York: Springer, 2006.
23. Weston AP, Badr AS and Hassanein RS. Prospective multivariate analysis of clinical, endoscopic, and histological factors predictive of the development of Barrett's multifocal high-grade dysplasia or adenocarcinoma. *Am J Gastroenterol* 1999; 94(12): 3413–3419.
24. Sharma P, Falk GW, Weston AP, et al. Dysplasia and cancer in a large multicenter cohort of patients with Barrett's esophagus. *Clin Gastroenterol Hepatol* 2006; 4(5): 566–572.
25. O'Conner JB, Falk GW and Richter JE. The incidence of adenocarcinoma and dysplasia in Barrett's esophagus. *Am J Gastroenterol* 1999; 94: 2037–2042.
26. Peters JH, Clark GW, Ireland AP, et al. Outcome of adenocarcinoma arising in Barrett's esophagus in endoscopically surveyed and non-surveyed patients. *J Thorac Cardiovasc Surg* 1994; 108(5): 813–822.
27. Falk GW, Rice TW, Goldblum JR, et al. Jumbo biopsy forceps protocol still misses unsuspected cancer in Barrett's esophagus with high-grade dysplasia. *Gastrointest Endosc* 1999; 49(2): 170–176.
28. Cameron AJ and Carpenter HA. Barrett's esophagus, high-grade dysplasia, and early adenocarcinoma: a pathological study. *Am J Gastroenterol* 1997; 92(4): 586–591.
29. Pourhasanzade F, Sabzpoushan S, Alizadeh AM, et al. An agent-based model of avascular tumor growth: immune response tendency to prevent cancer development. *Simulation* 2017; 93(8): 641–657.
30. Konda VJ and Waxman I. Low risk of prevalent submucosal invasive cancer among patients undergoing esophagectomy for treatment of Barrett's esophagus with high grade dysplasia. *J Gastrointest Oncol* 2011; 2(1): 1–4.
31. Heitmiller RF, Redmond M and Hamilton SR. Barrett's esophagus with high-grade dysplasia. An indication for prophylactic esophagectomy. *Ann Surg* 1996; 224(1): 66–71.
32. Mathieu LN, Kanarek NF, Tsai HL, et al. Age and sex differences in the incidence of esophageal adenocarcinoma: results from the Surveillance, Epidemiology, and End Results (Seer) Registry (1973–2008). *Dis Esophagus* 2014; 27(8): 757–763.

Author biographies

Ozgur Ozmen is a Research and Development Scientist in the Computational Sciences and Engineering division at the Oak Ridge National laboratory, Oak Ridge, TN. He

holds an Industrial Engineering degree from Yildiz Technical University in Istanbul, Master of Engineering Management degree from Galatasaray University in Istanbul, MISE, and PhD Degrees (both in Industrial and Systems Engineering) from Auburn University in Alabama.

James Nutaro is a Distinguished Research and Development Scientist in the Computational Sciences and Engineering Division of Oak Ridge National Laboratories. He holds a PhD in Computer Engineering from the University of Arizona. His email address is nutarojj@ornl.gov.

Sassan Ostvar is a Postdoctoral Scientist at Columbia University in the City of New York. He holds a PhD in Environmental Engineering from Oregon State University.

DR. Chin Hur is the Herbert and Florence Irving Professor of Medicine to Honor Dr. Jeffrey Alan Stein and Professor of Epidemiology and Director, Healthcare Innovations Research and Evaluation at Columbia University Medical Center. He received his MD and MPH from the Johns Hopkins University School of Medicine.

Chung Yin Kong is an Associate Professor of Medicine at the Icahn School of Medicine at Mount Sinai. He received his B.S. (with honor) in Physics from Stony Brook University, NY, USA and obtained his Ph.D. from the University of Massachusetts at Amherst, MA, USA. His research interests are in utilizing simulation modeling to inform medical decision making and health policy. He has developed multiple simulation models to determine the optimal strategies for the screening, surveillance, and treatment of cancer patients.