Proceedings of the 2024 Winter Simulation Conference
H. Lam, E. Azar, D. Batur, S. Gao, W. Xie, S. R. Hunter, and M. D. Rossetti, eds.

METAMODEL OF A SIMULATION MODEL OF COLORECTAL CANCER WITH DIVERSE CLINIC POPULATIONS AND INTERVENTION SCENARIOS

Ashley Stanfield[1], Maria E. Mayorga[1], Meghan C. O'Leary[2,3], and Kristen Hassmiller Lich[3]

[1]Dept. of Industrial and Systems Engineering, North Carolina State University, Raleigh, NC, USA
[2]Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC, USA
[3]Dept. of Health Policy and Management, University of North Carolina, Chapel Hill, NC, USA

ABSTRACT

Colorectal cancer (CRC) prevention is dependent on increasing screening rates, a strategy proven effective in reducing cancer cases and potential life years lost. Simulation models of CRC can be used to project expected outcomes associated with different evidence-based interventions. However, traditional simulation for each population of interest is computationally intensive and requires a model expert. To address this, we proposed a metamodeling approach, considering various techniques such as linear regression and random forest. By creating a metamodel of the simulation, decision makers can generate both individual and population-level estimates directly and instantaneously. We aimed to create a metamodel of an existing CRC simulation model that can be adapted for different interventions and populations to predict cancer cases averted and life years lost.

1 INTRODUCTION

Colorectal cancer (CRC) is the second most common cause of death by cancer in the United States. It was estimated that there would be approximately 153,000 new cases nationally in 2023, with 13 percent of cases among individuals under age 50, 32 percent among those 50-64 years, and 56 percent among those 65 years and older (Siegel et al. 2023). CRC is the only type of cancer in which increases in screening have been proven to reduce cancer mortality (Ladabaum et al. 2020). Additionally, it has been shown that a majority of CRC deaths in the United States can be attributed to lack of screening (Meester et. al 2015). Stool-based screenings, such as fecal immunochemical tests (FIT) and fecal occult blood tests (FOBT), are recommended to be completed annually, while colonoscopies are recommended every 10 years among average-risk individuals ages 45-75. (US Preventive Services Task Force 2021). The implementation of evidence-based interventions (EBIs) (CDC 2016), including mailed outreach, patient navigation, and provider reminders, have potential to increase CRC screening rates in diverse populations.

In this paper, we build on and create a metamodel of the NC-CRC simulation model (Koutouan et al. 2021; Davis et al. 2019), which was previously developed by members of our team and named because of our original focus on the population of North Carolina residents who were age-eligible for CRC screening. The NC-CRC model is a microsimulation model of the progression of CRC that included: polyp incidence and development, individual health states over time, and screening and surveillance. Individuals within a population are modeled from birth until death by CRC or by another cause. Individuals can develop noncancerous or cancerous polyps which, if not removed in time, can cause a CRC death. Modeling the implementation of EBIs within a population allows the possibility for polyps to be identified and removed before they develop into cancer or progress to a CRC death.

The NC-CRC simulation model has been validated and used in prior research to simulate screening prevalence and long-term CRC outcomes including cancer cases averted, cancer deaths averted, and life years gained under various intervention scenarios in specific input populations. For example, this model has previously been used extensively at the state population level to estimate the health impact of different policy scenarios such as Medicaid expansion and the Affordable Care Act (Hassmiller Lich et al. 2019; Powell et al. 2020), and interventions including endoscopy expansion, vouchers for individuals without health insurance, mailed reminders, mass media campaigns, and patient navigation (Hicklin et al. 2022; Davis et al. 2019; Hassmiller Lich et al. 2017).

More recently, Koutouan et al. (2021) tailored the NC-CRC model specifically to be used to estimate the lifetime health impact of the Centers for Disease Control and Prevention's (CDC) Colorectal Cancer Control Program (CRCCP), a national program focused on improving CRC screening in low-income, medically underserved populations by funding grantees to provide and promote CRC screening (Bitler et al. 2021; Hannon et al. 2013; Joseph et al. 2011). Koutouan used data for individuals who received a CRC screening either by stool test or colonoscopy through the CRCCP from the years 2009-2020, with a total of 62,682 unique patients.

Using this model specifically for the CRCCP population has been successful in estimating the substantial health impact of the direct screening services provided through the CRCCP. However, to date, members of our modeling and research team have needed to run the various intervention scenarios on behalf of our public health partners. Ultimately, to inform programmatic planning for a public health initiative like the CRCCP and to guide implementation of interventions in individual clinics with varying populations and contexts, it will be critical for clinical and program staff and leadership to have a tool that can help them make decisions efficiently and independently without the help of a modeling expert. Given that the contexts within these programs are diverse, and that there are many interventions that can be considered for supporting population-level screening, being able to run a sufficient number of input populations and intervention scenarios with the original model would be a large undertaking.

Therefore, in this paper, our goal was to create a metamodeling approach that allows us to use the existing simulation model to estimate the effectiveness of interventions for any population of interest and for a wide range of intervention scenarios. This metamodel can then be transformed into a user-interface which public health practitioners and clinicians without modeling experience can use to independently estimate outcomes of interest based on their program's own population and specific types of interventions being explored.

The need for this type of approach and practical tool was further highlighted in our prior work with diverse stakeholders. O'Leary et al. (2023) collected stakeholder feedback on the use of a CRC screening simulation tool, also based on the NC-CRC model, for informing their decision-making related to EBI implementation. Following an initial demonstration of the simulation tool and the model inputs and outputs, participants (n=17) expressed an interest in using tools like this for decision-making and believed the tool could help them to implement strategies for improving CRC screening. However, 76% of participants expressed that the simulation model was too "academic" or research focused (i.e., not a pragmatic tool that could be used by diverse practitioners with varying levels of modeling experience), and 71% of participants had a lack of confidence that EBI implementation results would be appropriate for local settings where the population inputs differed from the small set of population inputs in that model. Additionally, 35% of participants found that the microsimulation tool was not comprehensive of all EBIs they might consider. Thus, practitioners and other types of decision-makers need a tool that is more customizable to their local context and comprehensive of intervention scenarios, yet able to be used quickly and with limited technical support.

A metamodel, also known as a surrogate model, can be used to approximate the outputs of a more complex model. Metamodels are applied to models where an input is provided by the user, and the model provides an output. The inputs are the different parameters that can be changed in the system and the outputs are the measures of the model whose values can change depending on the input values. For example, in our work, the inputs were varying populations and screening rates, and the outputs were cancer cases and life

years lost. Since the progression of CRC looks different for individuals depending on their age, race and gender, changing the population changed the proportion of cancer cases as well as the number of life years lost. The original model is often a computationally expensive simulation such as the NC-CRC model. Original model outputs given the user input are considered to be accurate. Metamodels seek to produce highly similar outputs for the same input, often with the advantage of significantly reduced computational time.

Metamodel techniques include machine learning, regression, ensembles, and neural networks. Each metamodeling technique has a blend of advantages and limitations which must be considered to determine its appropriateness for a given situation. Some common characteristics of metamodels to consider include interpretability, ease of use, level of statistical expertise required, and performance matching the original model outputs given the same inputs (Degeling et al. 2020; McCandlish et al. 2022). Interpretability in this case refers to the ability of a user to understand the relationship between the inputs and outputs given the metamodel results. High-performing metamodels are able to closely estimate outputs for all input parameters across the input space.

Metamodels are developed through the following steps: sample input parameters to be run through the simulation model; run those parameters through the model; generate a metamodel using the inputs chosen and the associated outputs; and test the metamodel by running another scenario through both the original simulation model and the metamodel to compare estimates. A unique aspect of our model is that we utilized metamodels at the individual level and aggregated at the population level to estimate outcomes for any given population, see Figure 1.

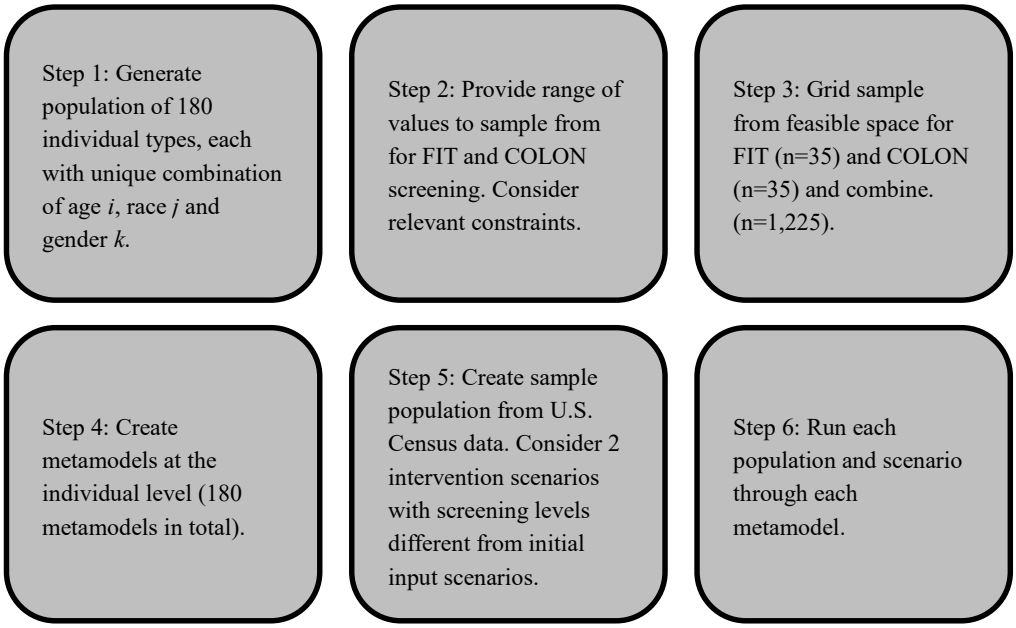

Figure 1: Metamodeling approach for cancer cases and life years lost at the population level.

In healthcare and public health settings, it is important to consider common inputs and outputs and the general tendencies of these variables. Health care decisions can be informed by new information related to new technologies or interventions (McCandlish et al. 2022), new information related to already existing technologies or interventions, or increased access to technology or interventions. Metamodels may be created based on simulation models with the goal of providing feedback for policymakers in real-time based on changes that have occurred since the original simulation model. Further, in health settings, metamodels can be used as the basis of online decision support systems which can be used by any user with relative ease to see the impact of a change in the input parameters on the outputs. These metamodeling techniques have previously been used in health settings to conduct value of information and cost effectiveness analysis, as

well as other analyses. However, there remains a dearth of applications of metamodels to health settings in the literature (Degeling et al. 2020).

Many simulation models have been generated to assess CRC screening delivery, typically for the purpose of assisting decision-makers with implementing interventions to improve screening. However, according to a systematic review that assessed 43 different publications related to CRC screening simulation models, only 12% have been used in practice for decision making (Smith et al. 2020). We believe that by creating a metamodel of a validated and well accepted simulation of CRC, such as the NC-CRC model, we will reduce the effort, time, and expertise required to use the model and thus increase model usage by stakeholders to inform decision-making.

2 LITERATURE REVIEW OF METAMODELING

Metamodel methods reviewed include linear regression, Gaussian process, machine learning, random forest, neural networks, and ensemble methods. Linear regression metamodels are effective in modeling the relationship between inputs and outputs in the case where the inputs and outputs have a moderately strong to strong linear relationship. Linear regression metamodels can handle sparse data (Koffijberg et. al 2021). These models are the easiest to interpret compared to other metamodels in this report. Their inclusion of coefficients, confidence intervals and *p* values allow for the understanding of parameter contributions to the model. It also allows users to understand the levels of uncertainty. Many software packages can easily perform linear regression in a time efficient way. One of the most significant limitations related to linear regression metamodels is that they lack the robustness to model nonlinear relationships effectively. Making the necessary adjustments to create linear regression metamodels which can better account for nonlinear relationships is time consuming and is more susceptible to biased estimates. For example, a full order second model with 40 input parameters would require 902 coefficients (Hastie et al. 2001). Additionally, decision-making models are becoming increasingly complex, and therefore including many coefficients may be required for linear regression metamodels, decreasing overall interpretability. Overall, if a simulation has a moderately strong to strong linear relationship between inputs and outputs and few coefficients, linear regression is an attractive metamodeling technique (McCandlish et al. 2022).

Gaussian process metamodels use the spatial distance between inputs and outputs. They are significantly better than linear regression metamodels at modeling nonlinear relationships between inputs and outputs. Additionally, they have a smooth prediction function, meaning that some of the noise and variation is reduced. However, one of the setbacks associated with Gaussian process metamodels is that they require statistical expertise to capture the relationship between inputs and outputs accurately. Additionally, most Gaussian process software cannot handle more than 30 parameters and may take days or weeks to train for datasets with at least 10,000 observations (Rojnik & Naveršnik 2008). These metamodels assume that the output distribution is a joint Gaussian distribution. The details of the Gaussian process metamodels are specified by the mean as well as the covariance. The output values are not overly sensitive to input values, meaning small changes in input values should result in small or no changes in output values (Zhong et al. 2022). An additional benefit is that Gaussian process metamodels are particularly suited for problems with a low number of observations and input parameters (<25) (Koffijberg et al. 2021).

Machine learning methods are known to perform better than traditional statistical models when high degrees of nonlinearity exist between input and output variables, or the variables are noncontinuous. In the context of healthcare, a frequently desired output is cost-effectiveness. This particular output is nonlinear, and the outputs are highly sensitive to inputs. There are few examples of metamodeling using machine learning in healthcare settings to date (McCandlish et al. 2022).

Random forest is an ensemble machine learning method (Zhong et al. 2022). It creates decision trees and uses a maximum voting or averages over trees to calculate outputs. Random forest is known to perform better than other methods such as linear regression or other machine learning methods (McCandlish et al. 2022), partially due to the fact that it makes no initial assumptions about the data (Zhong et al. 2022). Random forest can be conducted using the R software package *ranger*. While it is not as easily interpretable as linear regression metamodels with few coefficients, there still remains some degree of interpretability in

Random forest, which is not always true of machine learning metamodeling techniques. Additional advantages include measures of which input parameters are most significant and the ability to cluster observations based on characteristics displayed in the inputs and outputs (McCandlish et al. 2022).

3 METHODOLOGY

3.1 Population

The population input file parameters which impact the progression of CRC include age, race, and gender. The age range at which adults are eligible for CRC screening is 45-74, reflecting current screening recommendations (US Preventive Services Task Force 2021). The age in the model is the age of the individual in the year of intervention. Additionally, CRC progression differs between races (white, black and other) and between genders (male and female). Therefore, our metamodel must account for each unique combination of these parameters. Our input population consisted of 180 individuals, one representing each unique combination of these three parameters. A given output X could then be represented the following way.

X_{ijk} = value of outcomes X for a person of age *i* with race *j* and gender *k*.

3.2 Intervention Scenarios

The two types of screening tests included in this model are colonoscopy and FIT. There are two types of colonoscopies included in the model. A routine colonoscopy occurs when someone with colonoscopy modality undergoes a colonoscopy as their method of screening. A diagnostic colonoscopy occurs when someone with FIT modality gets a positive FIT result, and undergoes a diagnostic colonoscopy to check for colorectal cancer. In our model, we assume that each person starts with a FIT modality. Then, if they get a positive FIT and a diagnostic colonoscopy, their modality switches to colonoscopy. We define these as the interventions in this model because the screenings were provided to eligible patients through the CRCCP. In previous work, the model has been used to test the impact of an intervention within a given year on outcomes of interest. It has included a screening rate during the year of intervention for each modality, as well as a screening rate for the years post-intervention for each modality. In this work, we added a preintervention screening rate, once a given person in the population is age-eligible for screening. These screening rates are reflective of potential adherence levels within the population. In other words, these variables are estimates of what percentage of the age-eligible population will screen under their assigned modality each year. In our work, we consider 6 different intervention screening rate parameters. These parameters include FIT_{before}, FIT_{during}, FIT_{after}, $COLON_{before}$, $COLON_{during}$, $COLON_{after}$. FIT_{before} and $COLON_{before}$ represent screening rates for each modality before the year(s) of intervention. FIT_{during} and $COLON_{during}$ represent the screening rates for each modality during the year(s) of intervention. In this work, we consider a 5-year intervention period, starting with the intervention year provided to the model as year 1. We made this decision since grantees of the CRCCP are funded for 5 years. FIT_{after} and $COLON_{after}$ represent the screening rates for each modality after the years of intervention (for the unique lifetime of each individual). We consider the colonoscopy parameters as the adherence to routine colonoscopy, while we kept the adherence to diagnostic colonoscopy constant throughout our experimentation.

3.3 Sampling of Intervention Scenarios

We sampled intervention scenarios which we could run through the model and use the inputs and outputs generated to create a metamodel. In theory, these 6 different adherence levels could vary anywhere from 0 to 1 probabilistically; however, ranges seen in practice are much more narrow. For FIT, we considered the lower bound for annual screening adherence before, during or after the years of intervention as 6%, and the upper bound for screening adherence before, during or after the years of intervention as 11%. For colonoscopy, we considered 30% and 66% for the lower and upper bounds of all colonoscopy parameters,

respectively. These estimates came from the literature as explained in Koutouan (2023). We considered screening rates that fell approximately halfway between the lower and upper bound for both modalities as medium adherence. Screening rates that fell halfway between low and medium screening and medium and high screening were also considered as low-medium adherence and medium-high adherence, respectively. The different parameters that we sampled from are summarized below. For each parameter, one of the listed values was selected. (Table 1)

Table 1: Summary of sample space for each intervention parameter.

Intervention Parameter	Sample Space
FIT_{before}	(6,7,8,9.5,11)
FIT_{during}	(6,7,8,9.5,11)
FIT_{after}	(6,7,8,9.5,11)
$COLON_{before}$	(30,39,48,57,66)
$COLON_{during}$	(30,39,48,57,66)
$COLON_{after}$	(30,39,48,57,66)

Before combining these 6 parameters to create different scenarios we applied these constraints to avoid generating intervention scenarios that were unlikely to occur and thus less informative. From previous work, we observed that screening rates during the year(s) of intervention are at least as high as the year(s) of intervention and are most commonly higher. This makes sense, since you would expect a program that aims to increase screening rates in a given population to have a positive impact or, at worst, no impact. Additionally, we noted that the screening rates typically did not increase after the intervention period since funding is no longer being provided. However, while screening rates do not typically increase after the intervention from what they were during the intervention period, screening rates after year(s) of intervention were consistently higher than screening rates before the year(s) of intervention. In summary, our constraints ensured that prior screening was always less than or equal to screening post-intervention, and screening post-intervention was always less than or equal to screening during the intervention period (e.g. $FIT_{before} \leq FIT_{after} \leq FIT_{during}$).

From these screening parameter values and constraints we generated different scenarios. Each scenario consisted of a different set of the six parameters. We exhausted every combination of the FIT and colonoscopy parameters and then removed the ones that did not adhere to the constraints listed above. We considered 35 different FIT scenarios and 35 different colonoscopy scenarios and then evaluated each FIT scenario and each colonoscopy scenario together for a total of 1,225 different scenarios. (First 35 scenarios had FIT scenario 1 with colonoscopy scenarios 1-35, second 35 scenarios had FIT scenario 2 with colonoscopy scenarios 1-35 etc.)

3.4 Outcomes of Interest

For this paper, we are interested in cancer cases and life years lost. Cancer cases represent the number of CRC cases after screening. At the individual level, cancer cases take on the value of 0 or 1, representing whether an individual did not or did develop CRC after the screening, respectively.

Individuals are given a life expectancy based on life expectancy tables (Arias 2012) and these values are used to calculate life years lost, which is the difference between the life expectancy from the tables and the lifetime of an individual who dies from CRC. For a given individual, these outcomes can be summarized the following way:

CC_{ijk} = Cancer case estimate for a person at age *i* with race *j* and gender *k*.
LYL_{ijk} = Life years lost estimate for a person at age *i* with race *j* and gender *k*.

For each metamodeling technique, we show performance metrics but not results of the metamodel themselves, due to difficulty in showing the results in 6 dimensions (FIT_{before}, FIT_{during}, FIT_{after}, $COLON_{before}$, $COLON_{during}$, $COLON_{after}$).

3.5 Aggregating Outcomes at the Population Level

We estimate the model outputs for each person type in a population, as opposed to a population in its entirety. Unlike the previous works based on this model, our results are not population estimates of outcomes, but we want to find outcomes at the individual level. We are able to generate individual level estimates and then aggregate into population level estimates since individuals do not interact or share resources in our model, as they might in other types of health simulation models such as infectious disease modeling. Therefore, once we find estimates at the individual level, we can create a weighted average of the estimates for different individual types to represent a unique clinic population. The clinic level outcomes are calculated by equations 1 and 2.

P_{ijk} = number of persons of age i, race j and gender k in a given clinic population

$$\text{Clinic CC} = \sum_i \sum_j \sum_k CC_{ijk} P_{ijk} \quad (1)$$

$$\text{Clinic LYL} = \sum_i \sum_j \sum_k LYL_{ijk} P_{ijk} \quad (2)$$

3.6 Running the Simulation Model for Individual Level CC and LYL Estimates

We ran 5,000 replications (r) for each of the 180 individuals (i) in each of the 1,225 sampled scenarios (s). Therefore, we generated results for 1,102,500,000 individuals. We used the outcomes of the 1,225 different scenarios to create 180 different metamodels per metamodeling technique, one for each person type. Each scenario took approximately 104 seconds to complete. The total computational time is linearly associated with the number of sampled intervention scenarios, assuming the number of replications and the population size is the same for each intervention scenario. Therefore, the computational effort is dependent on how many samples are considered. As the number of replications and the population size increases, running one scenario through the model is computationally intensive. Creating population level estimates using the metamodel takes almost no time, since it is a matter of multiplying the individual level estimate for a given individual type by the number of individuals in the population. Once individual estimates are available, no computation time is required, and we can consider many different intervention scenarios.

4 RESULTS

4.1 Linear Regression Metamodel

The first metamodeling technique we selected was linear regression since it is commonly used and more interpretable than other metamodeling techniques. We created 180 different linear regression metamodels for both cancer cases and life years lost. In each linear regression model, our response variable was the output of interest (cancer cases or life years lost), and our 6 predicting variables were the screening rates of FIT_{before}, FIT_{during}, FIT_{after}, $COLON_{before}$, $COLON_{during}$, and $COLON_{after}$.

Consider $\beta_0{}^C$ and $\beta_0{}^{LY}$ as the intercepts and $\beta_l{}^C$ and $\beta_l{}^{LY}$ as the coefficients provided by a given metamodel for a person of age i, race j, gender k and screening parameter l for cancer cases and life years lost respectively. Consider S_l as the screening rate for each parameter.

i = 45,46,…,74, j=1,2,3 , k=1,2, l=1,2,3,4,5,6 where 1=FIT_{before}, 2=FIT_{during}, 3=FIT_{after}, 4=$COLON_{before}$, 5=$COLON_{during}$, and 6=$COLON_{after}$.

$$CC_{ijkl} = \beta_0{}^{C} + \sum_{l=1}^{6} \beta_l^{C} S_l \quad (3)$$

$$LYL_{ijkl} = \beta_0{}^{LY} + \sum_{l=1}^{6} \beta_l^{LY} S_l \quad (4)$$

For each linear regression metamodel, multiple R squared values based on the 1,225 scenarios were calculated to assess performance of the individual metamodels. The results are shown in Figure 2.

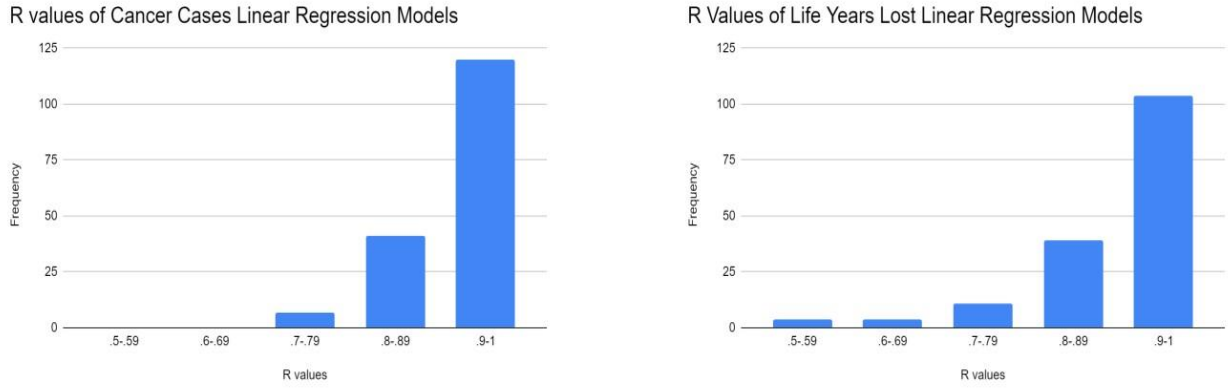

Figure 2: Linear regression R values for cancer cases and life years lost.

4.2 Polynomial Regression Metamodel

We also created 180 different polynomial regression metamodels for both cancer cases and life years lost. This metamodeling technique maintained some level of interpretability but was not as interpretable as the linear regression. Consider the following equations to estimate the cancer cases and life years lost for an individual of age *i*, race *j*, gender *k* and screening parameter *l*.

$$CC_{ijkl} = \beta_0^{C} + \sum_{l=1}^{6} \beta_l^{C} S_l + \beta_l^{C} S_l^2 \quad (5)$$

$$LYL_{ijkl} = \beta_0^{LY} + \sum_{l=1}^{6} \beta_l^{LY} S_l + \beta_l^{LY} S_l^2 \quad (6)$$

Similar to the linear regression metamodels, we found multiple R^2 values for the polynomial regression approach. These R^2 values, based on the 1,225 test scenarios, are shown in Figure 3.

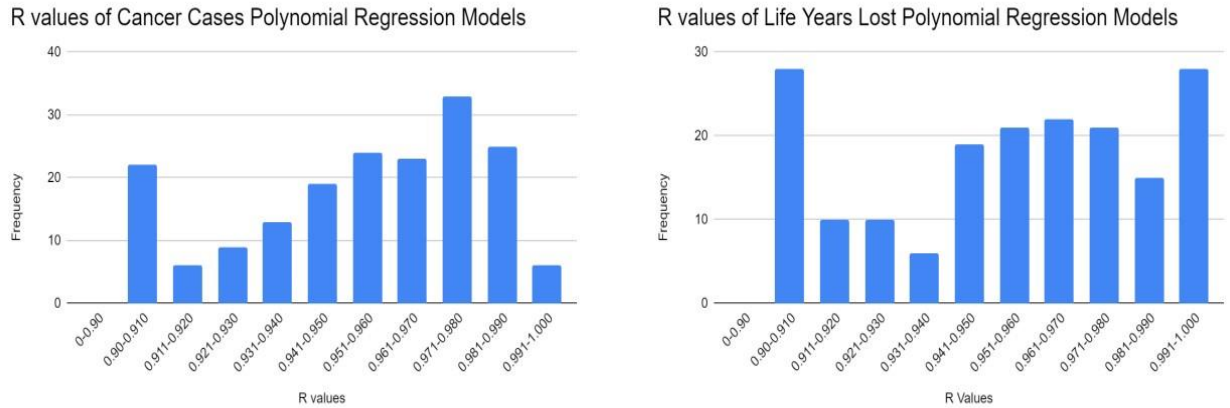

Figure 3: Polynomial regression R values for cancer cases and life years lost.

4.3 Random Forest Metamodel

The final metamodeling technique was random forest, which has lower interpretability but the potential of higher performance. For each person type, we used a random forest metamodel to estimate cancer cases and life years lost. Each random forest metamodel had 500 trees. All values of percent variance explained were above .95 for both cancer cases and life years lost.

4.4 Metamodel Performance on Test Population

The three different metamodeling approaches and their corresponding 180 metamodels, one per individual type, were used to generate population level estimates. These estimates were then compared to the actual values provided by the simulation model with the same input parameters (which were different than the input parameters used to build the model).

We considered a population of 1,500 individuals to test performance. We used data from the U.S. Census Bureau related to the proportion of people in different age brackets and different genders and races to inform our approach (Bureau 2022).

We considered 2 different intervention scenarios. Scenario 2 had screening levels very close to one of the scenarios we tested, and Scenario 1 had screening levels that were halfway between a few different scenarios we tested. This allowed us to test if proximity to sampled scenarios used to build the metamodels impacted metamodel performance. The two screening scenarios are provided in Table 2.

Table 2: FIT and COLON screening parameters for the two screening scenarios tested.

	FIT_{before}	FIT_{during}	FIT_{after}	$COLON_{before}$	$COLON_{during}$	$COLON_{after}$
Scenario 1	9.875	10.625	10.250	59.25	63.75	61.50
Scenario 2	10.9	10.9	10.9	65.4	65.4	65.4

We ran the population through each scenario in our original NC-CRC model as well as the three different metamodels provided. We then calculated the mean absolute percentage error (MAPE) per-person between the actual values and the estimates provided by each metamodel for both outcomes: cancer cases and life years lost. Figure 4 summarizes our findings for 10 replications of the NC-CRC simulation model.

For both outcomes we see very good performance by all models. For the cancer cases outcome, we see similar MAPE values across the metamodels for both scenarios. For the life years lost outcome, we observe that for both scenarios, the polynomial regression and random forest metamodels have lower average MAPE values than that of linear regression, but the range of MAPE values across the 10 replications are similar, suggesting that these metamodels perform similarly for life years lost.

5 CONCLUSION

This work provided an initial metamodeling strategy for the NC-CRC simulation model. It considered screening parameters before year(s) of intervention, during year(s) of intervention, and after year(s) of intervention. It also considered a variety of intervention scenarios, with each screening parameter ranging from low to high values. We built upon previous work using this simulation model to predict population outcomes and built metamodels that could provide useful information when it comes to predicting cancer cases and life years lost both at the individual and the population level. This work found that more computationally expensive metamodeling techniques (random forest, polynomial regression) had lower average MAPE values at the individual level as opposed to simpler techniques (linear regression) for life years lost, and similar performance for cancer cases. From this, we learned that the best metamodeling approach when combining models to create a population estimate may be different across outcomes, and depend on the desired level of interpretability.

6 FUTURE WORK

This work considered a population where each individual was assigned an initial modality of FIT. If they had a positive FIT result, they received a diagnostic colonoscopy and their modality switched to colonoscopy in future years. Future work could address individuals whose modality begins with routine colonoscopy. One limitation of this work is that it did not consider correlations in the outputs. This could be addressed in future work. Additionally, we could consider a "no intervention" scenario. This scenario would allow us to be able to generate averted outcomes, which represent the difference between the number of cancer cases after if there is an intervention with higher screening levels versus if there is no intervention. More test scenarios also have the potential of increasing metamodel performance. Future work could sample from a range of screening levels using other techniques such as Latin hypercube sampling.

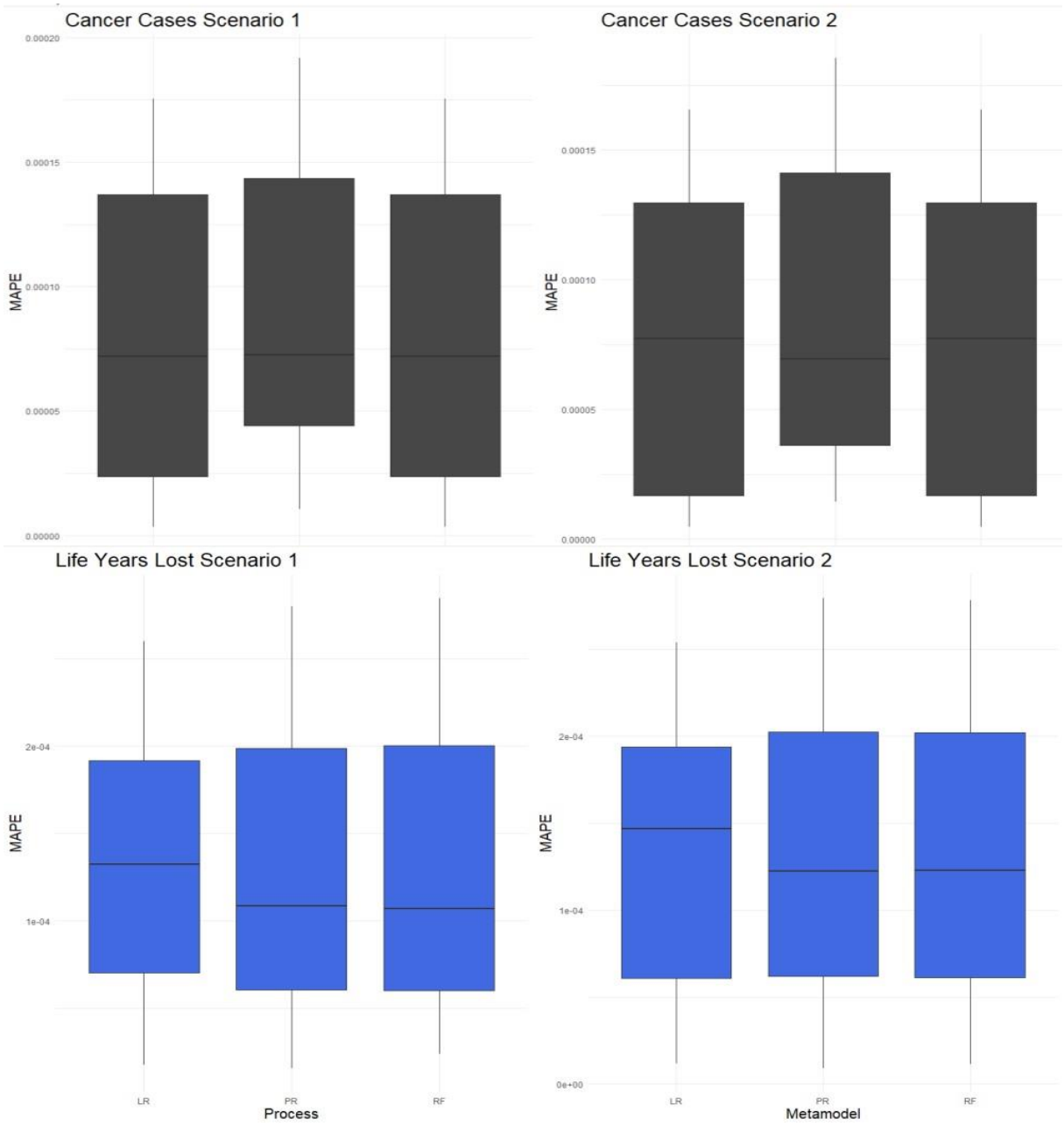

Figure 4: Cancer cases (top) and life years lost (bottom) mean absolute percent error for each intervention scenario tested.

ACKNOWLEDGMENTS

We would like to thank the Division of Cancer Prevention and Control at the Centers for Disease Control and Prevention (CDC) for sharing both the individual-level and clinic-level Colorectal Cancer Control Program data used in this analysis. This work is supported by the National Association of Chronic Disease Directors (NACDD), Grant No. 5- NU38OT000286. The content provided is solely the responsibility of the authors and does not necessarily represent the official views of the CDC or NACDD.

REFERENCES

Arias, E. 2012. "United States life tables, 2008." *National Vital Statistics Reports*, 61(3).

Bitler, M. P., C. S. Carpenter, and D. Horn. 2021. "Effects of the colorectal cancer control program". *Health Economics*, 30(11):2667–2685.

Centers for Disease Control and Prevention. 2016. "Guidance for Measuring Colorectal Cancer (CRC) Screening Rates in Health System Clinics". US Department of Health and Human Services.

Davis, M. M., S. Nambiar, M. E. Mayorga, E. Sullivan, K. Hicklin, M. C. O'Leary, et al. 2019. "Mailed FIT (fecal immunochemical test), navigation or patient reminders? Using microsimulation to inform selection of interventions to increase colorectal cancer screening in Medicaid enrollees". *Preventive Medicine*, 129.

Degeling, K., M. J. IJzerman, M. S. Lavieri, M. Strong, and H. Koffijberg. 2020. "Introduction to metamodeling for reducing computational burden of advanced analyses with Health Economic Models: A structured overview of metamodeling methods in a 6-step application process". *Medical Decision Making*, 40(3):348–363.

Haghighat, S., D. A. Sussman, and A. Deshpande. 2021. "US Preventive Services Task Force Recommendation Statement on screening for colorectal cancer". *The Journal of the American Medical Association*, 326(13):1328.

Hannon, P. A. , A. E. Maxwell, C. Escoffery, T. Vu, M. Kohn, J. Leeman, M. L. Carvalho, D. J. Pfeiffer, A. Dwyer, M. E. Fernandez, S. W. Vernon, L. Liang, and A. DeGroff. 2013. "Colorectal cancer control program grantees' use of evidence-based interventions". *American Journal of Preventive Medicine, 45*(5):644–648.

Hassmiller Lich, K. *et al.* 2019. "Estimating the impact of insurance expansion on colorectal cancer and related costs in North Carolina: A population-level simulation analysis". *Preventive Medicine*, 129:105847.

Hassmiller Lich, K., Cornejo, D. A., Mayorga, M. E., Pignone, M., Tangka, F. K. L., Richardson, L. C. et al. 2017. "Cost effectiveness analysis of four simulated colorectal cancer screening interventions, North Carolina". *Preventing Chronic Disease, 14*.

Hastie, T., R. Tibshirani, and J. H. Friedman. 2001. *The elements of statistical learning: Data mining, inference, and prediction*. New York: Springer.

Hicklin, K., M. C. O'Leary, S. Nambiar, M. E. Mayorga, S. B. Wheeler, Davis et al. 2022. "Assessing the impact of multicomponent interventions on colorectal cancer screening through simulation: What would it take to reach national screening targets in North Carolina?" *Preventive Medicine*, 162, 107126.

Joseph, D. A., A. S. DeGroff, N. S. Hayes, F. L. Wong, and M. Plescia. 2011. "The colorectal cancer control program: Partnering to increase population level screening". *Gastrointestinal Endoscopy*, 73(3):429–434.

Koffijberg, H., K. Degeling, M. J. IJzerman, V. M. Coupé, and M. J. Greuter. 2021. "Using metamodeling to identify the optimal strategy for colorectal cancer screening". *Value in Health*, 24(2):206–215.

Koutouan, P. R. 2023. *The Effect of Screening Behavior Assumptions and Heterogeneity on Preventive Health Interventions Evaluation using Simulation Modeling.* Ph.D. thesis, Department of Industrial Engineering, North Carolina State University, Raleigh, North Carolina.

Koutouan, P. R., M. E. Mayorga, M. C. O'Leary and K. Hassmiller Lich. 2021. "Creating Simulated Equivalents to Project Long Term Population Health Outcomes of Underserved Patients: An Application to Colorectal Cancer Screening". In *2021 Winter Simulation Conference (WSC)*, Phoenix, AZ, USA, 2021, pp. 1-12, doi: 10.1109/WSC52266.2021.9715374.

Ladabaum, U., J. A. Dominitz, C. Kahi, and R. E. Schoen. 2020. "Strategies for colorectal cancer screening". *Gastroenterology*, 158(2):418–432.

McCandlish, J. A., T. Ayer, & J. Chhatwal. 2022. "Cost-Effectiveness and Value-of-Information Analysis Using Machine Learning–Based Metamodeling: A Case of Hepatitis C Treatment". *Medical Decision Making*, 43(1):68–77.

Meester, R. G., C. A. Doubeni, A. G. Zauber, S. L. Goede, T. R. Levin, D. A. Corley, A. Jemal, and I. Lansdorp-Vogelaar. 2015. "Public health impact of achieving 80% colorectal cancer screening rates in the United States by 2018". *Cancer*, 121(13):2281–2285.

O'Leary, M. C., K. H. Lich, M. E. Mayorga, K. Hicklin, M. M. Davis, A. T. Brenner, D. S. Reuland, S. A. Birken, and Wheeler, S. B. 2023. "Engaging stakeholders in the use of an interactive simulation tool to support decision-making about the implementation of colorectal cancer screening interventions". *CCC. Cancer Causes & Control/CCC, Cancer Causes & Control, 34*(S1):135–148.

Powell, W., L. Frerichs, R. Townsley, M. Mayorga, J. Richmond, G. Corbie-Smith, S. Wheeler, and K. Hassmiller Lich. 2020. "The potential impact of the Affordable Care Act and Medicaid expansion on reducing colorectal cancer screening disparities in African American males". *PLOS ONE*, 15(1).

Rojnik, K., and K. Naveršnik. 2008. "Gaussian Process Metamodeling in Bayesian Value of Information Analysis: A Case of the

Complex Health Economic Model for Breast Cancer Screening". *Value in Health*, 11(2):240–250.

Siegel R. L., N.S. Wagle, A. Cercek, R. A. Smith, A. Jemal. 2023. "Colorectal cancer statistics, 2023". *CA Cancer J Clin*. 73(3):233254.

Smith H., P. Varshoei , R. Boushey, C. Kuziemsky. 2020. "Simulation modeling validity and utility in colorectal cancer screening delivery: A systematic review". *J Am Med Inform Assoc.* 27(6):908-916.

United States Census Bureau. 2022. 2020 Census. Census.gov. https://www.census.gov/2020census, accessed 3 February 2024.

Zhong H., M. L. Brandeau, G. E. Yazdi, et al. 2022. "Metamodeling for Policy Simulations with Multivariate Outcomes". *Medical Decision Making*. 42(7):872-884.

AUTHOR BIOGRAPHIES

ASHLEY STANFIELD is a doctoral student in the Edward P. Fitts Department of Industrial and Systems Engineering at North Carolina State University. She earned her B.S. in Industrial Engineering from North Carolina State University. Her research interests include simulation modeling for decision-making in the context of healthcare. Her email address is aestanfi@ncsu.edu.

MARIA E. MAYORGA is the Goodnight Distinguished Chair and Director of the Operations Research Graduate Program at the North Carolina State Univesirtsity where she is also a professor of personalized medicine in the Edward P. Fitts Department of Industrial and Systems Engineering. She earned her Ph.D. in Industrial Engineering and Operations Research at the University of California, Berkeley. Her research interests include predictive models in health care and health care operations management and humanitarian systems. Her e-mail address is memayorg@ncsu.edu. Her website is http://mayorga.wordpress.ncsu.edu.

MEGHAN C. O'LEARY is a post-doctoral research associate at the University of North Carolina Lineberger Comprehensive Cancer Center. She earned her Ph.D. in Health Policy and Management, with a concentration in Decision Sciences and Outcomes Research, from the University of North Carolina Gillings School of Global Public Health. Her research interests include the use of systems science within implementation research, healthcare decision-making, and advancing equity in cancer prevention and control. Her email address is mcoleary@live.unc.edu.

KRISTEN HASSMILLER LICH is an associate professor in the Department of Health Policy and Management at the University of North Carolina Gillings School of Global Public Health. She earned her Ph.D. in Health Services Organization and Policy and M.H.A. in Public Health from the University of Michigan. Her research interests include applying systems thinking, operations research, and simulation modeling methods to complex health problems. Her email address is klich@unc.edu.