Micro-Simulation of Medicaid Coverage Loss on Colorectal and Breast Cancer Screening

Executive Summary

Medicaid coverage changes directly affect cancer screening rates and long-term treatment costs. Yet most policy analyses lack the granularity to trace these effects at the individual level or to account for the geographic and demographic specificity of a regional population. This white paper introduces an integrated, modular approach to synthetic population modeling that enables decision-makers to explore the health and economic consequences of different Medicaid policies with unprecedented detail and flexibility.

We describe a four-stage pipeline that: (1) generates a geographically and demographically realistic synthetic population for the Hampton Roads region using Census and Public Use Microdata Sample (PUMS) data; (2) assigns cancer screening status using Behavioral Risk Factor Surveillance System (BRFSS)–derived behavioral patterns; (3) estimates individual-level disease risk using established epidemiological models; and (4) calculates the long-term healthcare costs of missed screening due to policy-driven Medicaid disenrollment.

This report represents our first effort to operationalize this approach. We emphasize three core advantages. First, the model operates at the individual level, allowing us to tell concrete stories about specific types of people and compute outcome metrics that mirror the sociodemographic structure of Hampton Roads, Virginia. Second, the architecture is explicitly modular, so each stage can be updated independently as better data become available. Third, the framework is extensible: although we focus primarily on colorectal cancer (CRC), the same structure can be adapted to breast cancer and other screenable conditions with relatively modest modification of risk and cost parameters.

The work is grounded in the best available data currently accessible to our team—Census American Community Survey (ACS) and PUMS for demographics, BRFSS and CDC PLACES for screening behavior, the Colorectal Cancer Risk Assessment Tool (CCRAT) framework for risk estimation, and published health economics literature for treatment costs. At the same time, we have active data requests for more granular sources, including Virginia Medicaid enrollment and claims, regional electronic health record (EHR) data, and local cancer registry statistics. The model has been designed so that these improved data sources can be "dropped in" without restructuring the pipeline, allowing us to make substantive analytic progress now while preserving a clear path toward higher-fidelity analyses in future work.

---

1. Introduction and Problem Motivation

1.1 Policy Context and Motivation

Medicaid expansion under the Affordable Care Act has provided insurance coverage to millions of low-income adults across the United States, including many residents of the Hampton Roads region of Virginia. Expanded coverage has improved access to primary care and preventive services, including screening for colorectal and breast cancer. However, ongoing policy debates and administrative changes continue to create uncertainty around the future of Medicaid coverage. Proposals to tighten eligibility thresholds, increase administrative burdens for renewal, or restrict coverage for specific subpopulations (e.g., recent immigrants) all risk reducing enrollment and increasing the frequency and duration of coverage gaps.

The concern is not purely financial. Loss of Medicaid coverage can disrupt established relationships with primary care providers, interfere with continuity of care, and reduce the likelihood that individuals receive age-appropriate cancer screening. When screening is delayed or foregone altogether, cancers that might have been detected at an early, more treatable stage are instead discovered later, when symptoms are more severe and treatment more complex and expensive. The consequences of these dynamics fall disproportionately on low-income communities, communities of color, and individuals with limited access to employer-sponsored or marketplace insurance options.

This report focuses on the Hampton Roads region because it combines substantial demographic diversity with a relatively high dependence on Medicaid among working-age adults and near-elderly individuals. The central policy question we seek to address is:

> What is the impact of Medicaid coverage loss on colorectal and (by extension) breast cancer screening, and how much will this ultimately cost the Hampton Roads region— both in terms of health outcomes and the long-run financial burden on society?

Traditional analyses often answer such questions using aggregate data at the state or national level. While informative, those approaches cannot capture the specific ways in which coverage changes affect the people who actually live in Hampton Roads, nor can they easily express differential impacts across demographic subgroups within the region. Our goal is to provide a more granular, data-driven view that preserves local detail and supports more nuanced policy discussion.

1.2 Why a Synthetic Population Model?

Most existing policy evaluations rely on aggregate indicators: statewide screening rates, statewide incidence rates, or average treatment costs stratified by broad age groups. Such metrics are useful but conceptually distant from the individual decisions and experiences that drive real-world outcomes. For example, knowing that "68 percent of adults aged 50–75 are up to date on colorectal cancer screening" tells us little about which specific communities are underserved, which subgroups are most vulnerable to policy-induced coverage loss, or how changes in Medicaid eligibility thresholds will alter the lived experience of people in Hampton Roads.

A synthetic population model addresses these limitations by constructing a micro-level representation of the region's population. Each record in the synthetic dataset represents one simulated individual, with attributes that jointly approximate the real distribution of age, sex, race/ethnicity, income, education, and insurance coverage. When calibrated using high-quality data sources, synthetic populations can serve as realistic platforms for simulating policy interventions and tracking their consequences over time.

In this project, we build a synthetic population for Hampton Roads that allows us to:

1. **Tell person-centered stories.** Rather than discussing the impact of a policy on a faceless aggregate, we can ask how it affects, for example, a 52-year-old Black woman living in a specific tract, with a defined income range, education level, and insurance status. While each such individual is synthetic, the joint distribution of attributes is calibrated to match real-world patterns. This alignment supports more intuitive and compelling narratives for stakeholders and decision-makers.

2. **Compute outcome metrics at the individual level.** By modeling outcomes at the person level and then aggregating, we can easily compute metrics by subgroup (e.g., stratified by race/ethnicity, income bracket, or age band) and drill down to understand who bears the largest burden of coverage loss. This is especially important for equity-focused policy design.

3. **Explore "what-if" policy scenarios.** The same synthetic population can be subjected to different hypothetical policy regimes—more restrictive income thresholds, administrative churning, targeted restrictions—and the resulting changes in insurance status, screening behavior, and long-term costs can be recomputed consistently. This enables scenario analysis and sensitivity testing.

4. **Make effective progress despite data limitations.** Crucially, synthetic population modeling allows us to begin analyzing the consequences of policy changes even while we await access to more detailed administrative data. At present, we do not have person-level Medicaid claims or local EHR data for Hampton Roads. However, we do have high-quality Census data, BRFSS-based estimates of screening behavior, and published epidemiological and cost models. By combining these sources in a principled way, we obtain a first generation of results that can be refined as better data become available.

In short, the synthetic population framework provides a bridge between abstract policy proposals and the concrete experiences of individuals in Hampton Roads. It permits both rigorous quantitative analysis and intuitive qualitative storytelling, which together have the potential to inform more thoughtful and equitable decision-making.

2. Model Architecture and Approach

Our approach is organized into four sequential stages, each of which is data-driven and designed to be modular. In the current implementation, these stages are implemented in separate Python modules that interact via CSV-based inputs and outputs. This separation allows each stage to be updated independently as data sources or modeling assumptions evolve.

2.1 Stage 1: Synthetic Population Generation

2.1.1 Objective

The first stage aims to generate a synthetic population of individuals whose demographic and socioeconomic characteristics match those observed in the Hampton Roads region. Specifically, we seek to replicate the joint distribution of age, sex, race/ethnicity, income, education, and health insurance status across census tracts in the region. This ensures that any downstream analysis starts from a realistic representation of the local population.

2.1.2 Data Sources

We rely on two primary data sources from the U.S. Census Bureau:

- **American Community Survey (ACS) 5-year estimates.** ACS provides tract-level marginal distributions of demographic and socioeconomic variables, including age, sex, race/ethnicity, household income, educational attainment, and health insurance coverage. These marginals define the target distributions that our synthetic population must match.

- **ACS Public Use Microdata Sample (PUMS).** PUMS provides individual-level records with joint distributions of demographics and socioeconomic characteristics. Although PUMS is not available at the tract level, it can be restricted to Public Use Microdata Areas (PUMAs) that approximate the Hampton Roads region. We use PUMS to estimate relationships such as the distribution of income given race/ethnicity and education, or the probability of insurance coverage given age and race/ethnicity.

By combining tract-level marginals from ACS with conditional distributions estimated from PUMS, we can reconstruct a plausible joint distribution of attributes at the tract level.

2.1.3 Method: Iterative Proportional Fitting

We use Iterative Proportional Fitting (IPF), a longstanding technique in transportation planning and demography, to reconcile marginal constraints with joint distributions. In this context:

1. We start with a pool of prototype individuals derived from PUMS, each with a set of attributes (age, race/ethnicity, income bracket, education level, insurance status).

2. For each tract, we apply IPF-style reweighting so that the weighted totals of the synthetic individuals match the ACS tract-level marginals (for example, the total number of individuals in each age group, race/ethnicity category, and sex).

3. We then sample integer numbers of individuals from these weighted prototypes, ensuring that each tract's synthetic population meets the marginal constraints within a specified tolerance.

This procedure preserves the realistic cross-variable relationships observed in PUMS while ensuring consistency with ACS tract-level totals. The result is a tract-resolved synthetic population for Hampton Roads.

2.1.4 Output and Interpretation

The output of Stage 1 is a CSV file in which each row corresponds to a synthetic individual and each column corresponds to an attribute. Core fields include:

- Demographic identifiers: tract GEOID, tract name, age group, sex, race/ethnicity.

- Socioeconomic indicators: income bracket, education level, health insurance status.

- Derived indicators: tract-level median income, total population count, and other contextual variables used later for analysis.

It is important to emphasize that these individuals are synthetic: they are not real people, and none of their records can be linked back to a specific person. However, the aggregate

distributions of attributes are designed to closely mirror the true population structure of Hampton Roads.

2.2 Stage 2: Cancer Screening Status Assignment

2.2.1 Objective

The second stage assigns a cancer screening status to each individual in the synthetic population. For colorectal cancer, this status indicates whether the individual is up to date with recommended screening, given their age and risk profile. The goal is to reflect both overall screening prevalence and known disparities by age, race/ethnicity, and insurance status, in a way that is consistent with the best available regional data.

Although the current implementation focuses on colorectal cancer screening, the same logic can be adapted to breast cancer mammography by adjusting eligibility age ranges and data sources.

2.2.2 Data Sources

We use two complementary sources:

- **CDC PLACES (Population-Level Analysis and Community Estimates) or BRFSS.** These resources provide tract- or county-level estimates of the proportion of adults who are current on colorectal cancer screening. For Hampton Roads, these estimates are used to define baseline screening rates at the geographic level.

- **BRFSS microdata (Virginia and regional subsets).** Where available, we use BRFSS microdata to estimate relative screening propensities by age group, race/ethnicity, sex, and insurance status. For example, we can estimate the odds ratio of screening for insured versus uninsured individuals, or for older versus younger adults.

Together, these sources allow us to capture both the overall screening level in each tract and the relative differences within demographic groups.

2.2.3 Method: Multilevel Probability Model

For each individual, we compute a screening probability using a multilevel probability model:

1. **Baseline tract-level rate.** We start with the tract-level baseline screening prevalence p_{tract} derived from PLACES or BRFSS.

2. **Adjustment factors.** We then apply multiplicative adjustment factors based on the individual's age group, race/ethnicity, and insurance status. These factors are estimated from BRFSS via logistic regression or stratified frequency analysis. Conceptually, we define:

$$p_{\text{indiv}} = \text{clip}\left(p_{\text{tract}} \times A_{\text{age}} \times A_{\text{race}} \times A_{\text{ins}}, 0.01, 0.99\right)$$

where A_{age}, A_{race}, and A_{ins} are multiplicative adjustments and the clip function enforces logical bounds on the probability.

3. **Random assignment.** Given the individual-level probability p_{indiv}, we draw a Bernoulli random variable to determine whether the individual is screened (success) or not screened (failure).

This modeling framework ensures that tract-level screening prevalence is preserved on average, while also incorporating subgroup differences that reflect known disparities in access and uptake.

2.2.4 Implementation and Output

In software, this logic is encapsulated in a dedicated `ScreeningCalculator` component. This module can be called both after Stage 1 to assign initial screening status and again in Stage 4 after coverage changes, ensuring consistent behavior across the pipeline.

The main outputs of Stage 2 include:

- A binary indicator for screening status (screened versus not screened).

- Continuous probabilities for screening at the individual level.

- An eligibility flag marking individuals who are in the recommended age range for screening.

These outputs form a bridge between the demographic structure constructed in Stage 1 and the disease risk and cost modeling in subsequent stages.

2.3 Stage 3: Cancer Risk Estimation

2.3.1 Objective

Stage 3 estimates, for each individual, the probability of developing colorectal cancer over a specified time horizon (e.g., five years), under two scenarios: no screening and screening as assigned in Stage 2. These risk estimates serve as inputs to the economic model in Stage 4 and also provide a clinically meaningful interpretation of the benefits of screening.

2.3.2 Data Sources

The risk estimation is anchored in the Colorectal Cancer Risk Assessment Tool (CCRAT) framework and related epidemiological literature. Parameterization draws on:

- Baseline age-specific incidence rates of colorectal cancer.

- Multiplicative risk factors associated with sex, race/ethnicity, income, and educational attainment.

- Screening effectiveness estimates derived from clinical studies of colonoscopy and other screening modalities.

These inputs are encoded in a configuration file of risk parameters, which can be updated as new evidence emerges.

2.3.3 Method: Multiplicative Risk Model

We employ a multiplicative risk model in which an individual's unscreened risk is represented as:

$$\text{UnscreenedRisk} = R_{\text{base}}(\text{age}) \times M_{\text{gender}} \times M_{\text{race}} \times M_{\text{income}} \times M_{\text{education}}$$

where:

- $R_{\text{base}}(\text{age})$ is the baseline risk associated with the individual's age group.

- M_{gender}, M_{race}, M_{income}, and $M_{\text{education}}$ are multiplicative factors representing the relative increase or decrease in risk attributable to each characteristic.

Screening is modeled as reducing this baseline risk by a factor corresponding to the effectiveness of screening in preventing or detecting early-stage disease:

$$\text{ScreenedRisk} = \text{UnscreenedRisk} \times (1 - E_{\text{screen}})$$

where E_{screen} is the effectiveness parameter.

The absolute risk reduction from screening is:

$$\text{ScreeningBenefit} = \text{UnscreenedRisk} - \text{ScreenedRisk}.$$

The model may also classify individuals into qualitative risk categories (e.g., low, medium, high) based on threshold values applied to these probabilities.

2.3.4 Output

For each individual, Stage 3 generates:

- An estimate of the five-year risk of colorectal cancer without screening.

- An estimate of the five-year risk with screening.

- The associated absolute risk reduction.

- A categorical risk label if desired.

These outputs are later combined with cost and stage distribution parameters in Stage 4 to quantify expected treatment costs and the financial value of screening.

2.4 Stage 4: Medicaid Policy Simulation and Economic Impact

2.4.1 Objective

The final stage simulates the downstream consequences of Medicaid policy changes on insurance coverage, screening behavior, and long-run cancer treatment costs. Decisions about Medicaid eligibility criteria and administrative processes are modeled as interventions that alter individuals' coverage status. These changes propagate through Stage 2 and Stage 3, ultimately affecting disease risk and cost outcomes.

2.4.2 Policy Specification and Medicaid Inference

Policies are defined externally in CSV files that specify:

- The target population (e.g., individuals currently inferred to be on Medicaid).

- The conditions that determine which subset of that population is affected (e.g., income between 100 and 138 percent of the federal poverty level, age between 18 and 55, or membership in a racial/ethnic group used as a proxy for immigrant status).

- The resulting coverage change (e.g., loss of Medicaid leading to an uninsured status).

Because we do not yet have direct access to person-level Medicaid enrollment data, we infer likely Medicaid status based on income, insurance type, and other socioeconomic indicators. For example, low-income individuals with public insurance are classified as probable Medicaid beneficiaries, while those with higher incomes are more likely to have private coverage. These inferences will be replaced with actual enrollment information as claims data become available.

2.4.3 Screening Recalculation After Policy Changes

When a policy scenario is applied, individuals who meet the policy's targeting criteria may lose Medicaid coverage. Their insurance status is updated to reflect this change (e.g., from insured to uninsured). The `ScreeningCalculator` is then invoked again to recompute screening probabilities given the new coverage configuration.

Individuals who lose coverage typically experience a reduction in screening probability, and a subset of them will transition from screened to unscreened status in the simulated population. This dynamic explicitly captures the link between Medicaid coverage and access to preventive cancer screening.

2.4.4 Economic Impact Model

To assess the financial implications of coverage and screening changes, we estimate expected treatment costs for each individual under both baseline and policy scenarios. The core components of the economic model are:

- **Stage-specific treatment costs.** Published literature provides estimated costs for treating colorectal cancer at different stages (I–IV). Early-stage cancers are less expensive to treat than late-stage cancers, which often require intensive chemotherapy, radiation, and extensive surgery.

- **Stage distribution by screening status.** Screening is associated with a shift toward earlier-stage diagnosis. The model uses different stage distributions for screened and unscreened populations, based on epidemiological studies and national cancer statistics.

- **Screening costs.** Screening itself incurs costs, such as the cost of colonoscopy and associated clinical services. These are included in the baseline scenario when individuals are screened.

- **Discounting and time horizon.** Expected costs are typically computed over a 10-year horizon and discounted at a standard rate to reflect the time value of money.

For each individual, the model computes:

- The expected cost of care under baseline coverage and screening (including screening costs and treatment costs weighted by stage probabilities).

- The expected cost of care under the policy scenario, in which some individuals are no longer screened and therefore face different stage distributions and higher expected treatment costs.

- The difference between these two expected costs, which represents the incremental financial burden associated with the policy for that individual.

By aggregating these differences across the population, we obtain an estimate of the total incremental treatment cost attributable to the policy, as well as per-capita figures for affected individuals and specific subgroups.

2.4.5 Interpretation for Decision-Makers

The outputs of Stage 4 can be summarized in ways that are meaningful to policymakers and stakeholders. For each policy scenario, we can report:

- How many individuals lose Medicaid coverage.

- How many of those individuals also lose access to colorectal cancer screening.

- The total additional cancer treatment costs expected over the modeled time horizon.

- The apparent savings from reduced screening expenditures.

- The net financial impact, highlighting whether coverage restrictions are ultimately cost-increasing once treatment costs are considered.

Because all of these estimates are derived from an individual-level model, they can readily be broken down by age group, race/ethnicity, income bracket, and other attributes. This allows decision-makers to understand not only the aggregate financial impact, but also which communities and demographic groups are most affected.

---

3. Modularity and Extensibility of the Approach

3.1 Rationale for a Modular Design

The model is intentionally decomposed into four stages rather than implemented as a single monolithic system. This modular design reflects both practical and scientific considerations.

From a practical standpoint, different data sources govern different parts of the pipeline. Census data are central to population generation; BRFSS and PLACES are critical for screening behavior; CCRAT and related literature inform risk estimation; and health economics studies underpin the cost model. These data sources evolve at different rates and may become available at different times. A modular architecture allows each stage to be updated independently as new data or methodological advances arise.

From a scientific perspective, modularity promotes transparency. Each stage has a clearly defined purpose, input, and output. Assumptions are localized, making it easier to evaluate and critique individual components. For example, if new evidence suggests revising the stage distribution of cancers among screened individuals, we can adjust the parameters in the economic model without altering the synthetic population or the screening assignment logic.

3.2 Making Progress with Current Data While Planning for Better Data

At present, we rely on publicly available data and published literature. These sources are, in many respects, the best information currently accessible without special data use agreements. Nevertheless, we recognize that they are imperfect, especially when it comes to local detail. For instance, national or state-level averages may not fully capture the unique demographic and healthcare environment of Hampton Roads.

Rather than delay analysis until ideal data are available, we have chosen to proceed with this first effort while simultaneously pursuing more specific sources. Data requests are in progress for:

- Virginia Medicaid enrollment and claims, which will refine our understanding of who is on Medicaid and how coverage changes affect utilization.

- Regional EHR and cancer registry data, which will improve estimates of local screening rates and stage distributions.

- Additional BRFSS microdata extracts, which will strengthen our ability to characterize screening disparities by race/ethnicity and insurance status at the regional level.

The modular design ensures that each of these new data streams can be integrated with minimal disruption. For example, once claims data are accessible, the inferred Medicaid status in Stage 4 can be replaced by observed enrollment. Similarly, local screening prevalence from EHR or registry sources can supplement or replace BRFSS-based estimates in Stage 2. This approach balances the urgency of informing current policy debates with the long-term goal of building a higher-fidelity model.

3.3 Extending the Framework to Breast Cancer

Although this report focuses on colorectal cancer, the same conceptual framework can be adapted to breast cancer with relatively modest changes.

In Stage 1, the synthetic population remains unchanged; the same demographic and socioeconomic attributes are relevant for breast cancer as for colorectal cancer. In Stage 2, the eligibility age range and screening intervals would be updated to reflect mammography guidelines, and breast cancer–specific screening prevalence and disparities would be modeled using BRFSS and other relevant sources. In Stage 3, the CCRAT-based risk model would be replaced or supplemented with a breast cancer risk assessment tool (such as BCRAT), with parameters calibrated to reflect breast cancer incidence patterns. In Stage 4, colorectal cancer treatment cost and stage distribution parameters would be replaced with breast cancer–specific values drawn from the health economics literature and local data where available.

Because all of these changes can be implemented via configuration files and modular components, the extension to breast cancer does not require redesigning the pipeline. This extensibility is particularly important for policy questions that involve trade-offs across multiple types of cancer screening or consider comprehensive preventive care strategies.

4. Key Design Decisions and Transparency

4.1 Individual-Level Modeling and Storytelling

A central design decision in this work is the choice to model outcomes at the individual level rather than relying solely on aggregate categories. This choice has several advantages.

First, it supports more intuitive storytelling. Policy debates can easily become abstract, focusing on percentages and total costs without conveying the human experience behind the numbers. By modeling synthetic individuals, we can describe how a specific combination of age, race/ethnicity, income, and insurance status translates into tangible changes in screening opportunities, cancer risk, and treatment burden. These narratives can be powerful tools for communication with both policymakers and the public.

Second, individual-level modeling provides flexibility in stratification. Because all attributes are defined at the person level, we can compute outcome metrics for any subgroup of interest without re-running the model. For example, we can ask how a given policy affects low-income Black women aged 50–64, or how it changes costs for uninsured men living in specific tracts. This is particularly important for evaluating equity and identifying disproportionate impacts.

Third, person-level modeling captures heterogeneity. Not all individuals in a given category share the same risk or respond identically to coverage changes. By simulating individual outcomes with probability models, we can represent variation within groups more convincingly than is possible with purely aggregate approaches.

4.2 Data Transparency and Provenance

Another key design principle is explicit documentation of data sources and assumptions. Each stage of the model is backed by clearly identified datasets and references, and parameters are stored in configuration files that can be reviewed and updated. The intent is to make it straightforward to trace how a particular numerical result was obtained and what evidence supports it.

This transparency is important for scientific credibility and for building trust with stakeholders. Decision-makers should be able to understand where the numbers come from, what uncertainty surrounds them, and how the model might change if new evidence emerges. To that end, the model is constructed in a way that facilitates parameter sensitivity analysis, scenario comparisons, and methodological scrutiny.

4.3 Acknowledging Uncertainty and Limitations

Because this is our first effort to develop a synthetic population–based policy model for Hampton Roads, it necessarily has limitations. Some of the most important include:

- The absence of person-level Medicaid claims and enrollment data, requiring us to infer coverage status from socioeconomic indicators.

- Reliance on national or state-level estimates for some parameters (such as stage-specific treatment costs and stage distributions) that may not perfectly reflect local practice patterns or costs.

- Use of relatively simple multiplicative models for risk and behavior, which may not capture all relevant interactions and contextual factors.

Recognizing these limitations is not a weakness but a prerequisite for responsible model use. We view the current model as a starting point that provides valuable insights despite imperfections. As better data and methods become available, we will refine and extend the analysis.

4.4 Equity and Policy Implications

The structure of the model is intentionally designed to highlight disparities. Because outcomes are computed at the person level and linked to demographic attributes, we can directly observe how policies affect different communities. This is critical when evaluating proposals that may disproportionately impact marginalized groups, such as immigrant-focused coverage restrictions or work requirements that predominantly affect low-wage workers.

By making these patterns explicit, the model can inform more equitable policy design. For example, if a given policy yields an overall cost savings but imposes significant additional cancer burden on specific racial or ethnic groups, that trade-off can be quantified and brought into the public discussion.

---

5. Application to Decision-Making

5.1 Intended Users and Use Cases

The modeling framework is intended to support a range of users, including:

- **State and local policymakers**, who need to understand the health and economic consequences of proposed Medicaid changes before implementing them.

- **Health systems and payers**, who must plan for the downstream impact of coverage shifts on service demand, treatment costs, and resource allocation.

- **Community and advocacy organizations**, which seek to quantify the burdens imposed by policy changes and advocate for more equitable alternatives.

- **Researchers and analysts**, who may use the model as a foundation for deeper investigations or as a benchmark for comparing alternative modeling approaches.

Typical use cases include evaluating the impact of specific proposals (e.g., lowering eligibility thresholds), exploring combinations of policies (e.g., administrative churning plus immigrant restrictions), and testing the sensitivity of results to key assumptions.

5.2 Examples of Policy Questions the Model Can Address

Because the model is scenario-based and individual-level, it can address a variety of questions, such as:

- How many individuals in Hampton Roads would lose Medicaid coverage if income eligibility were tightened from 138 percent to 100 percent of the federal poverty level, and how would that change colorectal cancer screening rates?

- Which demographic groups—by age, race/ethnicity, income, and geography—would be most affected by increased administrative churn due to more frequent eligibility redeterminations?

- If a policy were implemented that restricted Medicaid coverage for certain immigrant populations, how would this affect screening, cancer stage at diagnosis, and treatment costs for those communities?

- What is the long-term cost of not screening an average uninsured individual in a particular risk group, compared with the cost of maintaining screening coverage?

- How do policies that appear to save money by reducing screening costs compare, over a 10-year horizon, to the increased treatment costs associated with more advanced cancers?

By providing quantitative answers to these types of questions, the model can help move policy debates beyond anecdote and intuition toward evidence-informed decision-making.

5.3 Limitations and Responsible Use

Even as the model adds clarity, it is not a crystal ball. Users should bear in mind:

- The model simulates probabilities and expected values; actual outcomes will vary in ways that cannot be predicted exactly.

- Behavioral responses to policy changes can be complex and are only partially captured by changes in insurance status and screening probability.

- The cost and risk parameters are subject to revision as new evidence emerges.

Therefore, results should be interpreted as informed projections, not precise predictions. The most appropriate use of the model is to compare relative outcomes across scenarios, explore ranges of possible impacts, and identify patterns and trade-offs that warrant closer scrutiny.

---

6. First-Effort Positioning and Future Directions

This project represents our first coordinated effort to build an integrated, synthetic population–based model for assessing the impact of Medicaid coverage changes on cancer screening and treatment costs in Hampton Roads. Despite the limitations discussed above, the work already demonstrates several important strengths:

- It provides a coherent, end-to-end pipeline that links demographic structure to screening behavior, risk, and economic outcomes.

- It operates at the individual level, supporting both granular analysis and human-centered storytelling.

- It is explicitly modular and designed to incorporate improved data sources as they become available.

- It is extensible to other cancers, beginning with breast cancer, which is a natural next step given the importance of mammography in preventive care guidelines.

We emphasize that this is a first effort, not a final product. Our aim in presenting this work now is to:

- Invite feedback from clinicians, health economists, data scientists, policymakers, and community stakeholders.

- Identify opportunities for methodological refinement and data enhancement.

- Provide a concrete tool that can be used immediately to explore the implications of proposed Medicaid policies, even as we continue to push for better data access.

Ultimately, the goal is not simply to produce a single model but to establish an enduring analytic framework for understanding how coverage decisions affect the health and economic well-being of communities like Hampton Roads. Medicaid coverage is tightly linked to cancer screening; screening is tightly linked to early diagnosis and improved outcomes; and policy-driven coverage loss threatens to undermine these connections. By quantifying these relationships with a transparent, data-informed model, we hope to support decisions that better align with the long-term interests of both individuals and society.

References

Beckman, R. J., Baggerly, K. A., & McKay, M. D. (1996). Creating synthetic baseline populations. *Transportation Research Part A: Policy and Practice*, 30(6), 415–429.

Brown, M. L., Riley, G. F., Schussler, N., & Etzioni, R. (2002). Estimating health care costs related to cancer treatment from SEER-Medicare data. *Medical Care*, 40(8 Suppl), IV-104–IV-117.

Deming, W. E., & Stephan, F. F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Annals of Mathematical Statistics*, 11(4), 427–444.

Dominitz, J. A., Robertson, D. J., Ahnen, D. J., et al. (2021). Colonoscopy vs. fecal immunochemical test in reducing mortality from colorectal cancer (CONFIRM): A prospective randomized trial. *American Journal of Gastroenterology*, 116(1), 93–102.

Freedman, A. N., Slattery, M. L., Ballard-Barbash, R., et al. (2009). Colorectal cancer risk prediction tool for white men and women without known susceptibility. *Journal of Clinical Oncology*, 27(5), 686–693.

Mariotto, A. B., Yabroff, K. R., Shao, Y., Feuer, E. J., & Brown, M. L. (2011). Projections of the cost of cancer care in the United States: 2010–2020. *Journal of the National Cancer Institute*, 103(2), 117–128.

Siegel, R. L., Miller, K. D., & Jemal, A. (2023). Cancer statistics, 2023. *CA: A Cancer Journal for Clinicians*, 73(1), 17–48.

Tangka, F. K. A., Trogdon, J. G., Richardson, L. C., Xia, L. Z., & Sabatino, S. A. (2010). Cancer treatment cost in the United States: A systematic review of studies published in peer-reviewed journals. *Journal of Oncology Practice*, 6(6), 313–321.

Thorpe, K. E., Florence, C. S., & Joski, P. (2004). Which medical conditions account for the rise in health spending? *Health Affairs*, W4-437–W4-445.

Trogdon, J. G., Finkelstein, E. A., Tangka, F. K., Orenstein, D., & Richardson, L. C. (2008). State-level estimates of cancer screening, treatment, and management costs. *Cancer Epidemiology, Biomarkers & Prevention*, 17(9), 2540–2545.

Trogdon, J. G., Tangka, F. K., Finkelstein, E. A., & Stein, K. D. (2012). State-level estimates of cancer screening, treatment, and management costs: Health and economic implications. *Journal of the National Cancer Institute Monographs*, 2012(44), 38–45.

Winawer, S. J., Zauber, A. G., Ho, M. N., et al. (1993). Prevention of colorectal cancer by colonoscopic polypectomy. *New England Journal of Medicine*, 329(27), 1977–1981.

Zhang, X., Holt, J. B., Lu, H., et al. (2015). Multilevel regression and poststratification for small-area estimation of population health outcomes. *American Journal of Epidemiology*, 181(11), 899–907.