Columbia
Version: 1.0.00
Released: 2025-09-30

Columbia Gastric Cancer Simulation Model (GSiMo): Model Profile

Columbia University Irving Medical Center

Contact

Kevin Rouse (kr3164@cumc.columbia.edu)
Chin Hur (ch447@cumc.columbia.edu)

Funding

The development of this model was supported by the NIH/NCI CISNET Gastric Cancer Grant (U01CA265729).

Suggested Citation

Rouse K, Hur C. Columbia Gastric Cancer Simulation Model (GSiMo): Model Profile. [Internet] Sep 30, 2025. Cancer Intervention and Surveillance Modeling Network (CISNET). Available from: https://cisnet.cancer.gov/resources/files/mpd/gastric/CISNET-gastric-gsimo-columbia-model-profile-1.0.00-2025-09-30.pdf

Version Table

Version	Date	Notes
1.0.00	2025-09-30	Initial release

COLUMBIA UNIVERSITY

Reader's Guide

Core Profile Documentation

These topics will provide an overview of the model without the burden of detail. Each contains links to more detailed information if required.

Model Purpose
This document describes the primary purpose of the model.

Model Overview
This document describes the primary aims and general purposes of this modeling effort.

Assumption Overview
An overview of the basic assumptions inherent in this model.

Parameter Overview
Describes the basic parameter set used to inform the model, more detailed information is available for each specific parameter.

Component Overview
A description of the basic computational building blocks (components) of the model.

Output Overview
Definitions and methodologies for the basic model outputs.

Results Overview
A guide to the results obtained from the model.

KeyReferences
A list of references used in the development of the model.

Model Purpose

Summary

The **G**astric Cancer **Si**mulation **M**odel (GSiMo) is a state-transition microsimulation model that simulates the natural history of gastric cancer (GC) in the U.S. population. GSiMo models GC onset, progression, detection, and mortality, incorporating risk factors such as *H. pylori* (HP) infection status and demographic variations in risk by race and ethnicity. Developed to inform prevention and treatment strategies, GSiMo aims to identify optimal approaches to reduce GC incidence and mortality.

Purpose

Gastric cancer ranks as the fifth most common cancer globally and is the fifth leading cause of cancer mortality as of 2020 [1]. Certain racial and ethnic groups groups face a significantly higher risk of GC mortality and mortality than White populations[2] largely due to differences in HP infection rates, smoking prevalence, and access to preventive care. As part of the CISNET comparative modeling effort, this model seeks to inform public health policies to reduce incidence and mortality of the disease.

1. **Estimate gastric cancer outcomes for subgroups** by race and ethnicity in the U.S., focusing on subgroup-specific risk factors and competing mortality profiles.
2. **Assess the impact of risk factors and prevention strategies** on differences between subgroups, including the effects of *H. pylori* transmission dynamics and the cost-effectiveness of screen-and-treat interventions.
3. **Evaluate targeted secondary prevention strategies** to reduce early-onset gastric cancer incidence and mortality, including optimal screening and surveillance regimens for high-risk populations.
4. **Adapt the models for global application** to estimate the potential impact of prevention strategies on gastric cancer outcomes in various countries.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021;71(3):209–249.
2. American Association for Cancer Research. Cancer Disparities Progress Report 2024. 2024.

Model Overview

Model Overview

Summary

This document provides an overview of the structure of the Gastric Cancer Simulation Model (GSiMo).

Purpose

The GSiMo model is designed to assess the impact and cost-effectiveness of interventions for gastric cancer. See Model Purpose for more details.

Background

Gastric cancer (GC) ranks as the fifth leading cause of cancer death worldwide, with nearly 1 million new diagnoses each year [1]. Although age-standardized rates of GC have decreased since 1990, the absolute number of cases continues to rise [2]. In the US, the incidence and mortality of the disease vary by racial group[3]. H. pylori infection is a major risk factor, responsible for at least 80% of all gastric cancer cases. As differences in GC risk are largely attributable to differences in the prevalence of H. pylori infection and other risk factors [4], primary prevention strategies may be particularly effective at reducing GC burden among high-risk populations.

Gastric cancer, being a complex and multifactorial disease with a well-characterized precancerous process known as the Correa cascade [5], warrants both primary and secondary prevention efforts. Primary prevention efforts include screening and treatment of H. pylori, which has been shown to reduce GC incidence regardless of baseline GC risk [6]. Additionally, secondary prevention strategies (i.e. endoscopic screening for gastric intestinal metaplasia) targeted at reducing early-onset GC incidence and mortality is critical, as survival rates are low, with only 36% surviving at least five years post-diagnosis [7]. While recently proposed American Gastroenterological Association (AGA) guidelines recommend against routine surveillance in patients with gastric intestinal metaplasia (IM), they advocate for consideration of surveillance in high-risk groups (incomplete or extensive metaplasia, family history, racial/ethnic ancestry, country of origin) [8]. Despite these guidelines, evidence demonstrating the clinical benefits of specific screening modalities is limited, highlighting the need for decision modeling to address knowledge gaps.

Model Description

GSiMo is a state-transition microsimulation model that simulates the natural history of gastric cancer in the U.S. population. The model generates a population of individuals with varying risk of developing gastric cancer based on *H. pylori* (HP) infection status and demographic characteristics such as race and sex. Individuals then progress through health states with transition rates dependent on HP status, race, sex, and age. GSiMo was developed in Python (v3.11.8).

GSiMo simulates a population of individuals starting from age 18 to 100 for each demographic subgroup: Non-Hispanic (NH) Black females, NH Black males, NH White females, and NH White males. A proportion of the population is initialized as HP-positive and the remainder is initialized as healthy, aligning with estimated HP infection prevalence. Each month, patients transition to one of the following non-overlapping health states: healthy, HP infection, atrophic gastritis, intestinal metaplasia, dysplasia, undetected gastric cancer (Stages I-IV), detected gastric cancer (Stages I-IV), cancer death, and other death (Figure 1). Transition probabilities differ for each demographic subgroup and depend on HP infection status as well as age. See Assumption Overview for more details on model structure and parameter assumptions.

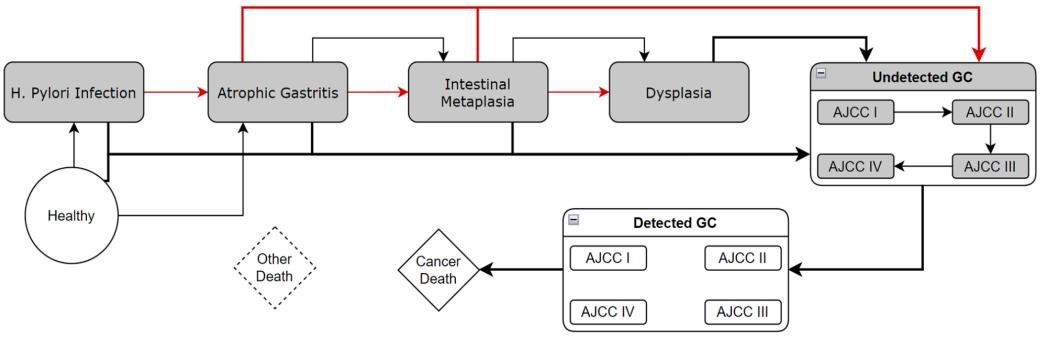

Figure 1. GSiMo model schematic. Red arrows indicate HP infection dependence in the transition probabilities. All states are connected to Other Death.

GSiMo's natural history module consists of two parts: a population-level Markov model to efficiently calibrate parameters, and an individual-level microsimulation model to capture greater clinical realism beyond the scope of the Markov model. See Component Overview for more details.

GSiMo derives fixed parameters from common model input generators including the HP Infection generator and Life Table generator, as well as survival data from the Surveillance Epidemiology and End Results (SEER) database. For unobserved "dark" states including progression through the Correa Cascade, progression through preclinical cancer, and detection of cancer, parameters are calibrated via a constrained simulated annealing process to SEER GC incidence data and precursor prevalence targets from literature. See Parameter Overview for more details on model inputs and parameters.

Primary outputs from GSiMo include GC incidence and mortality. Secondary outputs include prevalence of precursor lesions, dwell time, and progression rates. Comparative model validation utilizes the Maximum Clinical Likelihood Incidence Reduction (MCLIR) framework established by the CISNET Colorectal Group to highlight important differences between CISNET models. Once the screening and intervention component is fully implemented, outputs such as cancer cases and deaths averted, life years gained, quality-adjusted life years (QALYs) gained, and total costs will be added. See Output Overview and Results Overview for more details.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021;71(3):209–249.
2. Lyon, France: International Agency for Research on Cancer (IARC Working Group Reports, No. 8). Helicobacter pylori Eradication as a Strategy for Preventing Gastric Cancer. 2014;
3. American Association for Cancer Research. Cancer Disparities Progress Report 2024. 2024.
4. Ford A. C., Forman D., Hunt R. H., Yuan Y., Moayyedi P. Helicobacter pylori eradication therapy to prevent gastric cancer in healthy asymptomatic infected individuals: systematic review and meta-analysis of randomised controlled trials. BMJ. 2014;348:g3174.
5. Correa P., Piazuelo M. B. The gastric precancerous cascade. J Dig Dis. 2012;13(1):2–9.
6. Lee YC, Chiang TH, Chou CK, Tu YK, Liao WC, Wu MS, Graham DY. Association Between Helicobacter pylori Eradication and Gastric Cancer Incidence: A Systematic Review and Meta-analysis. Gastroenterology. 2016;150(5):1113–1124.
7. National Cancer Institute. Cancer Stat Facts. Surveillance, Epidemiology, and End Results Program. 2018. https://seer.cancer.gov/statfacts/. Accessed October 27, 2024. 2018;
8. Gupta S, Li D, El Serag HB, Davitkov P, Altayar O, Sultan S, Falck-Ytter Y, Mustafa RA. AGA Clinical Practice Guidelines on Management of Gastric Intestinal Metaplasia. Gastroenterology. 2020;158(3):693–702.

Assumption Overview

Summary

An overview of the basic assumptions inherent in this model.

Background

Although there is extensive data for certain measures such as gastric cancer (GC) incidence and survival, data on precursor prevalence, GC subtype prevalence, preclinical cancer progression rates, etc. — particularly regarding variations between demographic subgroups — remains relatively sparse. Thus, any model of GC will involve significant assumptions about the natural history of the disease. In developing GSiMo, assumptions were chosen to keep the model as simple as possible while maximizing the utility of the existing data.

Assumption Listing

- No regression
    - Risk factors
        - HP acts as a risk factor for precursor states up to undetected cancer
        - Smoking is not included as a risk factor
    - Precursors
        - Only atrophic gastritis, intestinal metaplasia, and dysplasia states
        - No distinction between low-grade and high-grade dysplasia
    - Preclinical cancer
        - Progression through stages only occurs in the undetected cancer states, once detected, patient stops progressing
    - GC
        - Only non-cardia cases included
        - No survival distinction between intestinal and diffuse cases
        - A small fraction of cases transition directly to cancer to account for the possibility that some diffuse cases may not be progressing through the Correa cascade
        - Patients that have survived cancer for more than 10 years are considered cancer-free and transitioned back to the Healthy or HP state, depending on their HP status.

Calibration Constraints

The calibration process utilized a bounded simulated annealing process. The purpose of this constrained stochastic calibration process was to limit degrees of freedom, improve identifiability and validity of screening and intervention simulations. A detailed list of calibrated parameters and corresponding constraints can be found in Parameter Overview.

- “Accelerating” Correa's cascade, transitions at later precursor states are faster than earlier precursor states
- Detection rates at higher AJCC stages are faster than at lower stages
- All transition probabilities differ by sex and age

    - HP status-dependent transitions differ by race as well
- Progression through preclinical cancer stages is bounded by sojourn time estimates from literature

Parameter Overview

Summary

Describes the basic parameter set used to inform the model, more detailed information is available for each specific parameter.

Background

GSiMo uses both fixed and calibrated parameters. Fixed parameters were derived from sources such as SEER, literature, and common model input generators (elaborated on below). Calibrated parameters were arrived at via a constrained parameter search, with constraints informed by literature and clinician input.

Parameter Listing Overview

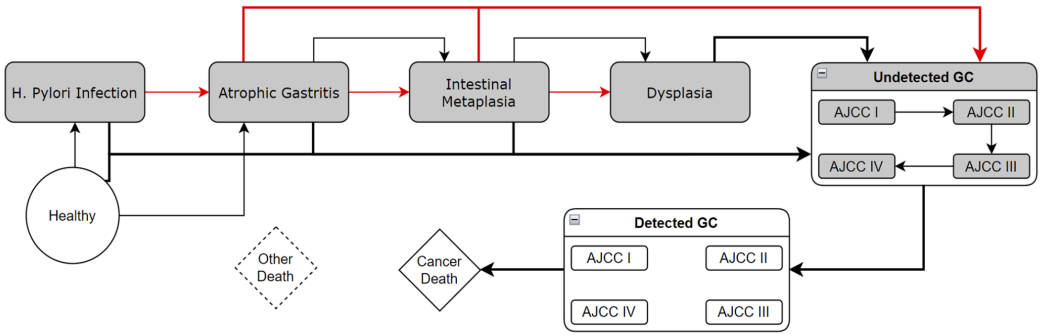

Figure 1. GSiMo model schematic. Red arrows indicate HP infection dependence in the transition probabilities. All states are connected to Other Death.

Natural History

To facilitate model comparison, common model input "generators" have been developed to ensure that all Gastric models are operating from the same base set of risk factor and competing mortality profiles. These input generators include the *H. pylori* (HP) Infection generator and the Life Table generator, whose outputs were incorporated into GSiMo as fixed parameters.

HP Infection Generator

Developed using National Health and Nutrition Examination Survey (NHANES) data and other sources, the HP generator models the age- and period-specific force of infection (FOI) -- the rate at which individuals acquire HP infection as a function of age -- by race/ethnicity subgroup. The FOI estimates outputted from the HP generator were used to derive the fixed, race-specific non-HP to HP transition probabilities. Additionally, HP generator data was used to derive starting HP infection prevalence in simulated subgroups.

Life Table Generator

The life table generator integrates mortality data from a wide variety of sources such as the National Center for Health Statistics (NCHS), CDC Wonder, Berkeley Mortality Database, US Social Security Administration and the American Cancer Society Cancer Prevention Study II, to create age-, sex-, period-, and race/ethnicity-specific all-cause mortality rates by smoking status. It addresses the lack of all-cause mortality data that is stratified by demographic subgroup as well as smoking status. While GSiMo does not currently incorporate smoking as a risk factor, mortality rates for non-smokers were used to derive all-cause mortality transition probabilities.

Survival Hazards

In order to model the competing risks of cancer mortality and mortality from other causes, GSiMo utilizes hazard functions generated from SEER data. Case listings data filtered for intestinal and diffuse, non-cardia gastric cancer cases with AJCC staging were pulled from the SEER 9 database [1]. These data include age at diagnosis, follow-up time, and cause of death, among other fields. Using the *rstpm2* package in R, flexible parametric survival models were fit to the case listings data to extract age- and duration-dependent hazard functions for cancer death and other death up to 10 years post-diagnosis. This was done for each race and sex subgroup. In the microsimulation, these hazard probabilities are used in place of probabilities derived from age-bucketed SEER survival rates data and life table generator data.

Table 1. Model Parameters

Parameter		Source	Varies by	Constraints
HP Infection	HP Prevalence at age 18	HP Generator	Race, Sex	N/A
	Healthy to HP	HP Generator	Race, Age	N/A
	AG to AG (HP)	HP Generator	Race, Age	N/A
	IM to IM (HP)	HP Generator	Race, Age	N/A
	Dys to Dys (HP)	HP Generator	Race, Age	N/A
	Healthy to AG	Calibrated	Sex, Age	Transition rates for women greater than those for men
Correa's Cascade	HP to AG	Calibrated	Race, Sex, Age	Less than double the Healthy to AG transition rates
	AG to IM	Calibrated	Sex, Age	Greater than Healthy to AG transition rates; Transition rates for women greater than those for men
	AG (HP) to IM (HP)	Calibrated	Race, Sex, Age	Less than double the AG to IM transition rates
	IM to Dys	Calibrated	Sex, Age	Greater than AG to IM transition rates; Transition rates for women greater than those for men
	IM (HP) to Dys (HP)	Calibrated	Race, Sex, Age	Less than double the IM to Dys transition rates
	Dys to Undetected GC I	Calibrated	Sex, Age	Greater than IM to Dys transition rates; Transition rates for women greater than those for men
	Dys (HP) to Undetected GC I	Calibrated	Race, Sex, Age	Less than double the Dys to Undetected GC I transition rates
Diffuse	Healthy to Undetected GC I	Calibrated	Race, Sex, Age	None
	HP to Undetected GC I	Calibrated	Race, Sex, Age	Equal to Healthy to Undetected GC I
	AG to Undetected GC I	Calibrated	Race, Sex, Age	Equal to Healthy to Undetected GC I
	AG (HP) to Undetected GC I	Calibrated	Race, Sex, Age	Equal to Healthy to Undetected GC I

Parameter		Source	Varies by	Constraints
	IM to Undetected GC I	Calibrated	Race, Sex, Age	Equal to Healthy to Undetected GC I
	IM (HP) to Undetected GC I	Calibrated	Race, Sex, Age	Equal to Healthy to Undetected GC I
Progression	Undetected GC I to Undetected GC II	Calibrated	Race, Sex, Age	Bound by progression rate values derived from sojourn times from literature [2]; Transition rates for women greater than those for men
	Undetected GC II to Undetected GC III	Calibrated	Race, Sex, Age	Greater than Undetected GC I to Undetected GC II transition rates; Bound by values derived from sojourn times from literature [2]; Transition rates for women greater than those for men
	Undetected GC III to Undetected GC IV	Calibrated	Race, Sex, Age	Greater than Undetected GC II to Undetected GC III transition rates; Bound by values derived from sojourn times from literature [2]; Transition rates for women greater than those for men
Detection	Undetected GC I to Detected GC I	Calibrated	Race, Sex, Age	None
	Undetected GC II to Detected GC II	Calibrated	Race, Sex, Age	Greater than Undetected GC I to Detected GC I transition rates
	Undetected GC III to Detected GC III	Calibrated	Race, Sex, Age	Greater than Undetected GC II to Detected GC II transition rates
	Undetected GC IV to Detected GC IV	Calibrated	Race, Sex, Age	Greater than Undetected GC III to Detected GC III transition rates
Cancer Death	Detected GC I to Cancer Death	SEER Survival	Race, Sex, Age	N/A
	Detected GC II to Cancer Death	SEER Survival	Race, Sex, Age	N/A
	Detected GC III to Cancer Death	SEER Survival	Race, Sex, Age	N/A
	Detected GC IV to Cancer Death	SEER Survival	Race, Sex, Age	N/A
Other Death	All states to Other Death	Life table Generator	Race, Sex, Age	N/A

AG – Atrophic Gastritis, IM – Intestinal Metaplasia, Dys – Dysplasia, GC – Gastric Cancer

Calibration Targets

Using simulated annealing with a sum-squared error-based objective function, GSiMo calibrates parameters to SEER incidence and precursor prevalence targets. Target data is ranked and assigned weights based on sample size and representativeness, so sources with smaller sample sizes and no stratification by sex/race contribute less to the objective function score. Therefore, GSiMo is calibrated primarily to SEER data followed by precursor prevalence data from literature.

Case listings data for the histology groupings listed in Table 2 were pulled from the SEER 18 database [3]. Stage at diagnosis, in accordance with AJCC (I-IV) or historical (local, regional, distant) staging, was also extracted from these case listings. In order to maximize the utility of the SEER data, missing data was imputed using the *mice* package in R, which utilizes the Multiple Imputation through Chained Equations (MICE) method to impute data. This imputation of data included reclassifying the NOS cases, of which there were a significant number, as intestinal or diffuse. AJCC stages were also imputed for cases with only historical staging or

missing stage information altogether. From the case listings data, stage distribution stratified by race/ethnicity, sex, and age was thus obtained.

Table 2. ICD codes used to filter for Gastric Cancer cases from SEER

Sites	Site ICD codes	Histology	Histology ICD codes
Non-cardia	16.1 - 16.9	Intestinal	8143 - 8144, 8210 - 8211, 8221, 8260 - 8263
		Diffuse	8141 - 8142, 8145, 8490
		NOS	8010, 8012, 8020 - 8021, 8140, 8201, 8230, 8310
		Other	All other cases of GC

Age-bucketed incidence rates for each subgroup were also extracted from SEER. Combining these incidence rates with the stage distribution data, incidence data stratified by race/ethnicity, sex, age, and stage were derived.

For the precursor prevalence targets, data from three sources in literature were used. For intestinal metaplasia (IM) prevalence, estimates from a U.S.-based national pathology database, provided by Dr. Robert Genta, included breakdowns by race/ethnicity subgroup, sex, and HP status. In this dataset, the "Other" race/ethnicity subgroup was primarily comprised of ≥90% non-Hispanic (NH) Whites and ≤10% African Americans. Although this group included a small proportion of non-White individuals, GSiMo's parameters for NH Whites were calibrated to the IM prevalence for this group in order to take advantage of this robust dataset.

For NH Blacks, only overall IM prevalence estimates were available from the literature, lacking further breakdowns by sex, age, and HP status. To estimate IM prevalence for NH Blacks across these categories, we applied the overall prevalence ratio of IM between Blacks and Whites in the U.S. from literature [4] to the sex- and age group-specific data used for NH Whites in Dr. Genta's dataset to approximate the corresponding values for NH Blacks.

Additional prevalence estimates for atrophic gastritis (AG) and dysplasia (DYS) in countries with low gastric cancer incidence were sourced from the literature [5]. These estimates were not stratified by any demographic characteristics.

Screening and Intervention

In addition to the natural history parameters, extra parameters are required to simulate screening and intervention strategies. These parameters are either taken from literature or estimated by expert opinion and vary with the strategy being tested. An incomplete list of parameters includes test performance characteristics such as sensitivity and specificity, costs, quality of life adjustments, and treatment efficacy.

References

1. National Institute of Health. Surveillance, Epidemiology, and End Results (SEER) Program SEER*Stat Database: Incidence - SEER Research Data, 9 Registries. 2021;
2. Broder M. S., Ailawadhi S., Beltran H., Blakely L. J., Budd G. T., Carr L., Cecchini M., Cobb P. W., Gibbs S. N., Kansal A., Kim A., Monk B. J., Schwartzberg L. S., Wong D. J. L., Yermilov I. Estimates of stage-specific preclinical sojourn time across 21 cancer types. Journal of Clinical Oncology. 2021;39(15_suppl):e18584–e18584.
3. National Institute of Health. Surveillance, Epidemiology, and End Results (SEER) Program SEER*Stat Database: Incidence - SEER Research Data, 18 Registries. 2021;
4. Altayar O., Davitkov P., Shah S. C., Gawron A. J., Morgan D. R., Turner K., Mustafa R. A. AGA Technical Review on Gastric Intestinal Metaplasia-Epidemiology and Risk Factors. Gastroenterology. 2020;158(3):732-744.e16.
5. Mulder DT, Hahn AI, Huang RJ, Zhou MJ, Blake B, Omofuma O, Murphy JD, Gutierrez-Torres DS, Zauber AG, O'Mahony JF, Camargo MC, Ladabaum U, Yeh JM, Hur C, Lansdorp-Vogelaar I, Meester R, Laszkowska M. Prevalence of Gastric Precursor Lesions in Countries With Differential Gastric Cancer Burden: A Systematic Review and Meta-analysis. Clin Gastroenterol Hepatol. 2024;22(8):1605–1617.

Component Overview

Summary

A description of the basic computational building blocks (components) of the model.

Overview

GSiMo consists of a natural history component, model stress testing component, and screening/intervention component.

Component Listing

Natural History

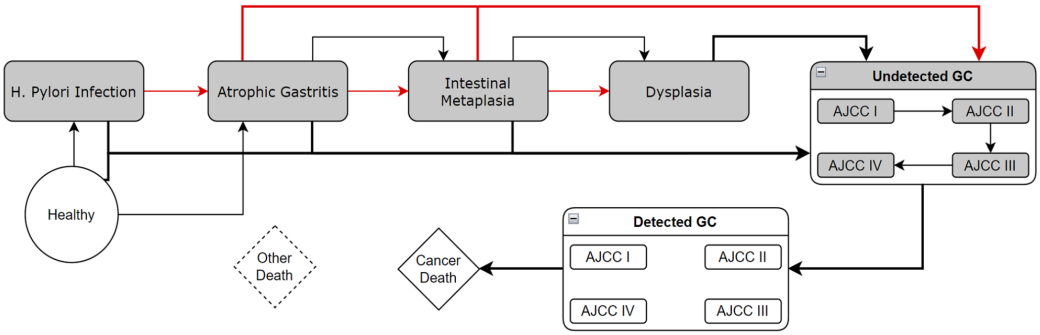

Figure 1. GSiMo model schematic. Red arrows indicate HP infection dependence in the transition probabilities. All states are connected to Other Death.

In the calibration component, a population-level Markov model is calibrated primarily to SEER GC incidence and stage distribution data, and secondarily to precursor prevalence targets. For each demographic subgroup, the starting population is initialized so that a proportion of the population starts in the HP state, in accordance with demographics-specific HP prevalence estimates among 18-year-olds, while the remainder start in the Healthy state. Populations are simulated from age 18 to age 84 to align with the availability of high-quality SEER incidence data prior to age 85. Fixed transition parameters are derived from common model input generators as well as SEER survival rates data. Calibrated transition parameters are determined via a bounded simulated annealing parameter search. See Parameter Overview and Assumption Overview for more details on model inputs and parameters.

A parameter set is calibrated for each race/ethnicity, sex, and age bracket (18-29, 30-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84) subgroup. Each set is a layer in a multidimensional transition probability matrix, allowing for parallelization of the calibration process and constraints across dimensions (Figure 3).

While the Markov model allows for time-efficient calibration, it is not sufficient to model transitions dependent on patient history beyond the most recent cycle due to its inherent memoryless property. To accurately model these transitions as well as screening and intervention strategies, a patient-level microsimulation was required.

In the microsimulation, individual patient trajectories are simulated. Patients are initialized with demographic characteristics, including race/ethnicity and sex, and HP infection status. As in the Markov, the proportion of HP-positive patients in the population aligns with HP prevalence estimates from the HP generator. From the Markov model's age-bucketed transition probabilities, single-age transition probabilities are smoothly interpolated using cubic spline interpolation (*csaps* package in R). The fitted splines are then used to extrapolate parameters for ages 85 to 100, a range for which there is a lack of high-quality target data, allowing for patients to be simulated from age 18 to 100. Figure 2 shows a representative transition matrix layer from the Markov and the microsimulation.

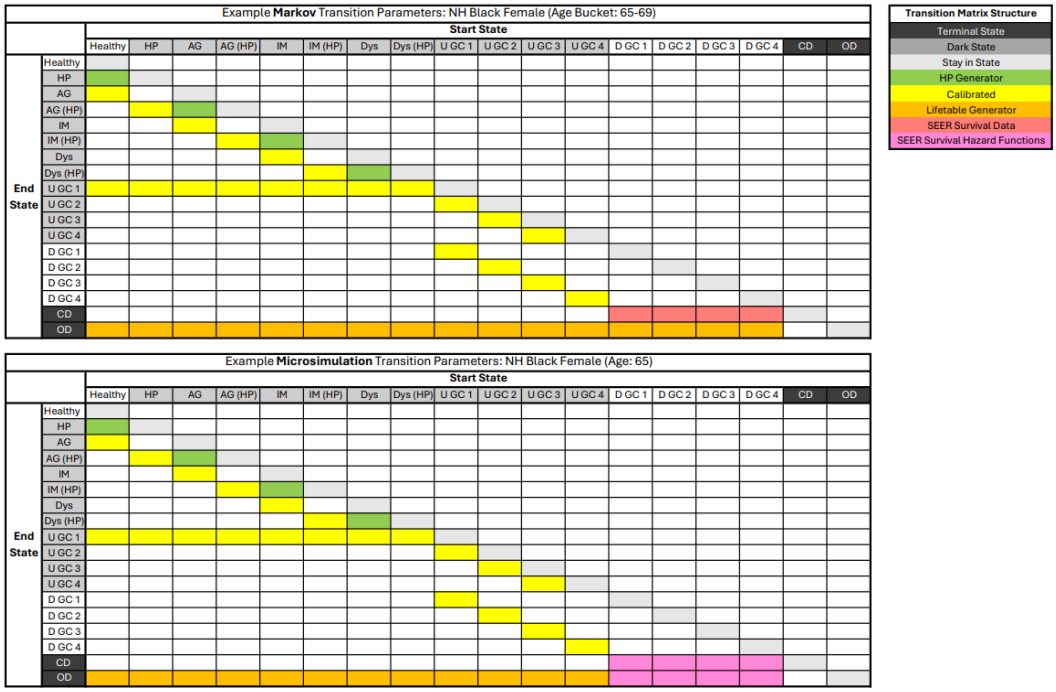

Figure 2. Transition Matrix Layers for the Markov vs the Microsimulation. Age-bucketed transition parameters from the Markov are smoothly interpolated to get single-age parameters which are then inputted into the microsimulation model. Color-coded cells indicate parameter sources.

As patients progress through the model, their state transition history along with duration in each state is recorded. This information in the microsimulation allows survival to be modeled as a function of both age and time since diagnosis.

Specifically, hazard functions dependent on race, sex, stage at diagnosis, age, and years survived with cancer (See Survival Hazards section in Parameter Overview) are used to determine a diagnosed cancer patient's competing risk of cancer mortality and other-cause mortality. At each cycle up to 10 years post-diagnosis, a patient's probability of dying from cancer and probability of dying from other causes at that point in time are used to sample an outcome from the following: cancer death, other death, and stay in state. After 10 years in the cancer state, the patient is assumed to have survived cancer and is moved back to either the healthy or HP-infected state.

Figure 3 provides an overview of the entire natural history model development process.

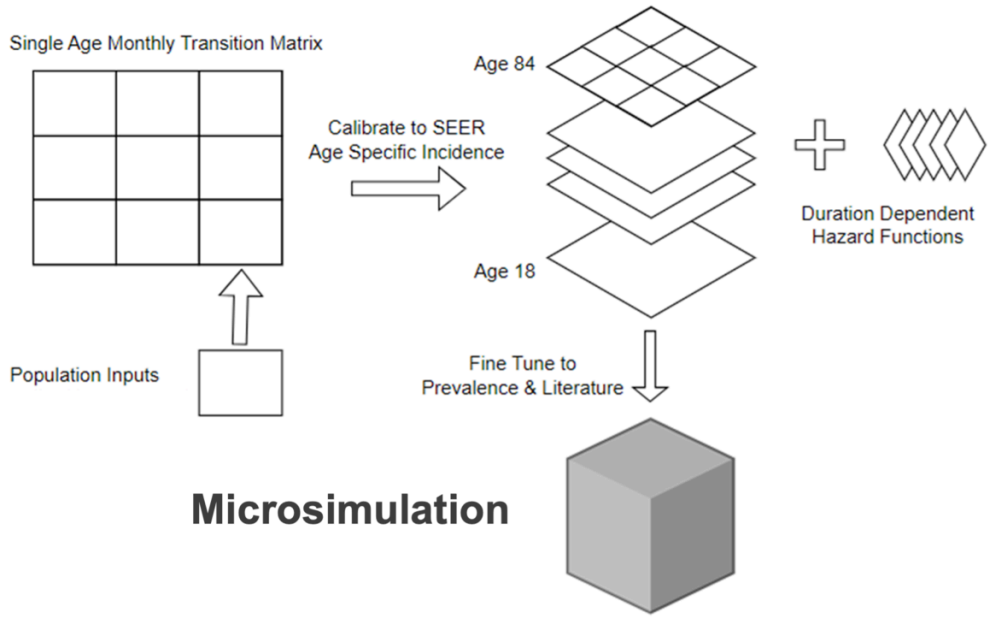

Figure 3. Model development process. Transition probability matrices are calibrated for each demographic and age-bucket grouping. The transition probabilities and duration-dependent hazard functions are then inputted into the microsimulation model.

After every patient is run in the simulated cohort, model outputs such as incidence, prevalence, dwell times, etc. are extracted from the patient state-transition logs. Outputs from the natural history simulation are used as a baseline for the assessment of screening and intervention strategies.

Model Stress Testing

The Maximum Clinical Incidence Reduction (MCLIR) framework is used to clarify how model assumptions and structure impact outcome predictions. Key factors influencing the effectiveness of cancer prevention methods include the onset and duration of preclinical disease, the probability of detecting preclinical disease, and the effectiveness of treatment following preclinical disease detection. MCLIR comprises four scenarios designed to evaluate differences in these aspects in an unrealistic, perfect screening and treatment context. Additional frameworks, Maximum Sensitivity Realistic Treatment (MSRT) and Realistic Clinical Incidence Reduction (RCLIR), were developed to assess model differences in more realistic screening and treatment contexts. Table 1 lists the parameters for all scenarios.

Table 1. MCLIR, MSRT, and RCLIR scenario definitions.

Scenario	Age	Screening sensitivity	Treatment population	HP treatment	Precursor disease treatment	Cancer treatment
MCLIR 1	20	100%	HP+	100% eradication	All precursors, 100% removal	100% removal
MCLIR 2	65	100%	HP+	100% eradication	All precursors, 100% removal	100% removal
MCLIR 3	65	100%	All	0% eradication	All precursors, 100% removal	100% removal
MCLIR 4	65	100%	All	100% eradication	All precursors, 100% removal	100% removal
MSRT 1	20	100%	HP+	80% eradication	Dysplasia only, 100% removal	100% removal
MSRT 2	65	100%	HP+	80% eradication	Dysplasia only, 100% removal	100% removal
MSRT 3	65	100%	All	0% eradication	Dysplasia only, 100% removal	100% removal
MSRT 4	65	100%	All	80%	Dysplasia only, 100%	100% removal

Scenario	Age	Screening sensitivity	Treatment population	HP treatment	Precursor disease treatment	Cancer treatment
				eradication	removal	
RCLIR 1	20	HP: 91% Dys: 71% EGC: 71% AGC: 92%	HP+	80% eradication	Dysplasia only, 100% removal	100% removal
RCLIR 2	65	HP: 91% Dys: 71% EGC: 71% AGC: 92%	HP+	80% eradication	Dysplasia only, 100% removal	100% removal
RCLIR 3	65	HP: 91% Dys: 71% EGC: 71% AGC: 92%	All	0% eradication	Dysplasia only, 100% removal	100% removal
RCLIR 4	65	HP: 91% Dys: 71% EGC: 71% AGC: 92%	All	80% eradication	Dysplasia only, 100% removal	100% removal

Scenarios are characterized by screening age, screening sensitivity, treatment target population, and treatment effectiveness. HP – *H. pylori*, Dys – Dysplasia, EGC – Early Gastric Cancer, AGC – Advanced Gastric Cancer.

Scenarios are implemented by imposing screening and treatment parameters on the natural history model. At the age of intervention, HP infection, precursor disease, and gastric cancer detection is probabilistically sampled based on the specified sensitivity. If the patient belongs to the treatment group and detection is successful, a treatment outcome is similarly sampled using the specified efficacy. Patients that are treated successfully for HP cannot be infected with HP again.

After running each scenario, incidence and incidence reduction relative to natural history incidence are calculated. Additional outputs include cancer prevalence, proportion of cancer cases attributable to HP infection, number of precursor disease cases successfully treated, and number of cancer cases averted.

Screening and Intervention

This section will be updated once the screening and intervention component is completed.

Columbia
Output Overview

Output Overview

Summary

This document provides an overview of the outputs produced by GSiMo.

Overview

GSiMo's outputs can be broadly divided into natural history outcomes and screening/intervention outcomes. Currently, only natural history outputs can be extracted from GSiMo. When the screening and intervention component is completed, additional outputs will be calculated. All outputs are stratified by demographic subgroup.

Output Listing

Natural History

- GC age-specific incidence rates
- GC mortality
- Proportion of GC cases attributable to HP infection
- Preclinical disease prevalence
- Progression rates/Dwell times

Screening and Intervention

Epidemiological	Benefits	Harms	Economic
• Number of precancerous lesions • Number of GC cases • Number of GC deaths • Number of screening tests • Number of surveillance procedures	• Cancer cases prevented • Cancer deaths averted • Life years (LY) gained • QALYs • QALYs gained	• Endoscopic complications • Surgical deaths	• Total costs

CISNET
Columbia
Results Overview

COLUMBIA UNIVERSITY

Results Overview

Summary

This document provides a summary of the model results from GSiMo's development and application.

Overview

Listed here are the natural history outputs and preliminary MCLIR results for the current iteration of GSiMo. Additional outputs and results will be included here as they become available.

Results List

Natural History

Incidence

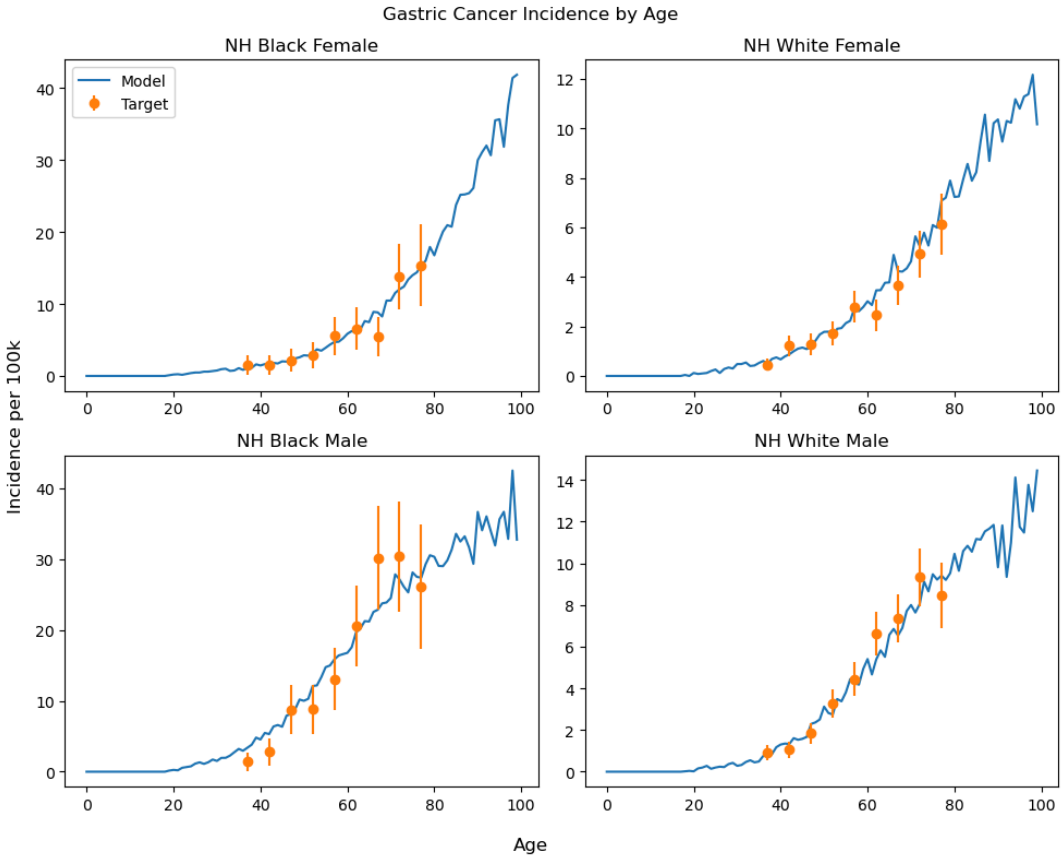

Figure 1. GC Age-Specific Incidence by demographic subgroup. The calibration target is SEER incidence data.

Dwell Times

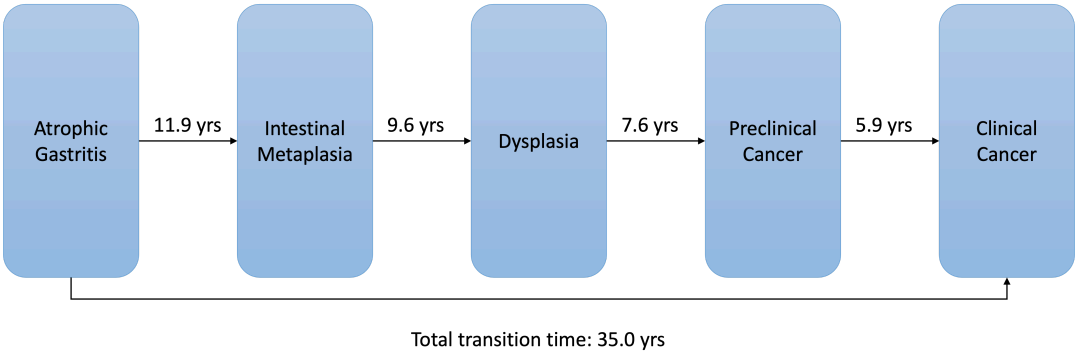

Figure 2. GSiMo mean dwell times.

Model Stress Testing

The following are a selection of results from the Maximum Clinical Incidence Reduction (MCLIR) analysis. GSiMo's outputs generally align with the other Gastric models. GSiMo deviates most from the other models in MCLIR Scenario 1, defined as screening/intervention at age 20 with perfect screening and perfect treatment in HP-positive patients. The comparatively higher incidence right after intervention age can be attributed to the proportion of diffuse cases that progress directly to cancer instead of through Correa's Cascade (Figure 3).

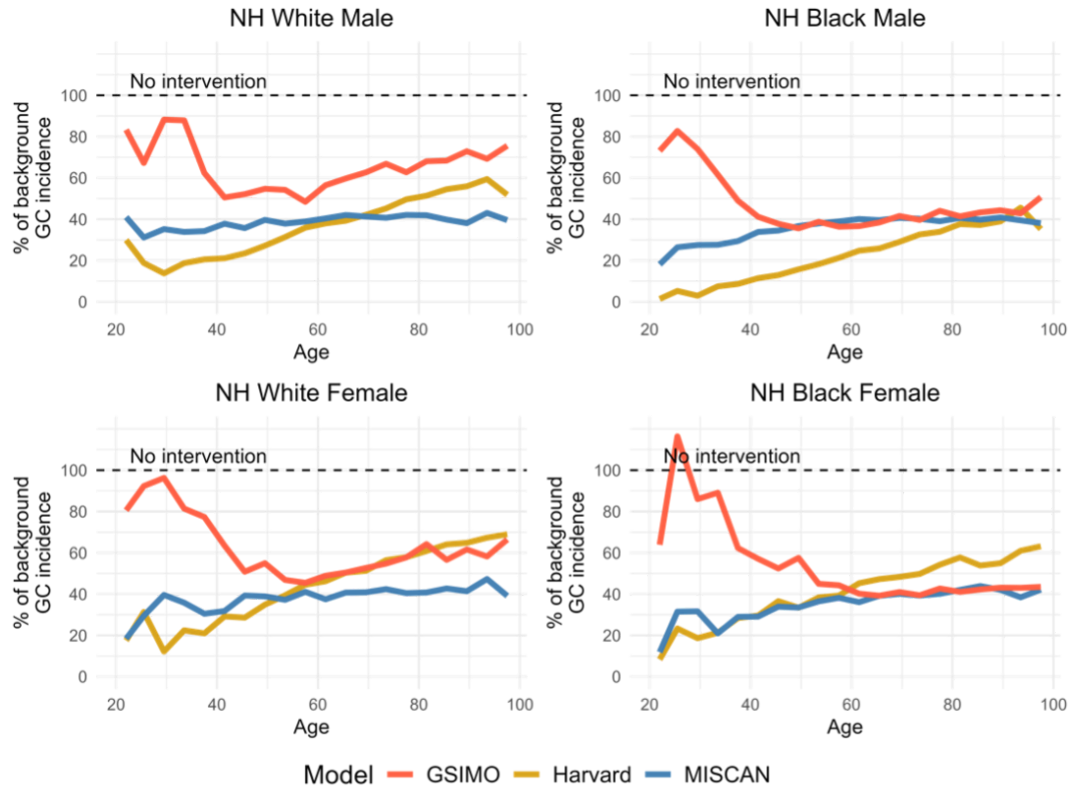

Figure 3. MCLIR Scenario 1 Incidence. Screening/intervention age: 20, screening sensitivity: 100%, treatment efficacy: 100%, and treatment population: HP-positive patients.

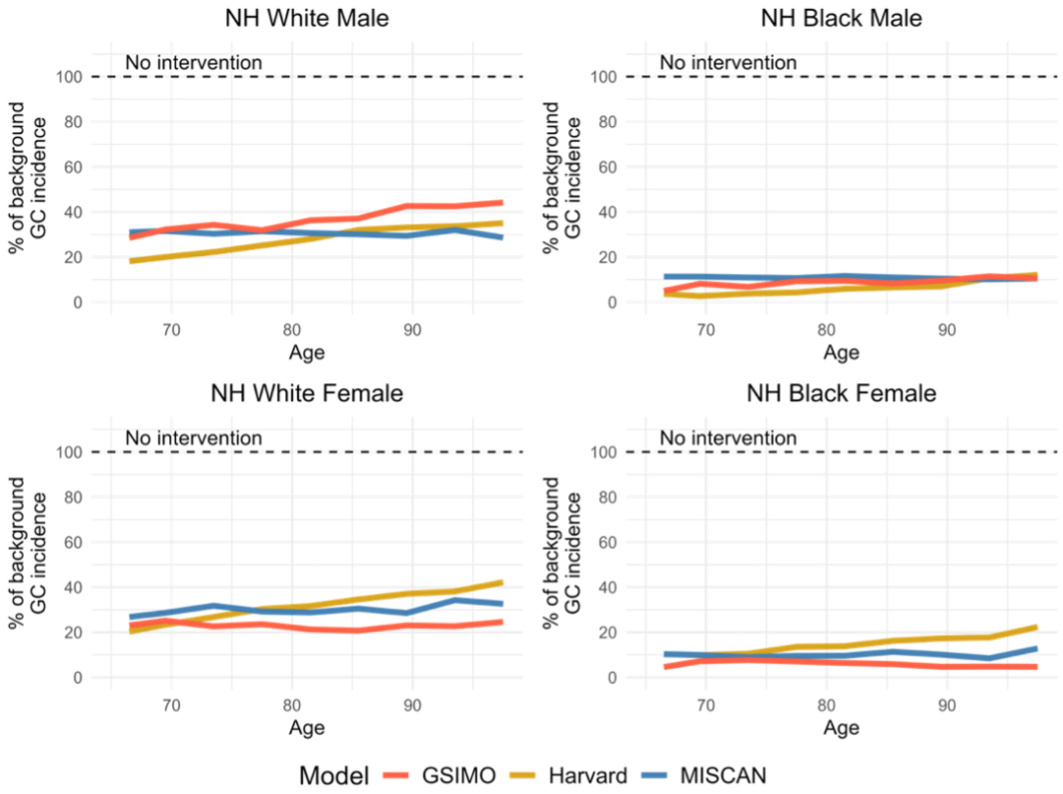

Figure 4. MCLIR Scenario 2 Incidence. Screening/intervention age: 65, screening sensitivity: 100%, treatment efficacy: 100%, and treatment population: HP-positive patients.

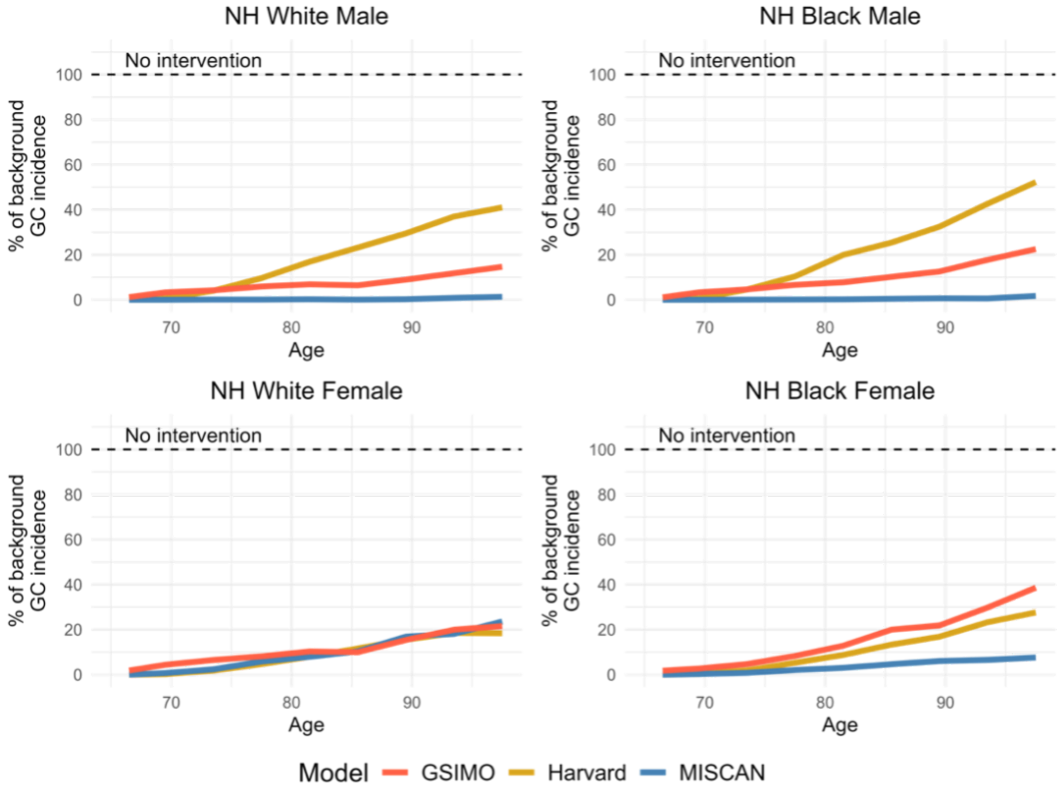

Figure 5. MCLIR Scenario 3 Incidence. Screening/intervention age: 65, screening sensitivity: 100%, treatment efficacy: 100%, and treatment population: all patients.

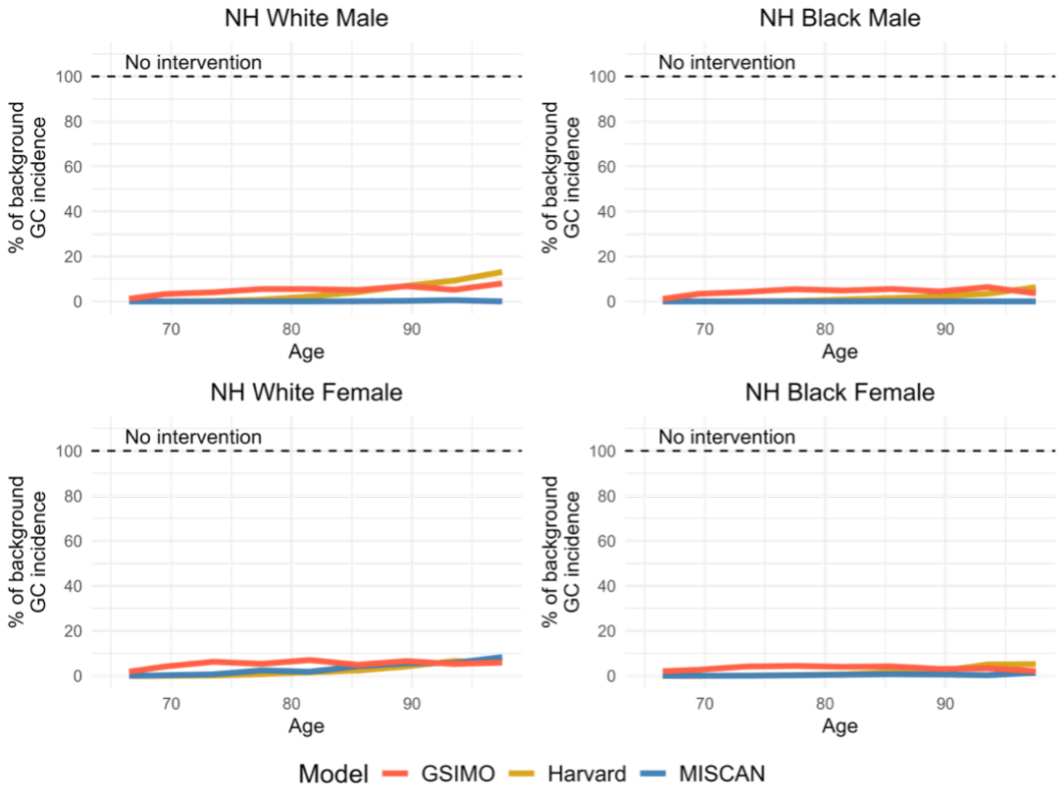

Figure 6. MCLIR Scenario 4 Incidence. Screening/intervention age: 65, screening sensitivity: 100%, treatment efficacy: 100%, and treatment population: all patients

Full results from the MCLIR analysis will be linked here once published.

Screening and Intervention

Results from screening and intervention analyses will be reported here as soon as they are available.

Key References

Altayar O., Davitkov P., Shah S. C., Gawron A. J., Morgan D. R., Turner K., Mustafa R. A. AGA Technical Review on Gastric Intestinal Metaplasia-Epidemiology and Risk Factors. Gastroenterology. 2020;158(3):732-744.e16.

American Association for Cancer Research. Cancer Disparities Progress Report 2024. 2024.

Broder M. S., Ailawadhi S., Beltran H., Blakely L. J., Budd G. T., Carr L., Cecchini M., Cobb P. W., Gibbs S. N., Kansal A., Kim A., Monk B. J., Schwartzberg L. S., Wong D. J. L., Yermilov I. Estimates of stage-specific preclinical sojourn time across 21 cancer types. Journal of Clinical Oncology. 2021;39(15_suppl):e18584–e18584.

Correa P., Piazuelo M. B. The gastric precancerous cascade. J Dig Dis. 2012;13(1):2–9.

Ford A. C., Forman D., Hunt R. H., Yuan Y., Moayyedi P. Helicobacter pylori eradication therapy to prevent gastric cancer in healthy asymptomatic infected individuals: systematic review and meta-analysis of randomised controlled trials. BMJ. 2014;348:g3174.

Gupta S, Li D, El Serag HB, Davitkov P, Altayar O, Sultan S, Falck-Ytter Y, Mustafa RA. AGA Clinical Practice Guidelines on Management of Gastric Intestinal Metaplasia. Gastroenterology. 2020;158(3):693–702.

Lee YC, Chiang TH, Chou CK, Tu YK, Liao WC, Wu MS, Graham DY. Association Between Helicobacter pylori Eradication and Gastric Cancer Incidence: A Systematic Review and Meta-analysis. Gastroenterology. 2016;150(5):1113–1124.

Lyon, France: International Agency for Research on Cancer (IARC Working Group Reports, No. 8). Helicobacter pylori Eradication as a Strategy for Preventing Gastric Cancer. 2014;

Mulder DT, Hahn AI, Huang RJ, Zhou MJ, Blake B, Omofuma O, Murphy JD, Gutierrez-Torres DS, Zauber AG, O'Mahony JF, Camargo MC, Ladabaum U, Yeh JM, Hur C, Lansdorp-Vogelaar I, Meester R, Laszkowska M. Prevalence of Gastric Precursor Lesions in Countries With Differential Gastric Cancer Burden: A Systematic Review and Meta-analysis. Clin Gastroenterol Hepatol. 2024;22(8):1605–1617.

National Cancer Institute. Cancer Stat Facts. Surveillance, Epidemiology, and End Results Program. 2018. https://seer.cancer.gov/statfacts/. Accessed October 27, 2024. 2018;

National Institute of Health. Surveillance, Epidemiology, and End Results (SEER) Program SEER*Stat Database: Incidence - SEER Research Data, 9 Registries. 2021;

National Institute of Health. Surveillance, Epidemiology, and End Results (SEER) Program SEER*Stat Database: Incidence - SEER Research Data, 18 Registries. 2021;

Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021;71(3):209–249.