Combined Model Profile
Version: 1.0.00
Released: 2025-09-30

Gastric Cancer Combined Model Profile

Individual Model Profiles

Columbia Gastric Cancer Simulation Model (GSiMo): Model Profile
Columbia University Irving Medical Center
Version: 1.0.00
Released: 2025-09-30

Harvard Gastric Cancer-United (GC-US) Microsimulation: Model Profile
Harvard University
Version: 1.0.00
Released: 2025-09-30

MIcrosimulation SCreening ANalysis Gastric Cancer Model (MISCAN-Gastric): Model Profile
Erasmus University Medical Center
Version: 1.0.00
Released: 2025-09-30

Suggested citation

CISNET Gastric Working Group. Gastric Cancer Combined 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-combined-model-profile-1.0.00-2025-09-30.pdf

Combined Model Profile Version Table

Version	Date	Notes
1.0.00	2025-09-30	Initial release

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
Readers Guide

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.

Columbia
Model Purpose

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.

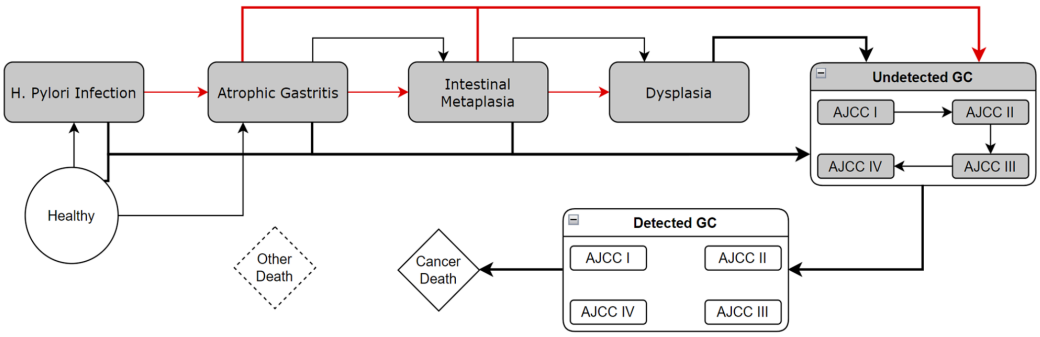

Figure 1. GSiMo model schematic. Red arrows indicate HP infection dependence in the transition probabilities. All states are connected to Other Death.

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

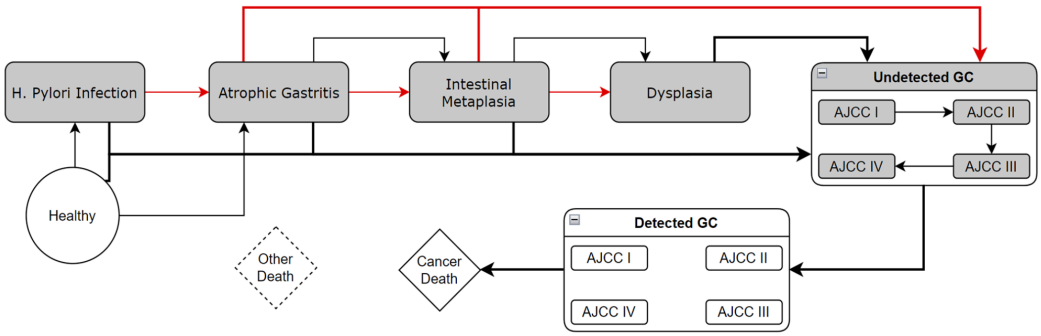

Figure 1. GSiMo model schematic. Red arrows indicate HP infection dependence in the transition probabilities. All states are connected to Other Death.

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
	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

Parameter		Source	Varies by	Constraints
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

Sites	Site ICD codes	Histology	Histology ICD codes
		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

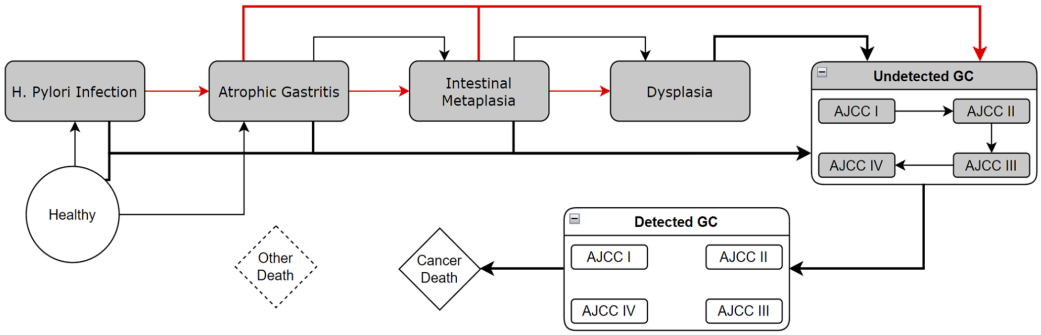

Figure 1. GSiMo model schematic. Red arrows indicate HP infection dependence in the transition probabilities. All states are connected to Other Death.

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.

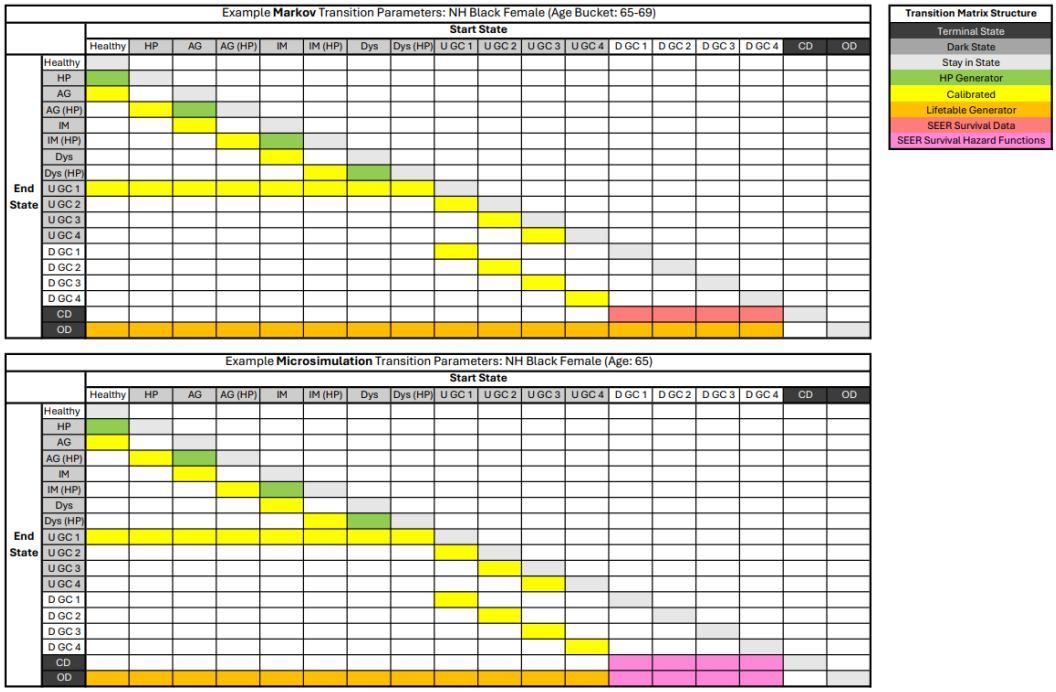

Figure 2. Transition Matrix Layers for the Markov vs the Microsimulation. Age-bucketed transition parameters from the Markov are smoothly interpolated to get single-age parameters which are then inputted into the microsimulation model. Color-coded cells indicate parameter sources.

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.

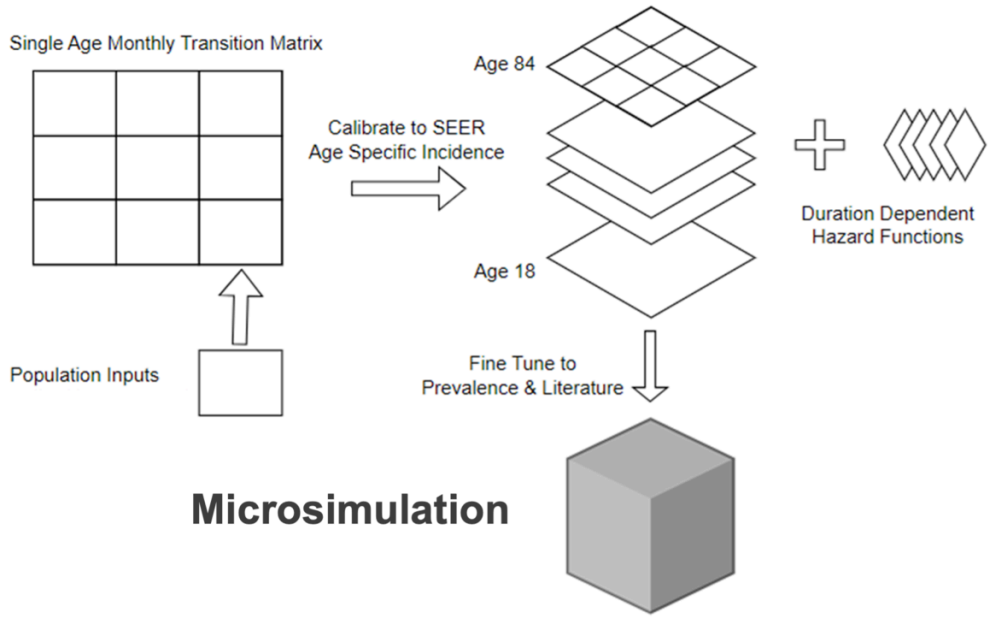

Figure 3. Model development process. Transition probability matrices are calibrated for each demographic and age-bucket grouping. The transition probabilities and duration-dependent hazard functions are then inputted into the microsimulation model.

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% eradication	Dysplasia only, 100% removal	100% removal
RCLIR 1	20	HP: 91%	HP+	80% eradication	Dysplasia only, 100% removal	100% removal

Scenario	Age	Screening sensitivity	Treatment population	HP treatment	Precursor disease treatment	Cancer treatment
		Dys: 71% EGC: 71% AGC: 92%				
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

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

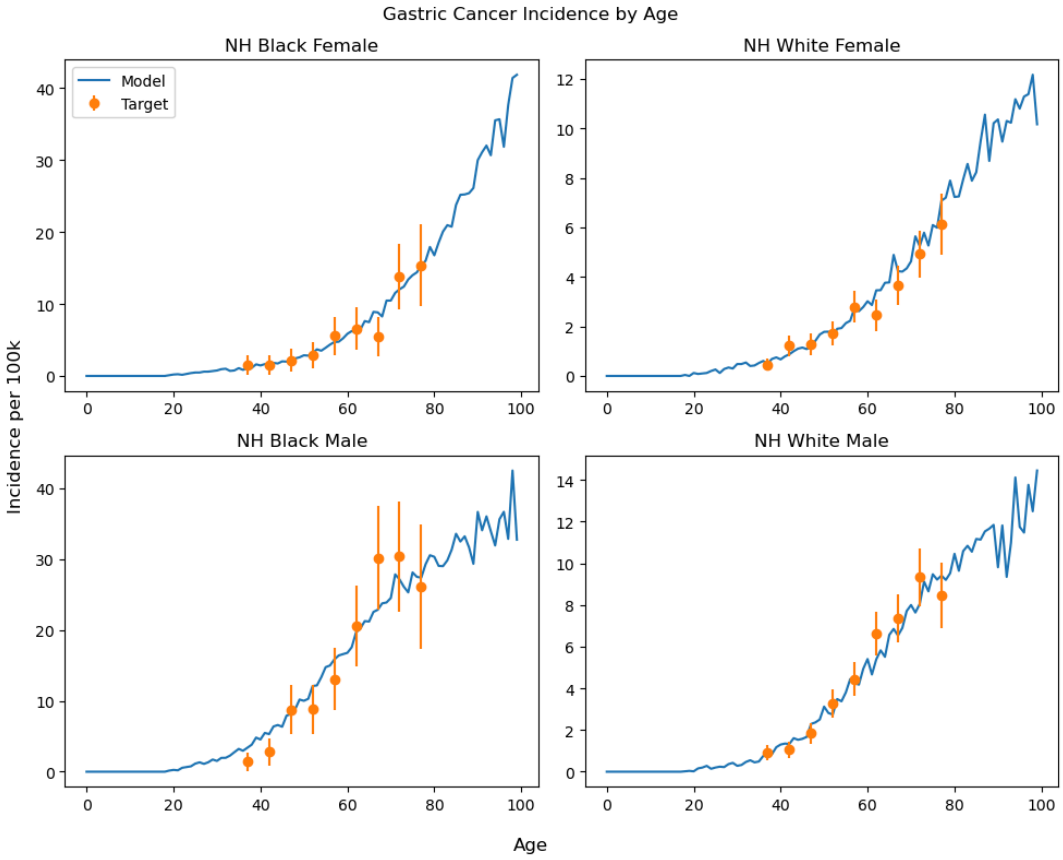

Figure 1. GC Age-Specific Incidence by demographic subgroup. The calibration target is SEER incidence data.

Dwell Times

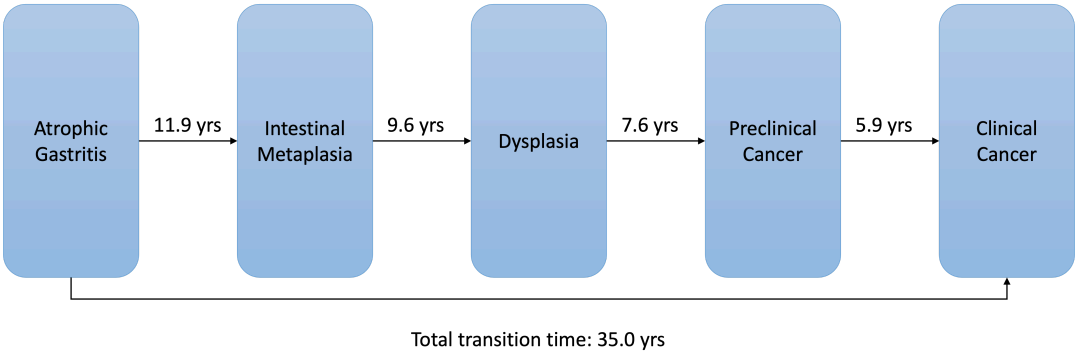

Figure 2. GSiMo mean dwell times.

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).

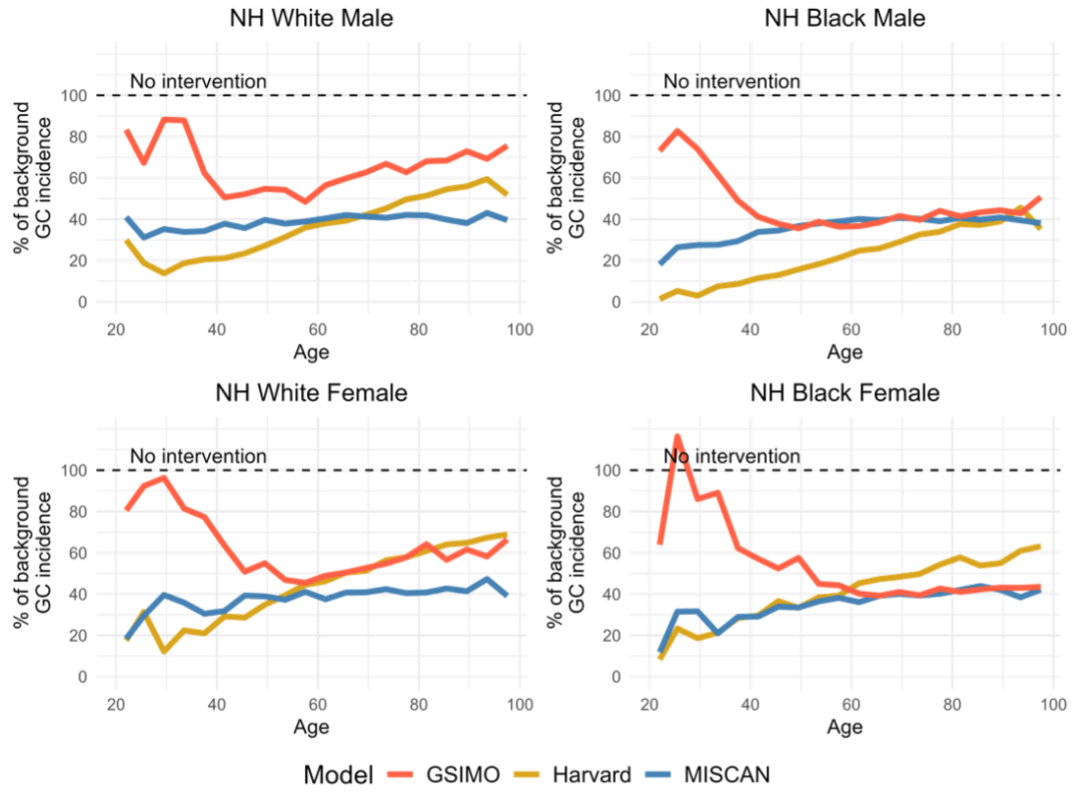

Figure 3. MCLIR Scenario 1 Incidence. Screening/intervention age: 20, screening sensitivity: 100%, treatment efficacy: 100%, and treatment population: HP-positive patients.

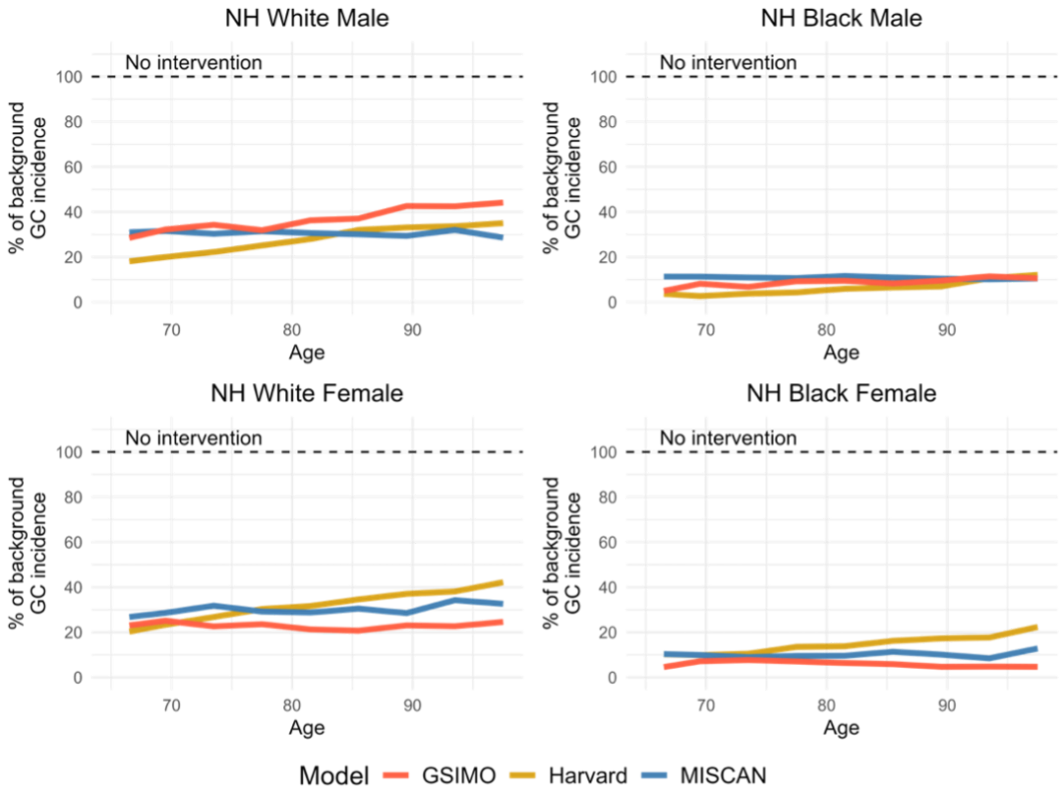

Figure 4. MCLIR Scenario 2 Incidence. Screening/intervention age: 65, screening sensitivity: 100%, treatment efficacy: 100%, and treatment population: HP-positive patients.

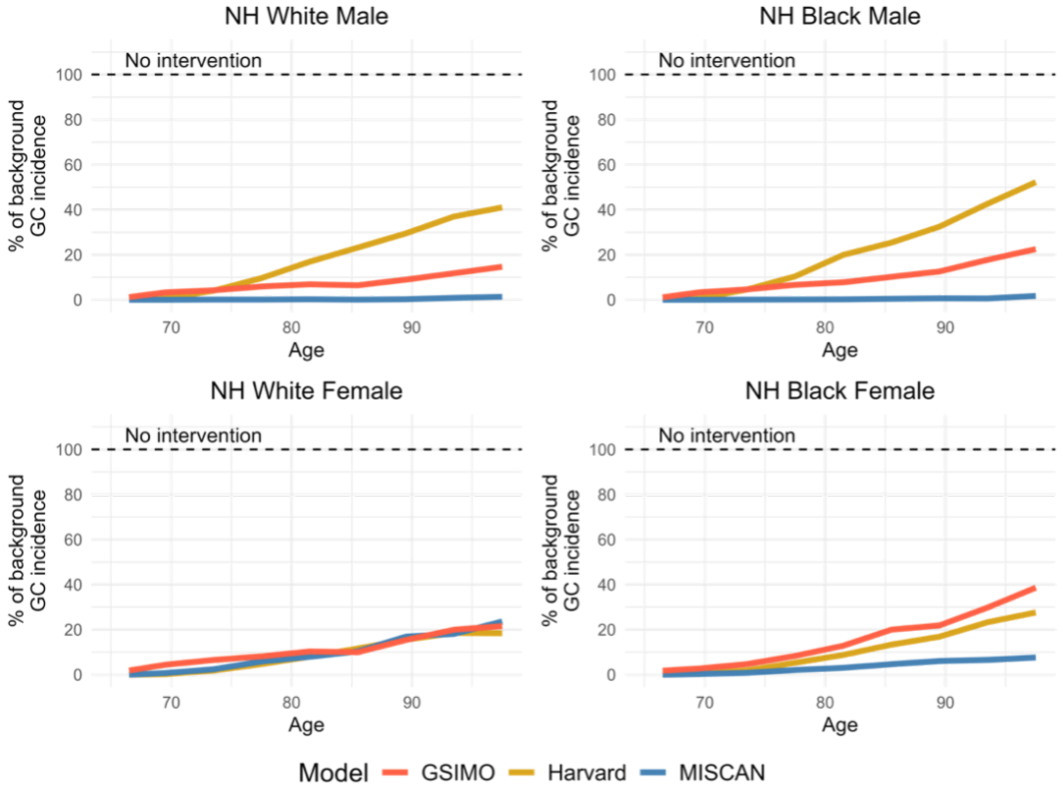

Figure 5. MCLIR Scenario 3 Incidence. Screening/intervention age: 65, screening sensitivity: 100%, treatment efficacy: 100%, and treatment population: all patients.

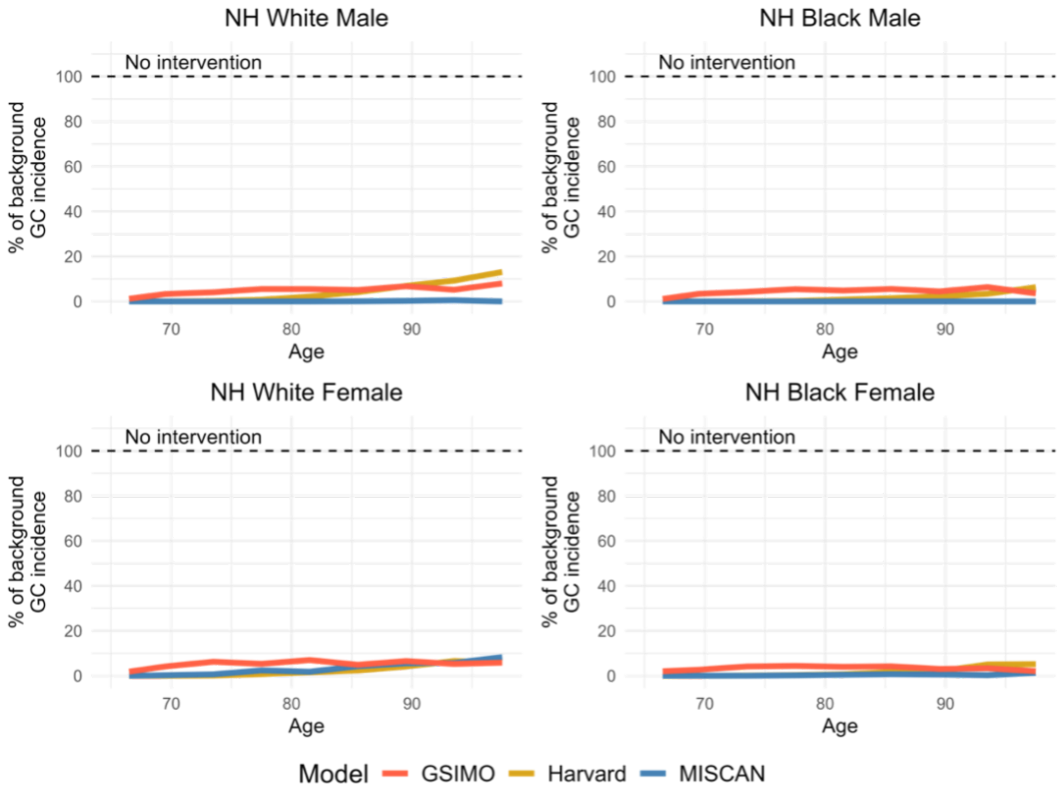

Figure 6. MCLIR Scenario 4 Incidence. Screening/intervention age: 65, screening sensitivity: 100%, treatment efficacy: 100%, and treatment population: all patients

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.

Columbia
Key References

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.

Harvard
Version: 1.0.00
Released: 2025-09-30

Harvard Gastric Cancer-United (GC-US) Microsimulation: Model Profile

Harvard University

Contact

Zach Ward (zward@hsph.harvard.edu)
Jennifer Yeh (jennifer.yeh@childrens.harvard.edu)

Funding

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

Suggested Citation

Zachary J. Ward, Chelsea S. Taylor, M. Constanza Camargo, M. Blanca Piazuelo, Jennifer M. Yeh. Harvard Gastric Cancer-United (GC-US) Microsimulation: 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-harvard-gc-model-profile-1.0.00-2025-09-30.pdf

Version Table

Version	Date	Notes
1.0.00	2025-09-30	Initial release

Harvard
Readers Guide

Reader's Guide

Core Profile Documentation

These topics will provide an overview of the model without the burden of detail. Each can be read in about 5-10 minutes. 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.

Further Reading

These topics will provide a intermediate level view of the model. Consider these documents if you are interested gaining in a working knowledge of the model, its inputs and outputs.

Model Purpose

Summary

This page describes the purposes for which the Harvard Gastric Cancer-United States (GC-US) model was developed.

Purpose

The Harvard GC-US model was developed for several purposes.

- As a population model with multiple demographic groups, the model aims to simulate the impact of demographic and epidemiologic trends on gastric cancer incidence and mortality.
- The model simulates 14 subtypes of gastric cancer, which allows us to examine trends in site- and histological-specific gastric cancers, and how these may vary by subgroups.
- For each cancer type, the model simulates the natural history, and also the impact of health system factors on diagnosis and treatment, allowing us to explore disparities in gastric cancer incidence and outcomes.
- The model includes GC risk factors such as *H. pylori* and smoking, allowing for analyses of attributable risk and impact of primary prevention strategies to be conducted.
- The impact and cost-effectiveness of secondary prevention (e.g., screening and surveillance) by various modalities can also be assessed.
- Lastly, we aim to develop a global version of the model (GC-Global) to perform similar analyses at a global level.

Model Overview

Summary

The Harvard-GC model is a microsimulation population model that currently includes 3 components:

- Demographic Generator
- GC Natural History
- Screening Module

Purpose

The Harvard-GC model was developed for several purposes - see Model Purpose.

Background

The Harvard-GC model is a microsimulation (individual-level) model that simulates the US population. The natural histories of 14 gastric cancer subtypes are simultaneously simulated for each individual, accounting for trends in risk factors and competing mortality that vary by demographic subgroup and over time. This allows subgroup- and subtype-specific analysis to be conducted for a virtual population that is representative of the United States.

Model Description

Below we briefly describe the main components of the model - for more details see Component Overview.

Demographic Generator

We model a full population (ages 0-100) of individuals from 1975 to 2020, accounting for sex, race/ethnicity, and nativity (US vs foreign-born), and account for subgroup-specific trends in competing mortality and risk factors.

Natural History

We model 14 subtypes of gastric cancer that together account for all diagnosed cases of GC. Adenocarcinoma natural history progression is modelled using Correa's cascade, while a simplified progression framework is used for other GC subtypes. The model is calibrated to empirical data on GC incidence (total and by type) from SEER (1975-2019), overall and by subgroup.

Screening Module

Screening for pre-cancerous lesions can be modelled, accounting for test characteristics, costs, and treatment efficacy. Risk factor screening (e.g., primary prevention) can also be simulated in the model.

Assumption Overview

Summary

An overview of the basic assumptions of the Harvard GC-US model.

Background

Each component of the model relies on some simplifying assumptions, detailed below.

Key Assumptions

Demographic Generator

- We assume that net migration is non-differential by smoking status (conditional on year, age, and sex).
- We assume that Former Smokers do not resume smoking after smoking cessation.
- We allowed smoking initiation/cessation rates to vary by native vs foreign-born (after accounting for sex and race/ethnicity)
- We assumed that foreign-born individuals faced the same background mortality rates as US-born individuals (conditional on age, sex, and race/ethnicity)
- Although we account for differential competing mortality by smoking status, we assume that *H. pylori* status does not impact competing (background) mortality.

Natural History

- We used Bayesian hierarchical models for all parameters to allow values to vary by subgroup (sex + race/ethnicity).
- We allowed *H. pylori* status to impact progression probabilities for Adenocarcinomas and MALT, while smoking was allowed to impact progression for all cancer types.
- We allowed the *H. pylori* hazard ratios on precursor progression to also vary by native vs foreign-born to account for potential differences in HP strains.
- We assume that detection probabilities are non-decreasing (i.e., weakly monotonic) by cancer stage.
- We assume that undetected Stage IV cancers have a risk of dying before diagnosis, with priors based on 5-year net survival estimates from SEER assuming a 2-year lead time bias. However, as these undetected patients are not undergoing treatment this may be a conservative assumption as they have no survival benefit from treatment.

Screening

- We assume that 100% of the population complies with recommended screening when simulating the impact of screening policies.

Parameter Overview

Summary

This page provides an overview of the Harvard GC-US model parameters.

Background

We group parameters by component below: Demographic Generator, Natural History, and Screening Module. We used Bayesian hierarchical models to allow parameter values to vary by demographic subgroup, and account for uncertainty around all model parameters by sampling from the best-fitting 100 parameter sets.

Parameter Summary

Demographic Generator

Cohort size: To model birth cohort sizes (i.e., age 0 by year), we used parameters to vary the size of the birth cohort (by race/ethnicity) in each year relative to the estimated 2000 birth cohort. Parameters were set using knots every 10 years which were interpolated using cubic splines.

Net migration: We estimated the % of foreign-born respondents (overall and by sex, race/ethnicity, age) for each year from the American Community Survey. We allow net migration rates to vary by year, sex, race/ethnicity, and age.

Baseline mortality: Based on US lifetables, we model baseline mortality (i.e., mortality for never smokers) by year, age, sex, and race/ethnicity.

Smoking initation: Annual probability of starting smoking (varies by year, age, sex, race/ethnicity, nativity).

Smoking cessation: Annual probability of quitting smoking (varies by year, age, sex, race/ethnicity, nativity).

Smoking mortality hazard ratio: Hazard ratio of mortality (i.e., modifies baseline mortality) by smoking status: current or former smoking (varies by year, age, sex, race/ethnicity).

HP infection: Annual probability of acquiring HP infection (varies by year, age, sex, race/ethnicity, nativity).

Natural History

Progression: Annual probability of progressing to the next pre-cancerous health state. Parameter values are allowed to vary by GC subtype and health state. Progression from Healthy is allowed to vary by year to account for secular trends. Progression from IM and Dysplasia is also age-dependent. All progression probabilities are also allowed to vary by sex and race/ethnicity.

Progression hazard ratio: Hazard ratio of progression by HP status and smoking status (allowed to vary by GC subtype, sex, and race/ethnicity).

Stage progression: Annual probability of progressing to the next stage of invasive gastric cancer (I-IV) - only simulated for undetected individuals. Allowed to vary by GC subtype, stage, sex, and race/ethnicity.

Diagnosis: Annual probability of having gastric cancer detected - assumed to increase with stage. Allowed to vary by GC subtype, sex, and race/ethnicity.

Undetected mortality: Annual probability of dying from undetected Stage IV GC. Allowed to vary by GC subtype.

Screening Module

Screening parameters include test characteristics, costs, specified frequency/eligibility of screenign policies, and treatment efficacy.

Component Overview

Summary

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

Overview

As described in the Model Overview, the Harvard GC model comprises 3 components: Demographic Generator, Natural History, and Screening.

Components

Demographic Generator

We model a full population (ages 0-100) of individuals from 1975 to 2020. We model each birth cohort starting in 1875 so that a full population of ages has been initialized in the model starting in 1975. We weight each race (i.e. oversampling smaller groups) to improve computational efficiency and stability of estimates. We model subgroups based on the following characteristics:

- Sex: Male/Female
- Race/Ethnicity: Based on 6 mutually exclusive race/ethnicity subgroups as defined by the US Census
    - White, non-Hispanic (White)
    - Black, non-Hispanic (Black)
    - Hispanic
    - American Indian/Alaska Native, non-Hispanic (AIAN)
    - Asian/Pacific Islander, non-Hispanic (API)
    - Two or more races, non-Hispanic (Multi)
- Nativity: US-born/Foreign-born

The combination of these characteristics yields demographic 24 subgroups, allowing us to assess trends in GC disparities over time in the US. Subgroup-specific competing mortality and risk factor trends are also taken into account.

Natural History

We model 14 subtypes of gastric cancer: 2 sites (cardia vs non-cardia) X 7 histology groups (Adenocarcinoma-Intestinal, Adenocarcinoma-Diffuse, NET, GIST, MALT, Lymphoma (non-MALT), Other). Adenocarcinoma natural history progression is modelled using Correa's cascade, while a simplified progression framework is used for other GC subtypes. We also model the impact of risk factors (*H. pylori* and smoking) on progression. Stage progression, and stage-,time-,subgroup-dependent detection probabilities are simulated to account for trends (and potential disparities) in cancer detection. The model is calibrated to empirical data on GC incidence (total and by type) from SEER (1975-2019), overall and by subgroup.

Screening Module

Screening for pre-cancerous lesions can be modelled, accounting for test characteristics, costs, and treatment efficacy. Risk factor screening (e.g., primary prevention) can also be simulated in the model.

Output Overview

Summary

This page provides an overview of the outputs that can be generated by the Harvard GC-US model.

Outputs

The mean and 95% uncertainty intervals are estimated for all model outputs, and can be reported by GC subtype, year, age group, sex, race/ethnicity, and foreign-born status.

Specific outputs include:

- Demographic and risk factor profiles (e.g, HP and smoking prevalence trends)
- Prevalence of precancerous lesions
- Incidence of gastric cancer, overall and by subtype: total and diagnosed
- Deaths from undetected GC
- Deaths from detected GC
- Stage distribution at diagnosis
- Lifeyears and quality-adjusted life years

Results Overview

Summary

This page described the results that can be obtaiend from the Harvard GC-US model

Overview

The Harvard GC-US model has been developed for both epidemiologic estimation and policy analyses, as described below.

Results

GC Trends: The model provides a comprehensive analytic framework to synthesize data from multiple sources and estimate trends in gastric cancer by subtype and subgroup. These analyses provided epidemiologic information by demographic subgroup, age, and GC subtype, highlighting important trends in gastric cancer disparities in the United States. The model can also be used to project trends into the future.

Risk Factor Analysis: The model also provides a framework to estimate attributable incidence and mortality from gastric cancer to specific risk factors: *H. pylori* and smoking. These estimates, and how they may vary by demographic subgroup, can inform policy decisions and planning.

Screening Cost-Effectiveness: The model simulates various screening strategies, allowing for the incremental cost-effectiveness of competing strategies to be assessed. This will allow for more rigorous evidence-based policy-making. The inclusion of demographic subgroups also allows distributional cost-effectiveness analyses to be performed to assess potential impacts of policies on both population health and health equity.

Harvard
Key References

Key References

Erasmus MC
Version: 1.0.00
Released: 2025-09-30

MIcrosimulation SCreening ANalysis Gastric Cancer Model (MISCAN-Gastric): Model Profile

Erasmus University Medical Center

Contact

Duco Mulder (d.t.mulder@erasmusmc.nl)

Suggested Citation

Mülder DT, O'Mahony J, Sun D, van Duuren L, van den Puttelaar R, Harlass M, Han W, Huang R, Spaander MCW, Ladabaum U, Lansdorp-Vogelaar I.. MIcrosimulation SCreening ANalysis Gastric Cancer Model (MISCAN-Gastric): 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-miscan-gc-model-profile-1.0.00-2025-09-30.pdf

Version Table

Version	Date	Notes
1.0.00	2025-09-30	Initial release

Erasmus MC
Readers Guide

Reader's Guide

Core Profile Documentation

These topics will provide an overview of the model without the burden of detail. Each can be read in about 5-10 minutes. 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

The Microsimulation Screening Analysis (MISCAN) gastric model is designed to evaluate the effect of gastric cancer screening and prevention strategies. These include endoscopic screening and *H. pylori* screen-and-treat interventions. MISCAN-gastric is developed within the Early Detection & Screening programme in the Department of Public Health at the Erasmus University Medical Center in Rotterdam, the Netherlands[1].

References

1. JDF Habbema, GJ Van Oortmarssen, JTN Lubbe, PJ Van der Maas. The MISCAN simulation program for the evaluation of screening for disease. Computer methods and programs in biomedicine. 1985;20(1):79–93.

Model Overview

As MISCAN-gastric is a microsimulation model, the model simulates independent individual life histories from birth until death, rather than as proportions of a cohort. This structure is similar across all MISCAN models, such as MISCAN-colon[1]. This allows future state transitions to depend on past transitions, giving individuals a memory function. Unlike most traditional Markov models, MISCAN-gastric does not use yearly/monthly transition probabilities. Instead, in each health state, individual durations to other health states are generated. The term stochastic implies that model uses probability distributions and durations to simulate events, rather than using fixed values. The results are therefore subject to random variation. In MISCAN-gastric, some individuals develop precursor lesions which may eventually progress to cancer.

MISCAN-gastric's natural history model is based on Correa's cascade, encompassing the states of atrophic gastritis (AG), intestinal metaplasia (IM), dysplasia, and ultimately carcinoma (Main Figure 1)[2]. A distinction between limited (non-extensive) and extensive IM was incorporated to permit future assessment of surveillance strategies based on the extent of IM, which often features in clinical guidelines[3].

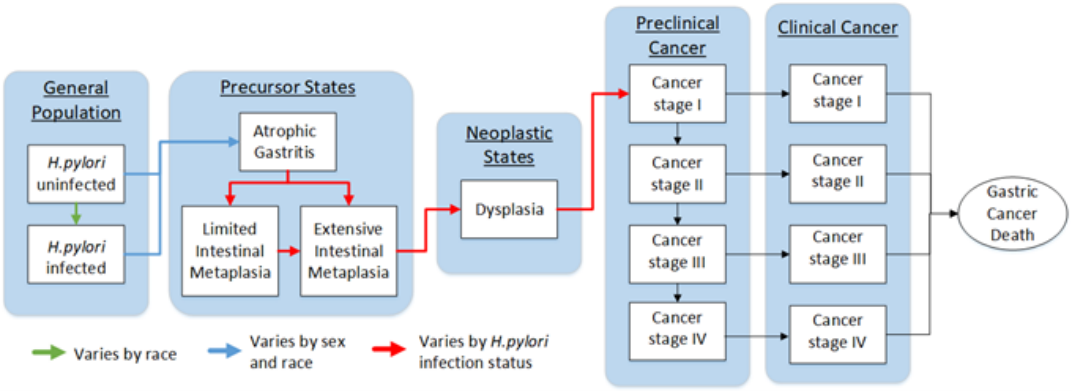

Figure 1: Natural history of the MISCAN-gastric model.

The arrows represent transitions between health states. The duration between transitions varies by race, sex and/or *H. pylori* infection status, depending on the color of the arrow.

H. pylori, *Helicobacter Pylori.*

References

1. R van den Puttelaar. Advancing Colorectal Cancer Screening: Challenges and Innovations [phdthesis]. Erasmus University; 2025.
2. P Correa. Human gastric carcinogenesis: a multistep and multifactorial process. Cancer research. 1992;52(24):6735–6740.
3. S Gupta, D Li, HB El Serag, others. AGA clinical practice guidelines on management of gastric intestinal metaplasia. Gastroenterology. 2020;158(3):693–702.

Assumption Overview

The model employs several assumptions, primarily due to insufficient data on precursor lesions in asymptomatic populations[1]. First, we assumed identical progression rates across racial groups. In contrast, the onset of precursor lesion was assumed to vary by race and sex. Second, *H. pylori* could elevate GC risk through two mechanisms: through an increased likelihood of developing precursor lesions (onset) or an accelerated progression of these lesions. The calibration decided how much each contributed. Regardless of the specific mechanism, the effect was assumed to be identical for all sex- and racial subgroups. Third, race-specific prevalence of *H. pylori* was based on estimates derived from literature and NHANES data [2]. Finally, we assumed the same stage distribution at clinical diagnosis across racial groups, consistent with SEER data.

Mathematically, the onset age of precursor lesions was based on an *H. pylori*-specific hazard and race- and sex-specific generalized logistic hazard functions (blue arrows in Figure 2). The dwell times of health states followed Weibull distributions (red and black arrows in Figure 2), similar to the approach in other cancer natural history models [3,4]. These transitions will now be further explained mathematically in the sequence of the model according to figure 2.

Atrophic gastritis onset risk

The onset age of precursor lesions was based on an *H. pylori*-specific hazard H_{Hp} and a race- and sex-specific generalized logistic hazard functions. The hazard for individuals uninfected with *H. pylori* was set to one. The logistic hazard function was rewritten for interpretability. We define a constant:

$$b_{s,r} = \frac{log(\frac{1}{v_{s,r}}(\frac{K_{s,r}}{L})^{v} - \frac{1}{v_{s,r}})}{G^{m}_{s,r} - M_{s,r}},$$

To define the onset, based on random number *x*:

$$Onset\ age_i(x) = \frac{K_{s,r}}{(1 + v_{s,r} \times exp(-b_{s,r}\,(x - M_{s,r})))^{1/v_{s,r}}}.$$

Where *L* is a small number larger than 0 (0.001).

Calibrated parameters: $K_{s,r},\ v_{s,r}, M_{s,r}, G^{m}_{s,r}$ for each sex *s* (male and female) and race *r* (black and white). x follows an exponential distribution with an *H. pylori* specific Hazard rate.

The parameters can be interpreted as follows:

- $K_{s,r}$ refers to the horizontal asymptote of the function.
- $v_{s,r}$, reflects the steepness in the inflection point.
- $M_{s,r}$ is the age of the inflection point
- $G^{m}_{s,r}$ is the age for which the function form equals L (0.001)

Atrophic gastritis progression:>

Two independent dwell times are drawn from a Weibull distribution with varying means $k_{AG,1}$ and $k_{AG,2}$ but the same scale parameter λ_{AG}:

$$Dwell\ AG_1 \sim WB(k_{AG,1} * \Omega_{AG}^{I\{Hp\}}, \lambda_{AG})$$

$$Dwell\ AG_2 \sim WB\left(k_{AG,2} * \Omega_{AG}^{I\{Hp\}}, \lambda_{AG}\right),$$

where $\Omega_{AG}^{I\{Hp\}}$ represents the effect of *Helicobacter pylori (H. pylori)* on the mean progression time. $\Omega_{AG}^{I\{Hp\}}$ is a calibrated factor between 0 and 1, depending on *H. pylori* infection.

If $AG_1 \leq AG_2$, the individual progresses to limited intestinal metaplasia (IM). Otherwise, they progress to extensive IM directly.

Calibration parameters: $k_{AG,1}$, $k_{AG,2}$, λ_{AG}, $\Omega_{AG}^{I\{Hp\}}$

IM progression:

For limited IM, a dwell time to extensive IM is drawn from a Weibull distribution:

$$Dwell\ limited\ IM\ \sim\ WB(k_{IM\ lim} *\ \Omega_{IM/dys}^{I\{Hp\}}, \lambda_{IM})$$

For extensive IM, we draw a dwell time to dysplasia with a different mean, but the same scale parameter:

$$Dwell\ extensive\ IM \sim\ WB(k_{IM\ ext} *\ \Omega_{IM/dys}^{I\{Hp\}}, \lambda_{IM})$$

The parameter $\Omega_{IM/dys}^{I\{Hp\}}$ again represents the effect of *H. pylori* on the mean dwell time. Note that this effect is the same for IM and dysplasia.

To account for the effect of age on progression of disease, these dwell times are multiplied by a factor:

$$1 +\ \rho * age$$

Where age is the midpoint between the onset age of IM and the next state.

Calibration parameters: $k_{IM\ lim}$, $k_{IM\ ext}$, λ_{IM}, $\Omega_{IM/dys}^{I\{Hp\}}$, ρ

Dysplasia progression:

For dysplasia, we draw a dwell time to preclinical cancer stage 1 from a Weibull distribution:

$$Dwell\ dysplasia \sim\ WB(k_{dys} *\ \Omega_{IM/dys}^{I\{Hp\}}, \lambda_{dys})$$

Again, we account for the effect of age by multiplying the dwell times by$\ 1 + \ \rho*age$, which is the same factor used in the dwell time of IM.

Additional calibration parameters: k_{dys}, λ_{dys}.

Preclinical cancer progression

In each preclinical cancer stage j, we draw a dwell time to preclinical stage j+1 from an exponential distribution:

$$Dwell\ cancer\ S_{j,\ j+1} \sim \exp\left(\lambda_j\right)$$

For j=1,2,3.

In each preclinical cancer stage, we also draw a dwell time until clinical detection:

$$Time\ until\ detection\ cancer\ S_{j,} \sim \exp\left(\alpha_j\right)$$

For j=1,2,3,4

Preclinical cancers progress to the next stage if:

$$Dwell\ cancer\ S_{j,\ j+1} \leq\ Time\ until\ detection\ cancer\ S_{j,}$$

Otherwise, cancers are clinically detected after the time until detection has elapsed.

Calibration parameters: λ_1, $\lambda_{2,}\lambda_3$ and α_1, α_2, α_3, α_4

Calibration targets

The model was calibrated to SEER incidence data and data from studies. An overview of the data used from clinical studies can be found in Table 1.

Table 1: Data used in the calibration of MISCAN-gastric

Data used in calibration	Value (95% CI)	Reference
Total mean sojourn time (time between onset of preclinical cancer stage I and clinical diagnosis)	3.7 (1.96-8.28 years)	(13)
OR of non-cardia GC with *H. pylori* infection	4.79 (2.39-9.60)	(14)
Prevalence of *H. pylori* at age 35	White people: 36% Black people: 60%	(15)
Overall prevalence of atrophic gastritis	2.1% (0.7-4.7%)	(16)
Overall prevalence of intestinal metaplasia	9.1% (6.9-12.0%)	(16)
Overall prevalence of dysplasia	0.2% (0.04%-1.5%)	(17)
Odds ratio of developing precursor lesions of *H. pylori*+ compared to *H. pylori*-	2.6 (1.5-3.3)	(16)
Odds ratios of intestinal metaplasia per age group	≤30 31-45 46-60 61-75 >75	Ref. 1.7 2.7 3.9 5.3
Odds ratios of developing precursor lesions for males compared to females	1.04	(18)
Proportion of extensive cases of all intestinal metaplasia cases	28%	(19)
Relative risk of intestinal metaplasia progression to subsequent precursors after *H. pylori* eradication	0.8	(20)
Relative risk of intestinal metaplasia progression to cancer following *H. pylori* eradication	0.7	(21)
Relative risk of atrophic gastritis progression to cancer following *H. pylori* eradication	0.28	(21)

References

1. RJ Huang, AR Ende, A Singla, others. Prevalence, risk factors, and surveillance patterns for gastric intestinal metaplasia among patients undergoing upper endoscopy with biopsy. Gastrointestinal endoscopy. 2020;91(1):70-77.e71.
2. RM Genta, A Sonnenberg. Characteristics of the gastric mucosa in patients with intestinal metaplasia. The American Journal of Surgical Pathology. 2015;39(5):700–704.
3. IMCM de Kok, J van Rosmalen, M van Ballegooijen. Description of MISCAN-cervix.
4. A. van der Steen, J. van Rosmalen, S. Kroep, others. Calibrating parameters for microsimulation disease models: a review and comparison of different goodness-of-fit criteria. Medical Decision Making. 2016;36(5):652–665.

Parameter Overview

An overview of the calibrated model parameters can be found in table 2.

Table 2: Calibrated values of the model parameters of MISCAN-gastric

Parameters	Interpretation	Calibrated Value
H_{Hp}	Hazard rate of developing atrophic gastritis for *H. pylori* infected individuals.	3.651151191
$K_{s,r}$	Horizontal asymptote of the onset age function	Black males: 0.135548 White males: 0.062793 Black females: 0.080278 White females: 0.083563
$v_{s,r}$	Inflection point of the onset age function	Black males: 19.54042 White males: 1.769942 Black females: 1.06667 White females: 1.242386
$M_{s,r}$	Age of the inflection point of the onset age function	Black males: 44.65589 White males: 46.55290 Black females: 52.38103 White females: 61.61536
$G^{m}_{s,r}$	Age for which the onset age function equals 0.001	Black males: 0.833748 White males: 8.716735 Black females: 1.437170 White females: 0.833748
$k_{AG,1}$	Weibull mean dwell time for atrophic gastritis to limited IM	23.43013009
$k_{AG,2}$	Weibull mean dwell time for atrophic gastritis to extensive IM	49.89100247
$\Omega_{AG}^{I\{Hp\}}$	Effect of *H. pylori* infection on progression time of atrophic gastritis	1
λ_{AG}	Scale of the Weibull distributions of atrophic gastritis dwell time	0.571882689
k_{IMlim}	Weibull mean dwell time for limited IM to extensive IM	74.07276355
$\Omega_{IM/dys}^{I\{Hp\}}$	Effect of *H. pylori* infection on the dwell time of IM and dysplasia	0.960589394
λ_{IM}	Scale of the Weibull distributions of IM dwell time	1.226024
k_{IMext}	Weibull mean dwell time for extensive IM to dysplasia	37.00643102
ρ	Parameter to model the effect of age on dwell time	7.980778226
k_{dys}	Weibull mean dwell time for dysplasia to preclinical cancer stage 1	4.388887775

Parameters	Interpretation	Calibrated Value
λ_{dys}	Scale of the Weibull distribution of dysplasia dwell time	0.661216
λ_j	Exponential distribution mean dwell time for preclinical cancer stage j to j+1	j=1: 3.297144 j=2: 0.794827 j=3: 1.137446
α_j	Exponential distribution mean dwell time for preclinical cancer to become clinically detected in stage j	j=1: 8.087736 j=2: 3.714227 j=3: 4.306321 j=4: 0.428254
$T_{Hp,AG}$	Multiplication factor of remaining atrophic gastritis dwell time after *H. pylori* eradication	3.554993
$T_{Hp,IM/dys}$	Multiplication factor of remaining IM and dysplasia dwell time after *H. pylori* eradication	1.125454

Component Overview

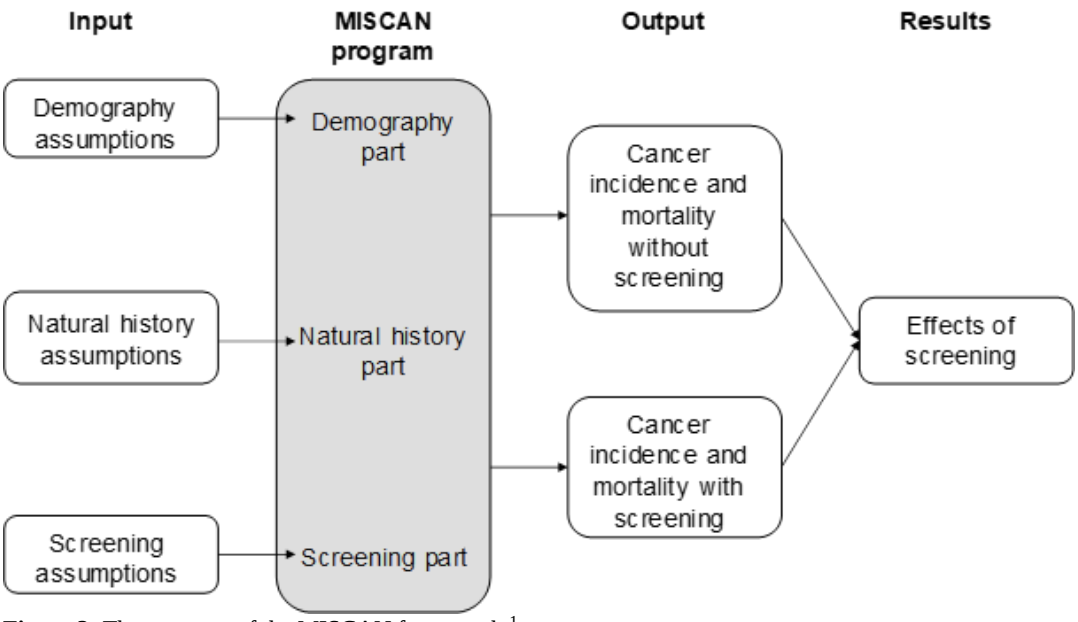

Figure 2: The structure of the MISCAN framework [1]

The MISCAN-framework consists of three modules: demography, natural history and screening (Figure 1)[1]. This framework has been extensively validated and applied for guiding cancer policy for other types of cancer in various contexts [2-4]. In the demography part, simulated individuals are born and die according to the population characteristics. The natural history part determines how many people develop precursor lesions and what proportion progresses to cancer. The screening part contains information about the screening test, such as the sensitivity and the effect of treatment. MISCAN models can be run with and without screening and prevention strategies. Comparing these scenarios can then inform the extent to which screening affects disease outcomes, such as incidence and mortality.

References

1. JDF Habbema, GJ Van Oortmarssen, JTN Lubbe, PJ Van der Maas. The MISCAN simulation program for the evaluation of screening for disease. Computer methods and programs in biomedicine. 1985;20(1):79–93.
2. F Van Hees, AG Zauber, H Van Veldhuizen, others. The value of models in informing resource allocation in colorectal cancer screening: the case of the Netherlands. Gut. 2015;64(12):1985–1997.
3. A Irzaldy, R Gvamichava, T Beruchashvili, others. Breast Cancer Screening in Georgia: Choosing the Most Optimal and Cost-Effective Strategy. Value in Health Regional Issues. 2024;39:66–73.
4. EAM Heijnsdijk, EM Wever, A Auvinen, others. Quality-of-life effects of prostate-specific antigen screening. New England Journal of Medicine. 2012;367(7):595–605.

Output Overview

The outputs that can be generated by MISCAN-gastric include:

- Incidence counts of each disease by calendar year
- Mean prevalence of each disease state in five year age groups
- Number of invitations for screen tests
- Number of positive/negative tests
- Number of specific deaths and non-specific deaths
- Total number of life years and life years lost due to gastric cancer
- Number of life-years gained due to screening

Results Overview

Estimation of the number needed to screen (NNS) to prevent one death by race

One-time endoscopic screening with subsequent surveillance demonstrated optimal efficacy when initiated at age 50. The NNS for the overall US population was 3506 to prevent a NI-GC death. However, the NNS for the non-Hispanic black male population was as low as 621 (Figure 3).

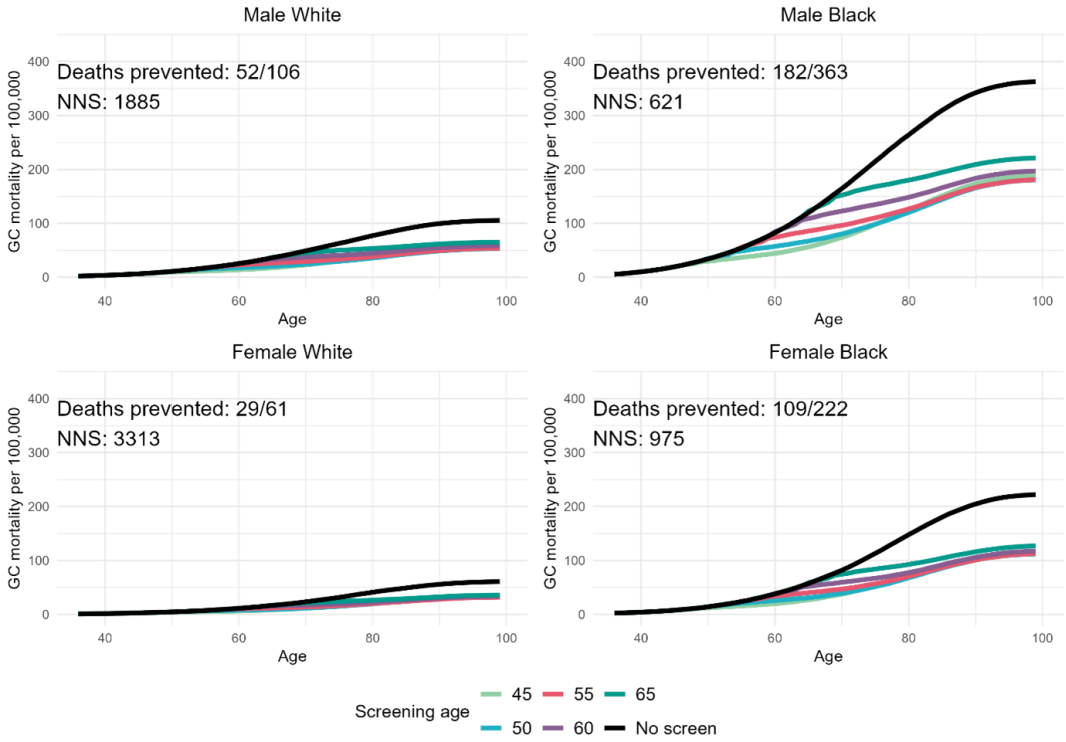

Figure 3: Effect of endoscopic screening on GC mortality

Optimal age of *H. pylori* screen-and-treat

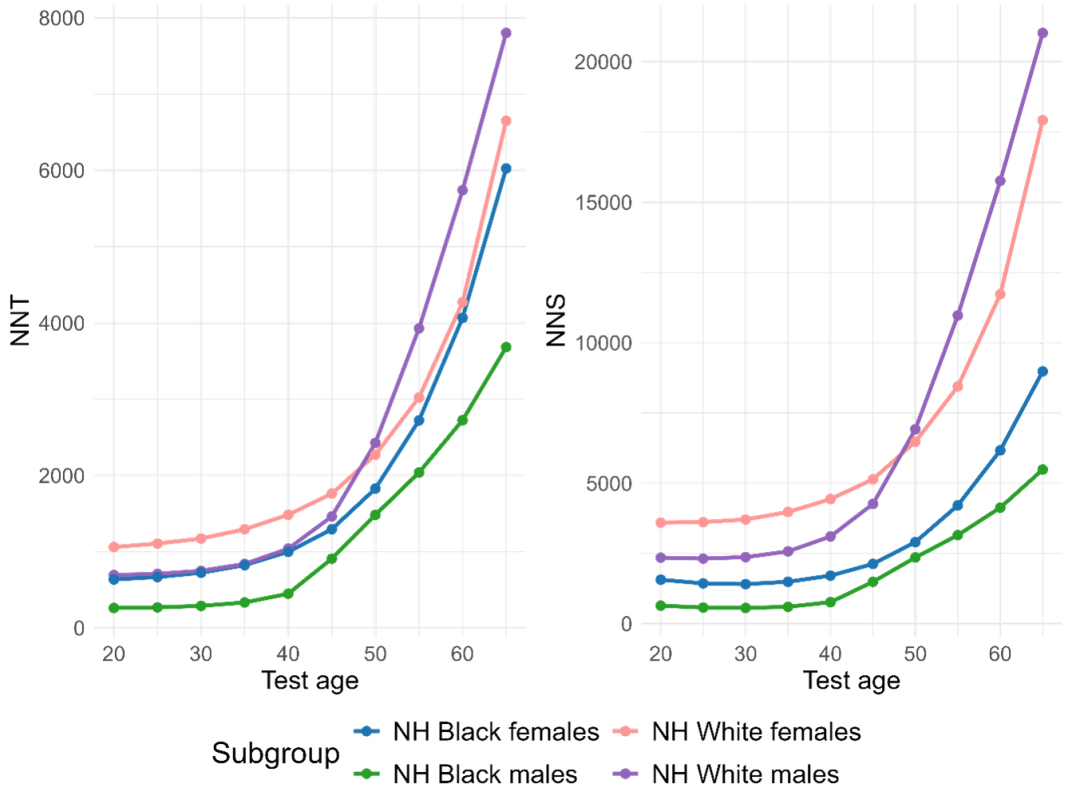

Figure 4: The Number Needed to Treat (NNT) and Number Needed to Screen (NNS) for *Helicobacter pylori* to prevent one case of gastric cancer. NH, non-Hispanic.

Key References

P Correa. Human gastric carcinogenesis: a multistep and multifactorial process. Cancer research. 1992;52(24):6735–6740.

RM Genta, A Sonnenberg. Characteristics of the gastric mucosa in patients with intestinal metaplasia. The American Journal of Surgical Pathology. 2015;39(5):700–704.

S Gupta, D Li, HB El Serag, others. AGA clinical practice guidelines on management of gastric intestinal metaplasia. Gastroenterology. 2020;158(3):693–702.

JDF Habbema, GJ Van Oortmarssen, JTN Lubbe, PJ Van der Maas. The MISCAN simulation program for the evaluation of screening for disease. Computer methods and programs in biomedicine. 1985;20(1):79–93.

EAM Heijnsdijk, EM Wever, A Auvinen, others. Quality-of-life effects of prostate-specific antigen screening. New England Journal of Medicine. 2012;367(7):595–605.

RJ Huang, AR Ende, A Singla, others. Prevalence, risk factors, and surveillance patterns for gastric intestinal metaplasia among patients undergoing upper endoscopy with biopsy. Gastrointestinal endoscopy. 2020;91(1):70-77.e71.

A Irzaldy, R Gvamichava, T Beruchashvili, others. Breast Cancer Screening in Georgia: Choosing the Most Optimal and Cost-Effective Strategy. Value in Health Regional Issues. 2024;39:66–73.

IMCM de Kok, J van Rosmalen, M van Ballegooijen. Description of MISCAN-cervix.

R van den Puttelaar. Advancing Colorectal Cancer Screening: Challenges and Innovations [phdthesis]. Erasmus University; 2025.

A. van der Steen, J. van Rosmalen, S. Kroep, others. Calibrating parameters for microsimulation disease models: a review and comparison of different goodness-of-fit criteria. Medical Decision Making. 2016;36(5):652–665.

F Van Hees, AG Zauber, H Van Veldhuizen, others. The value of models in informing resource allocation in colorectal cancer screening: the case of the Netherlands. Gut. 2015;64(12):1985–1997.