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.

*Harvard*
Model Overview



# 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