Gastric Cancer Mortality Reduction Calculator

Based on Equation 1 algorithm for projecting reduction in non-cardia gastric adenocarcinoma mortality

# Load required libraries

library(dplyr)

library(readr)

#' Calculate Reduction in Gastric Cancer Mortality

#'

#' This function implements Equation 1 to calculate the potential reduction in

#' gastric cancer mortality from screening and H. pylori eradication interventions

#'

#' @param data\_file Path to CSV file containing baseline parameters (default: "gastric\_cancer\_parameters.csv")

#' @param gastric\_cancer\_type Type of gastric cancer to analyze: "total" or "non-cardia adenocarcinoma"

#' @param racial\_ethnic\_groups Vector of racial/ethnic groups to analyze (default: all groups)

#' @param age\_groups Vector of age groups to analyze (default: all groups)

#' @param chi\_hp Compliance with H. pylori eradication therapy (0-1, default: varies by user)

#' @param chi\_screen Compliance with gastric screening (0-1, default: varies by user)

#' @param p\_hp\_effectiveness Effectiveness of H. pylori eradication therapy (0-1, default: 0.9)

#' @param r\_screen0 Relative risk reduction from screening with full compliance (default: 0.6)

#' @param r\_hp0 Relative risk reduction from H. pylori treatment with full compliance (default: 0.6)

#'

#' @return Data frame with columns:

#' - racial\_ethnic\_group: Racial/ethnic group

#' - age\_group: Age group

#' - cancer\_type: Type of gastric cancer analyzed

#' - baseline\_mortality\_per\_100k: Baseline projected mortality 2030-2034

#' - mortality\_reduction\_per\_100k: Reduction in mortality per 100,000 (Column AC equivalent)

#' - percent\_reduction: Percentage reduction in mortality (Column AD equivalent)

#'

#' @examples

#' # Analyze total gastric cancer

#' results\_total <- calculate\_gastric\_mortality\_reduction(

#' gastric\_cancer\_type = "total",

#' chi\_hp = 0.8, # 80% compliance with H. pylori treatment

#' chi\_screen = 0.7, # 70% compliance with screening

#' p\_hp\_effectiveness = 0.9 # 90% effectiveness of H. pylori treatment

#' )

#'

#' # Analyze non-cardia adenocarcinoma only

#' results\_noncardia <- calculate\_gastric\_mortality\_reduction(

#' gastric\_cancer\_type = "non-cardia adenocarcinoma",

#' chi\_hp = 0.8,

#' chi\_screen = 0.7

#' )

calculate\_gastric\_mortality\_reduction <- function(

data\_file = "gastric\_cancer\_parameters.csv",

gastric\_cancer\_type = c("total", "non-cardia adenocarcinoma"),

racial\_ethnic\_groups = NULL,

age\_groups = NULL,

chi\_hp = NULL, # User-defined: Compliance with H. pylori treatment (Cell AF4)

chi\_screen = NULL, # User-defined: Compliance with screening (Cell AG4)

p\_hp\_effectiveness = 0.9, # User-defined: Effectiveness of H. pylori therapy (Cell AI4)

r\_screen0 = 0.6, # Default: relative risk with full screening compliance

r\_hp0 = 0.6 # Default: relative risk with full H. pylori compliance

) {

# ============================================================================

# STEP 1: VALIDATE CANCER TYPE SELECTION

# ============================================================================

# Match the gastric\_cancer\_type argument

gastric\_cancer\_type <- match.arg(gastric\_cancer\_type)

# Determine which column prefixes to use based on cancer type

if (gastric\_cancer\_type == "total") {

# Use columns with "total\_gc\_" prefix for total gastric cancer

incidence\_col\_2015\_2019 <- "total\_gc\_incidence\_2015\_2019"

incidence\_col\_2030\_2034 <- "total\_gc\_incidence\_2030\_2034"

ibmortality\_col\_2015\_2019 <- "total\_gc\_ibmortality\_2015\_2019"

mortality\_col\_2015\_2019 <- "total\_gc\_mortality\_2015\_2019"

mortality\_col\_2030\_2034 <- "total\_gc\_projected\_mortality\_2030\_2034"

cat("Analyzing TOTAL gastric cancer mortality\n")

} else if (gastric\_cancer\_type == "non-cardia adenocarcinoma") {

# Use columns with "noncardia\_adeno\_gc\_" prefix for non-cardia adenocarcinoma

incidence\_col\_2015\_2019 <- "noncardia\_adeno\_gc\_incidence\_2015\_2019"

incidence\_col\_2030\_2034 <- "noncardia\_adeno\_gc\_incidence\_2030\_2034"

ibmortality\_col\_2015\_2019 <- "noncardia\_adeno\_gc\_ibmortality\_2015\_2019"

mortality\_col\_2030\_2034 <- "noncardia\_adeno\_gc\_projected\_mortality\_2030\_2034"

# Note: There's no noncardia\_adeno\_gc\_mortality\_2015\_2019 in the column list,

# so we'll use total if needed or skip if not essential

mortality\_col\_2015\_2019 <- NULL # Set to NULL if this column doesn't exist

cat("Analyzing NON-CARDIA ADENOCARCINOMA gastric cancer mortality\n")

}

# ============================================================================

# STEP 2: LOAD AND PREPARE DATA

# ============================================================================

# Read the main parameter file

# This corresponds to the Excel sheet with the gastric cancer parameters

df <- read\_csv(data\_file, show\_col\_types = FALSE)

# Select and rename the relevant columns based on cancer type

# Core columns that are always needed

df\_selected <- df %>%

select(

racial\_ethnic\_group,

age\_group,

gamma, # Column L equivalent - proportion attributable to IM

f11, # Column U equivalent - low-risk IM, H. pylori negative

f12, # Column V equivalent - high-risk IM, H. pylori negative

f21, # Column W equivalent - low-risk IM, H. pylori positive

f22, # Column X equivalent - high-risk IM, H. pylori positive

rr\_hp\_erad, # If these are used for validation or reference

rr\_screen, # If these are used for validation or reference

# Select the appropriate mortality column based on cancer type

baseline\_mortality = all\_of(mortality\_col\_2030\_2034)

)

# Filter out rows with missing essential data

# Some age groups may not have data for certain parameters

df\_selected <- df\_selected %>%

filter(

!is.na(gamma),

!is.na(f11),

!is.na(f12),

!is.na(f21),

!is.na(f22),

!is.na(baseline\_mortality)

)

# Filter for specified groups if provided

if (!is.null(racial\_ethnic\_groups)) {

df\_selected <- df\_selected %>%

filter(racial\_ethnic\_group %in% racial\_ethnic\_groups)

}

if (!is.null(age\_groups)) {

df\_selected <- df\_selected %>%

filter(age\_group %in% age\_groups)

}

# ============================================================================

# STEP 3: VALIDATE USER INPUTS

# ============================================================================

# Check compliance parameters are between 0 and 1

if (!is.null(chi\_hp)) {

if (chi\_hp < 0 || chi\_hp > 1) {

stop("chi\_hp (H. pylori compliance) must be between 0 and 1")

}

} else {

stop("chi\_hp (H. pylori compliance) must be specified")

}

if (!is.null(chi\_screen)) {

if (chi\_screen < 0 || chi\_screen > 1) {

stop("chi\_screen (screening compliance) must be between 0 and 1")

}

} else {

stop("chi\_screen (screening compliance) must be specified")

}

if (p\_hp\_effectiveness < 0 || p\_hp\_effectiveness > 1) {

stop("p\_hp\_effectiveness must be between 0 and 1")

}

# ============================================================================

# STEP 4: CALCULATE INTERVENTION EFFECTS

# ============================================================================

# Calculate r\_screen: the actual risk reduction from screening given compliance

# This corresponds to the formula: r\_screen = χ\_screen × r\_screen0 + (1 - χ\_screen)

# When compliance is 0, r\_screen = 1 (no effect)

# When compliance is 1, r\_screen = r\_screen0 (full effect)

r\_screen <- chi\_screen \* r\_screen0 + (1 - chi\_screen)

# Calculate r\_hp: the actual risk reduction from H. pylori treatment given compliance

# This corresponds to the formula: r\_hp = χ\_hp × r\_hp0 + (1 - χ\_hp)

r\_hp <- chi\_hp \* r\_hp0 + (1 - chi\_hp)

# ============================================================================

# STEP 5: APPLY EQUATION 1 TO CALCULATE MORTALITY WITH INTERVENTIONS

# ============================================================================

df\_results <- df\_selected %>%

mutate(

# Store the cancer type being analyzed

cancer\_type = gastric\_cancer\_type,

# Calculate the effect on IM-attributable gastric cancer

# This is the first part of the equation: γ × [f11 + f12×r\_screen + f21×r\_hp + f22×r\_screen×r\_hp]

im\_attributable\_effect = gamma \*

(

f11 + # Low-risk IM, H. pylori negative (no intervention effect)

f12 \* r\_screen + # High-risk IM, H. pylori negative (screening effect only)

f21 \* r\_hp + # Low-risk IM, H. pylori positive (H. pylori treatment effect only)

f22 \* r\_screen \* r\_hp # High-risk IM, H. pylori positive (both interventions)

),

# Calculate the effect on non-IM-attributable gastric cancer

# This is the second part: (1 - γ) × p\_hp\_effectiveness × r\_hp

# Note: p\_hp\_effectiveness represents (p\_Hp+|IM-) in the equation

non\_im\_attributable\_effect = (1 - gamma) \* p\_hp\_effectiveness \* r\_hp,

# Total effect combining both IM and non-IM pathways

total\_intervention\_effect = im\_attributable\_effect + non\_im\_attributable\_effect,

# Calculate mortality with interventions

# The baseline mortality is multiplied by the intervention effect

mortality\_with\_intervention = baseline\_mortality \* total\_intervention\_effect,

# Calculate the reduction in mortality (Column AC equivalent)

# This is the absolute reduction per 100,000 population

mortality\_reduction\_per\_100k = baseline\_mortality - mortality\_with\_intervention,

# Calculate the percentage reduction (Column AD equivalent)

# This is the relative reduction as a percentage

percent\_reduction = (mortality\_reduction\_per\_100k / baseline\_mortality) \* 100

)

# ============================================================================

# STEP 6: PREPARE AND RETURN OUTPUT

# ============================================================================

# Select and reorder columns - parameters first, then results

results <- df\_results %>%

mutate(

# Add the intervention parameters as columns

chi\_hp = chi\_hp,

chi\_screen = chi\_screen,

p\_hp\_effectiveness = p\_hp\_effectiveness

) %>%

select(

# Identification columns

racial\_ethnic\_group,

age\_group,

cancer\_type,

# Baseline value

baseline\_mortality\_per\_100k = baseline\_mortality,

# Intervention parameters

chi\_hp,

chi\_screen,

p\_hp\_effectiveness,

total\_intervention\_effect,

# Results columns at the end (rightmost)

mortality\_reduction\_per\_100k,

percent\_reduction

)

# Add intervention parameters to results as attributes for reference

attr(results, "intervention\_parameters") <- list(

gastric\_cancer\_type = gastric\_cancer\_type,

chi\_hp = chi\_hp,

chi\_screen = chi\_screen,

p\_hp\_effectiveness = p\_hp\_effectiveness,

r\_screen0 = r\_screen0,

r\_hp0 = r\_hp0,

r\_screen\_actual = r\_screen,

r\_hp\_actual = r\_hp

)

# Print summary

cat("\n=== Gastric Cancer Mortality Reduction Analysis ===\n")

cat(sprintf("Cancer Type: %s\n", gastric\_cancer\_type))

cat(sprintf("H. pylori treatment compliance: %.1f%%\n", chi\_hp \* 100))

cat(sprintf("Screening compliance: %.1f%%\n", chi\_screen \* 100))

cat(sprintf("H. pylori treatment effectiveness: %.1f%%\n", p\_hp\_effectiveness \* 100))

cat(sprintf("Actual risk reduction from screening: %.3f\n", r\_screen))

cat(sprintf("Actual risk reduction from H. pylori treatment: %.3f\n", r\_hp))

cat(sprintf("Number of groups analyzed: %d\n", nrow(results)))

# Check for calculation issues

na\_count <- sum(is.na(results$mortality\_reduction\_per\_100k))

if (na\_count > 0) {

cat(sprintf("WARNING: %d rows have NA values in calculations\n", na\_count))

cat("This may be due to missing data in the input file.\n")

}

cat("\n")

return(results)

}

# ============================================================================

# EXAMPLE USAGE

# ============================================================================

# Example 1: Analyze total gastric cancer with specific compliance rates

# results\_total <- calculate\_gastric\_mortality\_reduction(

# gastric\_cancer\_type = "total",

# chi\_hp = 0.8, # 80% compliance with H. pylori treatment

# chi\_screen = 0.7, # 70% compliance with screening

# p\_hp\_effectiveness = 0.9 # 90% effectiveness (default)

# )

# print(results\_total)

# Example 2: Analyze non-cardia adenocarcinoma for specific groups

# results\_noncardia <- calculate\_gastric\_mortality\_reduction(

# gastric\_cancer\_type = "non-cardia adenocarcinoma",

# racial\_ethnic\_groups = c("Hispanic", "Non-Hispanic Asian American/Pacific Islander"),

# age\_groups = c("60-64", "65-69", "70-74"),

# chi\_hp = 0.75,

# chi\_screen = 0.65,

# p\_hp\_effectiveness = 0.85

# )

# print(results\_noncardia)

# Example 3: Compare both cancer types

# compare\_cancer\_types <- function(chi\_hp, chi\_screen) {

# # Run for total gastric cancer

# results\_total <- calculate\_gastric\_mortality\_reduction(

# gastric\_cancer\_type = "total",

# chi\_hp = chi\_hp,

# chi\_screen = chi\_screen

# )

#

# # Run for non-cardia adenocarcinoma

# results\_noncardia <- calculate\_gastric\_mortality\_reduction(

# gastric\_cancer\_type = "non-cardia adenocarcinoma",

# chi\_hp = chi\_hp,

# chi\_screen = chi\_screen

# )

#

# # Combine results

# combined <- bind\_rows(results\_total, results\_noncardia)

#

# # Summary comparison

# summary\_comparison <- combined %>%

# group\_by(cancer\_type) %>%

# summarise(

# mean\_baseline\_mortality = mean(baseline\_mortality\_per\_100k, na.rm = TRUE),

# mean\_reduction = mean(mortality\_reduction\_per\_100k, na.rm = TRUE),

# mean\_percent\_reduction = mean(percent\_reduction, na.rm = TRUE)

# )

#

# return(list(

# detailed\_results = combined,

# summary = summary\_comparison

# ))

# }

# Example 4: Scenario analysis for different compliance levels

# scenario\_analysis <- function(gastric\_cancer\_type = "non-cardia adenocarcinoma") {

# scenarios <- list(

# low\_compliance = list(chi\_hp = 0.3, chi\_screen = 0.2),

# medium\_compliance = list(chi\_hp = 0.6, chi\_screen = 0.5),

# high\_compliance = list(chi\_hp = 0.9, chi\_screen = 0.8)

# )

#

# results\_list <- lapply(names(scenarios), function(scenario\_name) {

# s <- scenarios[[scenario\_name]]

# res <- calculate\_gastric\_mortality\_reduction(

# gastric\_cancer\_type = gastric\_cancer\_type,

# chi\_hp = s$chi\_hp,

# chi\_screen = s$chi\_screen

# )

# res$scenario <- scenario\_name

# return(res)

# })

#

# combined\_results <- bind\_rows(results\_list)

# return(combined\_results)

# }

# ============================================================================

# UTILITY FUNCTIONS

# ============================================================================

#' Summarize results by racial/ethnic group

#'

#' @param results Output from calculate\_gastric\_mortality\_reduction

#' @return Summary statistics by group

summarize\_by\_group <- function(results) {

summary <- results %>%

group\_by(racial\_ethnic\_group, cancer\_type) %>%

summarise(

mean\_baseline\_mortality = mean(baseline\_mortality\_per\_100k, na.rm = TRUE),

mean\_reduction = mean(mortality\_reduction\_per\_100k, na.rm = TRUE),

mean\_percent\_reduction = mean(percent\_reduction, na.rm = TRUE),

total\_lives\_saved\_per\_100k = sum(mortality\_reduction\_per\_100k, na.rm = TRUE),

.groups = "drop"

)

return(summary)

}

#' Summarize results by age group

#'

#' @param results Output from calculate\_gastric\_mortality\_reduction

#' @return Summary statistics by age

summarize\_by\_age <- function(results) {

summary <- results %>%

group\_by(age\_group, cancer\_type) %>%

summarise(

mean\_baseline\_mortality = mean(baseline\_mortality\_per\_100k, na.rm = TRUE),

mean\_reduction = mean(mortality\_reduction\_per\_100k, na.rm = TRUE),

mean\_percent\_reduction = mean(percent\_reduction, na.rm = TRUE),

.groups = "drop"

)

return(summary)

}

#' Plot intervention effects

#'

#' @param results Output from calculate\_gastric\_mortality\_reduction

#' @return ggplot object

plot\_intervention\_effects <- function(results) {

library(ggplot2)

# Create subtitle based on cancer type(s) in the data

cancer\_types <- unique(results$cancer\_type)

subtitle\_text <- paste("Cancer Type:", paste(cancer\_types, collapse = " & "))

p <- ggplot(results, aes(x = age\_group, y = percent\_reduction,

fill = racial\_ethnic\_group)) +

geom\_bar(stat = "identity", position = "dodge") +

theme\_minimal() +

labs(

title = "Percentage Reduction in Gastric Cancer Mortality (2030-2034)",

subtitle = subtitle\_text,

x = "Age Group",

y = "Mortality Reduction (%)",

fill = "Race/Ethnicity"

) +

theme(axis.text.x = element\_text(angle = 45, hjust = 1))

# Add faceting if multiple cancer types

if (length(cancer\_types) > 1) {

p <- p + facet\_wrap(~ cancer\_type, ncol = 1)

}

return(p)

}

#' Export results to Excel-friendly format

#'

#' @param results Output from calculate\_gastric\_mortality\_reduction

#' @param filename Name of output file

export\_results <- function(results, filename = "mortality\_reduction\_results.csv") {

# Get intervention parameters

params <- attr(results, "intervention\_parameters")

# Add parameters as columns for easy reference

results\_export <- results %>%

mutate(

chi\_hp = params$chi\_hp,

chi\_screen = params$chi\_screen,

p\_hp\_effectiveness = params$p\_hp\_effectiveness

)

# Write to CSV

write\_csv(results\_export, filename)

cat(sprintf("Results exported to %s\n", filename))

return(invisible(results\_export))

}