

RICE-MDA (TSCE)

FLEXKB DOCUMENT Version: HI.001.01132012.83681 Document generated: 01/13/2012

Important note: This document will remain archived as a technical appendix for publications. New versions will be added periodically as model refinements and updates are completed. The most current version is available at http://cisnet.cancer.gov/profiles. Note that unlike most PDF documents, the CISNET model profiles are not suitable for printing as they are not typically written or read in sequential fashion.



Readers Guide We recommend you let your interests guide you through this document, using the Model Overview navigation tree as a general guide to the content available.

Assumption Overview

Parameter Overview The intent of this document is to provide the interested reader with insight into Component Overview ongoing research. Model parameters, structure, and results contained herein Output Overview should be considered representative but preliminary in nature.

Results Overview

Key References We encourage interested readers to contact the contributors for further information.

Go directly to the: Reader's Guide.



READERS 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.

- Smoking History Generator Component
- Survival Mortality Component

Output Overview

Definitons and methodologies for the basic model outputs.

Results Overview

A guide to the results obtained from the model.

Key References

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

Readers Guide Model Overview Assumption Overview Parameter Overview Component Overview Output Overview Results Overview Key References



MODEL PURPOSE

SUMMARY

This document describes the primary purpose of the model.

RICE[®] MDAnderson **Readers Guide** Model Overview

PURPOSE

The Rice-MD Anderson model was formulated to model lung cancer mortality in a population in the absence of screening programs. The general objective is to make risk Assumption Overview predictions for individuals based on their unique smoking histories. The Rice-MDA Parameter Overview model is further used to simulate lung cancer mortality in individuals based on their **Component Overview** given smoking histories.

Output Overview Results Overview Key References

The goal of this project is to use carcinogenesis modeling, specifically the two-stage clonal expansion (TSCE) model estimate the effects of different risk factors on the development of lung cancer and use this model to make risk predictions for individuals. Since the TSCE model is incidence based, it is normally fit to prospective cohort data. For this study, cohort data is unavailable but case-control data on risk factor exposure and tabled age-specific mortality rates are available. For the males the model is fit using least square methods while for females a re-sampling based maximum likelihood method is used.

The main limitation of this model is that it predicts lung cancer mortality directly without including incidence or tumor growth or development. Although it can predict an individuals risk of lung cancer death and can simulate an age at death, it cannot provide information about the age when the lung cancer was diagnosed, or the hystology/stage of the lung tumor.

CATEGORIES **Core Docs**



MODEL OVERVIEW

SUMMARY

This document describes the the underlying Rice-MDA model for use in the prediction of lung cancer risk and how the model is used to simulate lung cancer mortality in the U.S. population. Further details about data sources, model fitting and parameter estimates can be found in the Parameter Overview section.

Readers Guide Model Overview Assumption Overview Parameter Overview Component Overview Output Overview Results Overview Key References

PURPOSE

W The purpose of the Rice-MDA model is to create a model for lung cancer mortality that
 W is based on individual risk histories. The models are fit to data on risk factors collected
 W in a case-control study combined with mortality rate data from prospective cohort
 S studies. These models are then used to simulate lung cancer mortality for use in the smoking base case project.

BACKGROUND

Lung cancer is the second leading cancer in terms of incidence for both men and women, second to prostate cancer for men and breast cancer for women. However, because of its serious health implications, lung cancer is the leading cancer killer for both men and women worldwide². Smoking accounts for 90% of lung cancer cases³. This modeling effort makes use of data from a lung cancer case-control study being conducted at MD Anderson Cancer Center. Using data on risk factors from this case-control study, we created a time-to-event risk prediction model.

MODEL DESCRIPTION

A two-stage clonal expansion (TSCE) model is used to predict lung cancer risk in individuals based on his/her unique smoking history and age. Moolgavkar et al.⁴ established a two-stage clonal expansion (TSCE) model. This model is depicted as follows:



The TSCE model assumes that a normal cell (NC) mutates into an initiated cell (IC) in the first transition, according to a Poisson process with intensity $\nu(t)$ where t denotes the age. There are χ normal cells in the tissue at birth or maturity, depending on the tissue. Then the initiated cell duplicates or dies according to a birth-death process with parameters $\alpha(t)$, and $\beta(t)$ and forms a clone of initiated cells. Each initiated cell can also



mutate into a malignant cell (MC) for the second transition according to a Poisson process with parameter $\mu(t)$. After some lag-time, this malignant cell is assumed to develop into a cancerous tumor with probability one. Smoking is related to the parameters of the TSCE model through the use of response functions. For piece-wise constant parameters, the exact formulas for the hazard and survival functions of the TSCE model were derived by Heidenreich in 1997⁵. More details on the assumptions of the model can be found in Assumption Overview.

The TSCE model is normally fit to prospective cohort data. The fitting routine was augmented to allow for fitting the model to data that come from an MD Anderson case-control study. Details on the data sources, model fitting, and parameter estimates can be found in Parameter Overview.

The resulting TSCE models are then used to simulate lung cancer mortality in the US population by simulating individuals using the method described in Component Overview. The smoking history generator is used to simulate individuals with complete smoking histories and death of any other cause times. These individuals are then inputted into the model to simulate lung cancer mortality. 50,000 individuals are simulated per birth cohort 1891-1970. Then the age distribution by calendar year is adjusted to match the US population using re-weighting.

CATEGORIES

Core Docs

REFERENCES:

- ¹ Coleman, M.P., Esteve, J., Demieka, P., Arslan, A., Renard, H., "Trends in cancer incidence and mortality" in International Agency for Research on Cancer 1993;
- ² NIH "What you need to know about lung cancer" in Publication No. 07-1553 2007;
- ³ Alberg, Anthony J., Samet, Jonathan M. "Epidemiology of Lung Cancer" in Chest 2003; 123: 1: 21S-49S
- ⁴ Moolgavkar, S.H., Venzon, D.J., "Two-event models for carcinogenesis: Incidence curves for childhood and adult tumours" in Mathematical Biosciences 1979; 47: : 55-77
- ⁵ Heidenreich, W.F., Jacob, P., Paretzke, H.G., "Exact Solution of the clonal expansion model and their application to the incidence of solid tumors of atomic bomb survivors" in Radiat Environ Biophys 1997; 36: : 45-58



ASSUMPTION OVERVIEW

SUMMARY

A description of the assumptions of involved in the Rice-MDA model are listed in this document.

RICE MDAnderson Cancer Center ASSUMPTION LISTING

Readers Guide Model Overview Assumption Overview Assumption Overview Nore details on the model can be found in Model Parameter Overview

Component Overview Output Overview Results Overview Key References



• There are *X* Normal Cells in the tissue at maturity that can mutate into intermediate cells.

• Intermediate Cells can either duplicate, die off, or further mutate into Malignant Cells.

• Once a Malignant Cell arises, cancer will develop after some-lag-time with probability 1.

Model Calibration Assumptions: More details on data sources and model calibration can be found in Parameter Overview.

Identifiability

One deficiency of the model is that only 4k - 1 of the 4k biological parameters ν, μ, α and β are identifiable when fitting to data in the piecewise-constant parameters over kdistinct time intervals. This issue is dealt with by setting the background mutations rates equal to each other, $\nu_0 = \mu_0$, and assuming a likely number of normal cells such as, $X = 10^7$. This approach makes use of the fact that the only the product ($\nu\mu$) appears in the survival and hazard functions. So, using this assumption will not affect estimates of incidence rates and risk.



• Lag-time between appearance of the first malignant cell and lung cancer In order to simplify the model and reduce the number of estimated parameters, a lag time of zero was assumed between the appearance of the first malignant cell and death of lung cancer as done in Deng et al¹ and Luebeck et al². This is assumption is justifiable since the TSCE model is insensitive to lag-time assumptions², ie the parameters will adjust based on assumptions about the lag-time. In other words, overall risk predictions will not be different for models calibrated to the same data but with different assumed lag-times.

• The Resampling based approach that is used to calibrate the model for females assumes that given the matching stratum from a case-control study, cases and controls are randomly sampled from the underlying population.

Simulation Based Assumptions: Details about the simulation routine can be found in Component Overview

• The Smoking History generates accurate smoking histories and death of other cause times for individuals based on the inputs of race, gender, and birth year.

• Simulating 50,000 individuals per birth cohort 1891-1970, and then re-weighting then scaling the population to match the age by calendar year distribution in the U.S. population can accurately reflect the U.S. population.

CATEGORIES

Core Docs

REFERENCES:

¹ Deng, L., Kimmel, M., Foy, M., Spitz, M., Wei, Q., Gorlova, O. "Estimation of the effects of smoking and DNA repair capacity on coefficients of a carcinogenesis model for lung cancer." in Int J Cancer 2009; 124: 9: 2152-8

² Luebeck, E. Georg , Moolgavkar, Suresh H. "Multistage carcinogenesis and the incidence of colorectal cancer." in Proc Natl Acad Sci U S A 2002; 99: 23: 15095-100



PARAMETER OVERVIEW

SUMMARY

This document provides information on the data sources used to build the Rice-MDA model, as well as, describes how the model was fit.

RICE MDAnderson Cancer Center BACKGROUND

Readers Guide Model Overview Assumption Overview Parameter Overview control data follow.

Model Overview Assumption Overview Parameter Overview Component Overview Output Overview Results Overview Key References



Rice-MDA (TSCE) Parameter Overview Parameter Listing Overview Data Sources:

The models are fit to case-control data on risk factors combined with external incidence/mortality rate data. A case-control study is currently underway in the M.D. Anderson Cancer Center Department of Epidemiology. In this study, measurements of DNA repair capacity, as well as data on other risk factors such as smoking are being recorded. Cases of lung cancer are matched with cancer-free controls on age (within 5yrs), gender, ethnicity, and smoking status. The MD Anderson case-control data contains information on over 6,000 matched cases and controls.

Characteristics of cases (n=3433) and controls (n=3132) available from the parent lung cancer case-control study (R01 CA55769, Spitz, PI)

Characteristic	C as es	Controls	
Mean Age (SD)	62.3 (11.1)	59.4 (10.9)	
Race/Ethnicity	n (%)	n (%)	
White	27.44 (79.9)	2488 (79.4)	
Black	488 (14.2)	411 (13.1)	
Hispanic	173 (5.0)	210 (6.7)	
Other	28 (0.8)	23 (0.7)	
Sex			
Male	1848 (53.8)	1587 (50.7)	
Female	1585 (46.2)	1545 (49.3)	
Smoking Status			
Current	1371 (39.9)	1120 (35.8)	
Former	1453 (42.3)	1274 (40.7)	
Never	556 (16.2)	702 (22.4)	

A subset of 272 and 919 cases and controls were used to fit the model for males and females respectively.

Since the TSCE model is an incidence-based model, data on mortality or incidence rates are needed. For males, tabled age-specific mortality rate data by smoking intensity and duration from the CPS-II study were used. This tabled data can be found in Smoking and Tobacco Monograph 8¹. For females, tabled age-specific (5yr age bins) incidence by gender, race and smoking status (current, former, and never) from the Nurses Health Study were used. Data from the case-control study and the LC rates from the cohort studies were combined to fit the models.

PARAMETER LISTING OVERVIEW

Details on the underlying TSCE model can be found in Model Overview.

MODEL FITTING

For males and females the TSCE models were fit in 2 different ways. For males, the TSCE model was fit using a Least Squares method while for females a re-sampling based methodology is used.



Rice-MDA (TSCE) Parameter Overview Parameter Listing Overview

Males

A complete description of the model for males can be found in Deng et al 2009. In this model information on not only smoking but also DNA repair capacity as measured in biological assays were included as risk factors. Optimal and Suboptimal DRC were defined as 1 or ½ determined by the cutoff of the median DRC level measured amongst controls. To remove this effect DRC was defined as ¾ for all individuals in this model.

For this model, a least squares approach was used based on the following objective function using the MD Anderson case-control data supplemented by tabled age-specific (5 yr bins) lung cancer mortality stratified by smoking status and duration. The objective function is the combination of 2 components. The first is a chi-square statistic comparing the model-predicted death counts to CPS-II observed death counts. The second is a similar chi-square type of statistic comparing model predicted death counts in the optimal DRC group with the model-predicted death counts in the suboptimal DRC group, multiplied by the estimated relative risk of lung cancer in the optimal DRC group which was based on the MD case-control data. The objective function and assumptions follow.

$$f = f_1 + f_2 f_1 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

 E_i =number of subjects in the *i*th age group at enrollment to CPS-II ($P(t_i < T \le t_i + duration of the study (6 years)|T > t_i)wheret_i = median age of the$ *i*th age group at enrollment to CPS-II and*k*is the number of age groups for a given smoking intensity (20 or 40 cigarettes per day).

The probability $P(t_i < T \le t_i + 6|T > t_i)$ concerns the CPS-II population. In the computation to approximate this probability, a mix of optimal and suboptimal DRC groups with equal weight was used. Furthermore,

$$f_{2} = \frac{((P(35 \le T \le 80|optimalDRC) - P(35 \le T \le 80|suboptimalDRC)) \times R_{21}) \times n)^{2}}{P(35 \le T \le 80|optimalDRC) \times n}$$
where,
$$R_{21} = \frac{P_{2}}{P_{1}} = \frac{P(35 \le T \le 80|optimalDRC)}{P(35 \le T \le 80|suboptimalDRC)} = \frac{P(35 \le T \le 80;optimalDRC)/P(optimalDRC)}{P(35 \le T \le 80;suboptimalDRC)/P(suboptimalDRC)}$$

 R_{21} is an estimate of the relative risk of developing lung cancer given optimal DRC when compared with suboptimal DRC, assuming equal frequencies of individuals with optimal and suboptimal DRC in the population. R_{21} is estimated as the ration of the number of patients with optimal DRC to the number of patients with suboptimal DRC in the corresponding smoking status group.

For more details please refer to Deng et al.². After removing the DRC effect the following are fitted response functions relating smoking intensity measured in cigarettes per day (cpd) to the parameters of the TSCE model.

Never smokers

$$\begin{split} \nu X(t) &= 0.0114 \\ \mu(t) &= 4.845 \times 10^{-7} \\ \alpha(t) &= 1.12 \\ \beta(t) &= 1 \end{split}$$

Smokers while not smoking

$$\begin{split} \nu X(t) &= 0.0934 \\ \mu(t) &= 1.5023 \times 10^{-7} \end{split}$$



$$\alpha(t) = 1.12$$
$$\beta(t) = 1$$

Rice-MDA (TSCE) Parameter Overview Parameter Listing Overview

Smokers when smoking

$$\begin{split} \nu X(t) &= 0.2568 \\ \mu(t) &= 4.1317 \times 10^{-7} \\ \alpha(t) &= 1.12 \times (1 + 0.1655 \times (\log(cpd) - 1)) \\ \beta(t) &= 1 \times (1 + 0.1655 \times (\log(cpd) - 1)) \end{split}$$

Females

For females, the MD Anderson case-control data on smoking histories was supplemented with incidence rate data from Nurses Health Study. In order to adjust for the fact that the MD Anderson cases and controls are matched by both age (within 5 years) and smoking status (current, former, and never smokers), data on age-specific incidence by smoking status are needed to adjust for the biases introduced by matching.

The TSCE model is usually fit to prospective cohort data using maximum likelihood. The cohort likelihood is defined as the product of the individual likelihoods, $L = \prod L_j$

Each L_j depends on the time of entry into the study, s_j , censoring or failure time, t_j , and the individual's exposure history.

$$L_{j}\left(t_{j}, s_{j}\right) = \begin{cases} h\left(t_{j} - t_{lag}\right) S\left(t_{j} - t_{lag}\right) / S\left(s_{j} - t_{lag}\right) & \text{if diagnosed with cancer} \\ S\left(t_{j} - t_{lag}\right) / S\left(s_{j} - t_{lag}\right) & \text{otherwise} \end{cases}$$

In order to fit the TSCE model to case-control data a new method was developed to reconstruct cohort data using the combination of case-control data and tabled incidence/mortality data using re-sampling. The goal of the method is to re-sample case-control cases and controls in proportions reflected in the mortality data to recreate cohort data. Each re-sampled cohort is referred to as a pseudo-cohort and is created by simulating individuals. Each individual is sampled as follows:

- 1. Smoking status (current, former, or never) is sampled using the rates from NHIS for the year 2000.
- 2. Randomly sample which 5-year age bin the individual belongs to by sampling based on the number of individuals in each age bin of the controls, with that smoking status, in the case-control study.
- 3. Using the corresponding incidence table for the smoking status, in the age bin generated above, randomly sample whether the individual has cancer or not based on the estimated probability of an individual with the sampled smoking status within the sampled age bin getting cancer within the 5 years spanning the age bin.
- 4. Once we have a smoking status, age bin, and cancer status we then sample an individual from the MD Anderson dataset with the same characteristics.



5. The censoring or failure time of the individual is assigned as their age from the MD Anderson dataset and the age at entry is assigned as 5 years prior for individuals who do not develop cancer and at a randomly distributed age in the previous 5 years for those who develop cancer.

Ages of enrollment and exit were assigned this way because the cancer status was sampled from the probability of getting cancer over a 5-year interval. If the individual does get cancer during the interval the timing is sampled as uniform over the interval.

10,000 individuals are re-sampled from the case-control dataset for each pseudo-cohort created. Then each pseudo-cohort is fit to the TSCE model by maximizing the cohort likelihood in the usual way. 200 pseudo-cohorts are created and fitted. This provides 200 joint estimates of the parameters for each simulated case-control study. The overall fit is assumed to be the mean estimates over the 200 runs. The following are the fitted response functions relating the parameters of the TSCE model to smoking intensity measured in packs per day (ppd=cpd/20).

$$\begin{split} X &= 10^{7} \\ \nu X(t) &= 1.5 \times (1 + 2.2 \times ppd) \\ \mu(t) &= 1.5 \times 10^{-7} \times (1 + 2.2 \times ppd) \\ \alpha(t) &= 3.2 \times (1 + 0.32 \times ppd) \\ \gamma(t) &= \alpha(t) - \beta(t) - \mu(t) = 0.072 \times (1 + 0.32 \times ppd) \\ \hline \mathbf{CATEGORIES} \\ \hline \mathbf{Core Docs} \end{split}$$

REFERENCES:

- ¹ Thun, M.J., Myers, D.G., Day-Lally, C., Myers, D., Calle, E.E., et al. "Trends in tobacco smoking and mortality from cigarette use in Cancer Prevention Studies I (1959 through 1965) and II (1982 through 1988)." in National Cancer Institute, Smoking and Tobacco Control 1997;
- ² Deng, L., Kimmel, M., Foy, M., Spitz, M., Wei, Q., Gorlova, O. "Estimation of the effects of smoking and DNA repair capacity on coefficients of a carcinogenesis model for lung cancer." in Int J Cancer 2009; 124: 9: 2152-8



COMPONENT OVERVIEW

SUMMARY

This document provides an overview of the components involved in the Rice-MDA model.

RICE MDAnderson OVERVIEW

Readers Guide The main components of the Rice-MDA model are the model describing lung cancer Model Overview Assumption Overview mortality in individuals based on smoking history, and the simulation of lung cancer Parameter Overview mortality in the US population using the smoking history generator.

Component Overview Output Overview

COMPONENT LISTING Results Overview

Key References Smoking history generator

The smoking history generator is used to simulate individual data on smoking histories and death of other cause times to feed into the model and produce simulated LC mortality. The Smoking History Generator uses National Health Interview Survey data to generate for each individual a smoking history (age at initiation, age at cessation, and number of cigarettes smoked per day) and age at death from all causes other than lung cancer, based on the inputs of gender, race and birth year.

TSCE model of lung cancer mortality

The TSCE models described in the Model Overview section are used to calculate risks of lung cancer death in individuals. Using the smoking history generator, the model is then used to simulate LC death on the individual basis.

LC mortality Simulation

Using the simulated smoking histories and death of other cause times, the model of lung cancer mortality is used to simulate LC death as described in the Component Overview section. 50,000 individuals are simulated per birth cohort (1891-1970). Once a population is simulated, then the age distribution by year is adjusted, using reweighting, to match the US population. Details on the simulation of lung cancer in individuals is described in the Survival Mortality Component section.

CATEGORIES

Core Docs



Rice-MDA (TSCE) Smoking History Generator Component

SMOKING HISTORY GENERATOR COMPONENT

SUMMARY

RICE[®] MDAnderson **Readers Guide** Model Overview Assumption Overview lung cancer consortium. Parameter Overview Component Overview OVERVIEW **Output Overview**

The smoking history generator (SHG) is a shared precursor micro-simulation model that produces cohort-specific smoking histories and deaths due to causes other than lung cancer as inputs for the dose-response models used by members of the CISNET

Results Overview The core SHG software was parameterized using three tobacco control scenarios to Key References produce the requisite input data for the models. The first, called the actual tobacco control (ATC) scenario, is a quantitative description of actual smoking behaviors of males and females born in the United States between 1890 and 1984. The second, called no tobacco control (NTC), is a quantitative description of predicted smoking behaviors of males and females in the United States under the assumption that tobacco control efforts starting mid-century had never been implemented. The third, called complete tobacco control (CTC), is a quantitative description of predicted smoking behaviors of males and females in the United States under the assumption that tobacco control activities yielded perfect compliance, with all cigarette smoking coming to an end in the mid-sixties. The ATC scenario used inputs derived directly from observed data in the National Health Interview Surveys (NHIS) and the Substance Abuse and Mental Health Services Administration (SAMHSA) National Survey on Drug Use and Health. The NTC scenario used inputs derived by extrapolating from trends in the observed histories before 1954, i.e., before any tobacco control in the decade leading up to the publication of the Surgeon General's Report in 1964. The CTC scenario was simulated by setting cessation rates to one (i.e., transferring all current smokers to former smokers) and allowing no further initiation starting in 1965 while using the observed values in earlier years.

DETAIL

The SHG accepts parameters supportive of the three tobacco control scenarios described above (see Table SGH-I below). The ATC scenario uses initiation, cessation and smoking intensity (CPD) rates directly derived from the NHIS and SAMHSA datasets. The NTC scenario uses initiation and cessation rates derived by fitting an age-



Rice-MDA (TSCE) Smoking History Generator Component Detail period-cohort model to the ATC rates upto 1954, i.e., before the apperance of any tobacco control measures, and by projecting those into the future maintaining them consistent with the patterns observed in 1954. The CTC scenario uses initiation and cessation rates identical to those of the ATC scenario upto 1965, and then sets the cessation rates equal to one and the initiation rates equal to zero, i.e., all smokers are forced to quit in 1965, and no new smokers are allowed to appear thereafter. All scenarios use smoking dependent other cause mortality (OCD) rates derived from several sources as mentioned above.

Computational process in the usage of the SHG

The CISNET SHG is implemented in C++ and consists of a single simulation class, that receives file system paths to five parameter files, four integer pseudorandom number generator (PRNG) seeds, and an optional immediate smoking cessation year parameter. The SHG simulation class employs four independent random selection processes that are implemented via a class-based wrapper of the Mersenne Twister PRNG.¹

Here we briefly describe the outline for computational process in the usage of the SHG:

1. Initialization

- a. Load input data
- b. Initialize random number streams
- 3. Start Simulation
 - a. Validate inputs
 - b. Determine Initiation Age (if any)
 - c. Determine Cessation Age (if any)
 - d. Compute cigarettes smoked per day (CPD) vector for those who initiate
 - 1. Determine smoking intensity group (based on initiation age)
 - 2. Determine CPD based on smoking intensity and age at initiation
 - 3. Determine uptake period and attenuate CPD during uptake period
 - 4. Generate CPD vector from initiation to cessation or simulation cutoff
 - e. Compute other cause of death (OCD) age
- 5. Write individual outputs
- 6. Loop simulation if repeats are specified



Rice-MDA (TSCE) Smoking History Generator Component Relevant Parameters

RELEVANT PARAMETERS

The SHG utilizes input data from several sources: the NHIS data from 1965 to 2001, the SAMHSA data, the Berkeley mortality database cohort life-tables, the National Center for Health Statistics (NCHS), the Cancer Prevention Study I and II (CPS-I and CPS-II), and the Nutrition follow-up studies sponsored by the American Cancer Society. The NHIS and the SAMHSA datasets provide estimates for prevalence of never, former (by years quit) and current smokers by age and year, and data on smoking intensity (in terms of the average number of cigarettes smoked per day (CPD)). These data were used to create implicit initiation and cessation rates. Using the average initiation rate, the SHG is able to determine the likelihood that a never smoker becomes a smoker. For those individuals that are smokers, the cessation rates are used to determine the likelihood that a smoker becomes an ex-smoker. The Berkeley life-tables, combined with smoking prevalence estimates from NHIS and the relative risks of death for smokers and former smokers in comparison to never smokers from CPS-I and CPS-II, are used to produce the probability of death from causes other than lung cancer based on age, sex, birth cohort, and smoking status. Table SHG-I summarizes the input source for the SHG for the three CISNET tobacco control scenarios.

Table SHG-I

Inpupt	ATC	NTC	СТС
Initiation rates	NHIS	Derived	Derived
			(no new smokers after 1965)
Cessation rates	NHIS	Derived	Derived
			(all smokers quit in 1965)
CPD ¹			NHIS,SMAHSA
OCD ²	Berkely life-tables, NCHS, NHIS, CPS-I, CPS-III, Nutrition Follow-up studies		
Birth year	User Defined		
(1890-1984)			
Gender	User Defined		
(Male/Female)			
Race			User Defined
(All race)			

¹ Cigarettes smoked per day,²Other Cause of Death

ATC: actual tobacco control, NTC: no tobacco control, CTC: complete tobacco control. To simulate life histories for individuals using the SHG, for any given run, the following parameters must be provided:



Table SHG-II

Rice-MDA (TSCE) Smoking History Generator Component Dependent Outputs

Parameter	Valid Values
Seed value for PRNG used for Initiation, Cessation, OCD ¹ , Smoking	Integer from -1 to 2147483647
intensity quintile	(A value of -1 uses the clock time as the
	seed)
Race	0 = All Races
Sex	0=Male, 1=Female
Year of Birth	Integer from 1890 to 1984
Immediate Cessation year ²	0 or Integer from 1910 to 2000
Repeat ³	Integer >1 (number of times to repeat
	simulation)
File paths to Initiation,Cessation, OCD,	As derived from NHIS depending on the
Smoking intensity quintile and CPD ⁴ data files	scenario

¹Other cause of death, ² This variable is set to 0 except for CTC scenario. To apply immediate smoking cessation for CTC scenario, the year for immediate cessation must be supplied to the simulator. If the year value supplied is 0, immediate cessation will not be used in the run. If a year value is supplied, immediate cessation will occur on January 1st of year provided. ³Key is optional and can be excluded. If the Repeat value is included and is not a vector value, each set of parameters will be repeated by the amount specified. If the Repeat value is included and is a vector value, the repeat value will pertain to the value set that it corresponds to. ⁴Cigarettes smoked per day.

DEPENDENT OUTPUTS

The inputs of the SHG are used to simulate life histories (up to age 84) for individuals born in the United States between 1890 and 1984. These life histories include a birth year, and age at death from causes other than lung cancer, conditioned on smoking histories. For each simulated individual, the generated life histories include whether the individual was a smoker or not and, if a smoker, the age at smoking initiation, the smoking intensity in cigarettes per day (CPD) by age, and the age of smoking cessation. Smoking relapse, the probability that a former smoker starts smoking again, is not modeled. Table SHG-III summarizes the output of the SHG. Fig. SHG-1 shows two examples of smoking histories simulated by the SHG; a) an individual born in 1910 who begins smoking at age 17, quits at age 56 and dies at age 67 due to causes other than lung cancer, and b) an individual born in 1920 who begins smoking at age 22 and dies at age 53 due to causes other than lung cancer.

Table SHG-III

Table SHG-III

Initiation Age	Age at smoking initiation
Cessation Age	Age at smoking cessation
OCD ¹ Age	Age at death from cause other than lung cancer
Smoking	Smoking intensity quintile (5 quintiles ranging from light to heavy smoking), Yearly smoking dose
History	(CPD ²)

¹Other cause of death, ²Cigarettes smoked per day.



Figure SHG-1: Examples of the SHG-Generated Events

Simulation results by the SHG can be formatted in four different ways:

- 1. Text (formatted, human readable text depicting smoking history);
- 2. Tab Delimited Data (plain text, suitable for post-processing);
- 3. Annotated text-based timeline (visual representation in text);
- 4. XML (plain text, suitable for parsing). The outputs from the SHG are made up of individual life histories, each of which includes the following variables: birth year, age of smoking initiation, the corresponding smoking intensity (CPD) by age, age of smoking cessation, and age at death from causes other than lung cancer, conditioned on smoking histories.

REFERENCES:

¹ Matsumoto M., Nishimura T. "Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator." in ACM Transactions on Modeling and Computer Simulation 1998; 8: 1: 3-30





SURVIVAL MORTALITY COMPONENT

SUMMARY

This document describes how lung cancer mortality is simulated for individuals using RICE[®] MDAnderson the Rice-MDA model. Readers Guide

Model Overview Assumption Overview OVERVIEW **Component Overview Output Overview**

Parameter Overview The model uses smoking histories and age at death of any other cause generated using the smoking history generator as inputs. If the individual does not die of other causes Results Overview by the year 2000 then their age in 2000 is considered their censoring time. Using these Key References inputs the model then simulates whether the individuals die of lung cancer in their

lifetimes. If they do die of lung cancer then the model produces the age at death of lung cancer. Individual mortality is simulated as follows:

- 1. Complete smoking histories and death of other cause or censoring times are generated using the smoking history generator.
- 2. The probability an individual will not die of lung cancer by their death of other cause or censoring time, t_d , is calculated according to the model, $pTSCE = S(t_d, d)$
- 3. Then a uniform(0,1) random variable, *u*, was drawn
- 4. If $u \le pTSCE$ then the time of censoring is td and no cancer death occurs in the person's lifetime
- 5. If u > pTSCE then lung cancer death occurs during the individual's lifetime and occurs at age, t, computed by inverting the survival function, t = S(t, d).



OUTPUT OVERVIEW

SUMMARY

This document describes the output generated by the Rice-MDA model. Details about the underlying model can be found in Model Overview.

RICE MD Anderson OVERVIEW

Readers Guide Using the TSCE model of lung cancer mortality, predictions can be made about an Model Overview Assumption Overview individual's risk of lung cancer death. The model is used to simulate LC mortality in Parameter Overview the US population as described below.

Component Overview

Output Overview Results Overview OUTPUT LISTING

Key References For each individual age at LC death is simulated based on the smoking history and death of other cause time generated from the smoking history generator.

Simulating LC mortality:

The corresponding male and females TSCE models are used to simulate lung cancer mortality in individuals based on given smoking histories. Simulation of the smoking base case scenarios followed the path depicted below.

- 1. The smoking history generator is used to generate smoking histories and death of other cause ages.
- 2. Given the smoking history and death of other cause times are inputted into the model to generate lung cancer mortality.
- 3. The age distribution for each year of the simulated population is adjusted to match the US population's age-distribution by year.

Details about the simulation of lung cancer mortality can be found in the Component Overview section.

50,000 individuals in each birth cohort 1891-1970 are simulated. Once a complete population is generated, then the age distribution by year is adjusted to match the US population.

CATEGORIES **Core Docs**



RESULTS OVERVIEW

SUMMARY

This document describes the results from the Rice-MDA model in the efforts to determine the impact of tobacco control policy on the rate of lung cancer mortality in the US population.

Readers Guide OVERVIEW

Assumption Overview There were 2 major components of the Smoking Base Case project for which the Rice-Parameter Overview MDA model was used to produce results. Details on the model can be found in Model Component Overview Overview while details about model calibration can be found in Parameter Overview. Output Overview Results Overview

Key References First, for the Hypothetical Scenarios the model was used to produce mortality curves for an individual based on a given hypothetical smoking history. These curves were generated directly from the model.

Second, the models were used to simulate lung cancer mortality in the US population. Using individuals generated from the Smoking History Generator as inputs the model was was used to simmulate whether the individuals died from lung cancer and the age at death. Fifty thousand individuals were simulated per birth cohort 1891-1970, and the resulting simulated population was re-weighted to match the age distribution per calendar year of the U.S. population. Three different scenarios were simulated, Actual, based on the observes smoking histories, Counterfactual based on smoking histories reflecting predicted smoking trends if the Surgeon General's Report of 1965 warning about the dangers of smoking was not published, and lastly the Complete Tobacco Control forcing all people to quit smoking in 1965 after the report.

RICE MDAnderson



RESULTS LIST



Hypothetical Scenarios

The following graphs show the predicted lung cancer mortality rates per 100,000 for the smoking base case hypothetical smoking histories. All hypothetical smoking histories are based on a birth year of 1921. For our model however, birth year does not effect predictions. The predicted mortality rates increase based on the amount and duration of smoking. For former smokers the predicted lung cancer rate decreases but always remains elevated compared to never smokers.



Mortality Rates- Males





Simulation Scenarios The following graphs show the simulated mortality rates per 100,000 for individuals



aged 30-84 in the US population for years 1975-2000. Even though the models were not fully calibrated to the US population, the model still produces reasonable predictions in the Actual Scenario.





Lung Cancer Mortality Rates- Females (aged 30-84)



CATEGORIES Core Docs