FLEXKB DOCUMENT
Version: HI.001.03302011.70024
Document generated: 03/30/2011

CISNET Lung Cancer Collaborators

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.

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

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

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

Erasmus MC	Erasmus MC (Lung)
	Fred Hutchinson Cancer Research Center
	Pacific Institute for Research and Evaluation
	MGH Institute for Technology Assessment
	Yale University
RICE MD Anderson Cancer Center	Rice-MDA (TSCE)

FLEXKB DOCUMENT
Version: HI.001.03302011.70010
Document generated: 03/30/2011

ERASMUS MC (LUNG)

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.

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

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

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. 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
- Population Component
- Risk Factors Component
- Smoking Generator Component
- Natural History Component
- Screening Component

Output Overview

Definitons and methodologies for the basic model outputs.

Results Overview

A guide to the results obtained from the model.

Specific implementations

- Smoking Base Case14Mar06 describes the 14 March 2006 version of our model assumptions for the Smoking Base Case.
- Smoking Base Case16Feb09 describes the 16 February 2009 version of our model assumptions for the Smoking Base Case.

Key References

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

MODEL PURPOSE

SUMMARY

This document describes the purposes of the MISCAN-lung model and the types of questions it was designed to answer.

PURPOSE

The MISCAN-lung model is intended to simulate population trends in lung cancer for comprehensive surveillance of the disease and to estimate the impact of cancer-control interventions (smoking, diet, screening).

Comprehensive surveillance of population trends in lung cancer

The model is primarily intended to simulate observed trends in incidence and mortality from lung cancer in the U.S. population in order to investigate to what extent observed trends can be explained by (earlier) trends in exposure to risk factors, in particular smoking and diet.

The model is also intended to project trends in lung cancer incidence and mortality to years of (future) observation not yet reported. Where the simulation of observed trends concentrates on exposure to risk factors, future trends may be influenced also by trends in screening and therapy.

Evaluation of interventions

The model can simulate the effects of different intervention scenarios in order to compare results in terms of population trends as well as health outcomes, including life years lost due to lung cancer.

More background information on the purposes and aims the model can be found under Model Overview, which also provides a model description including its limitations.

MODEL OVERVIEW

SUMMARY

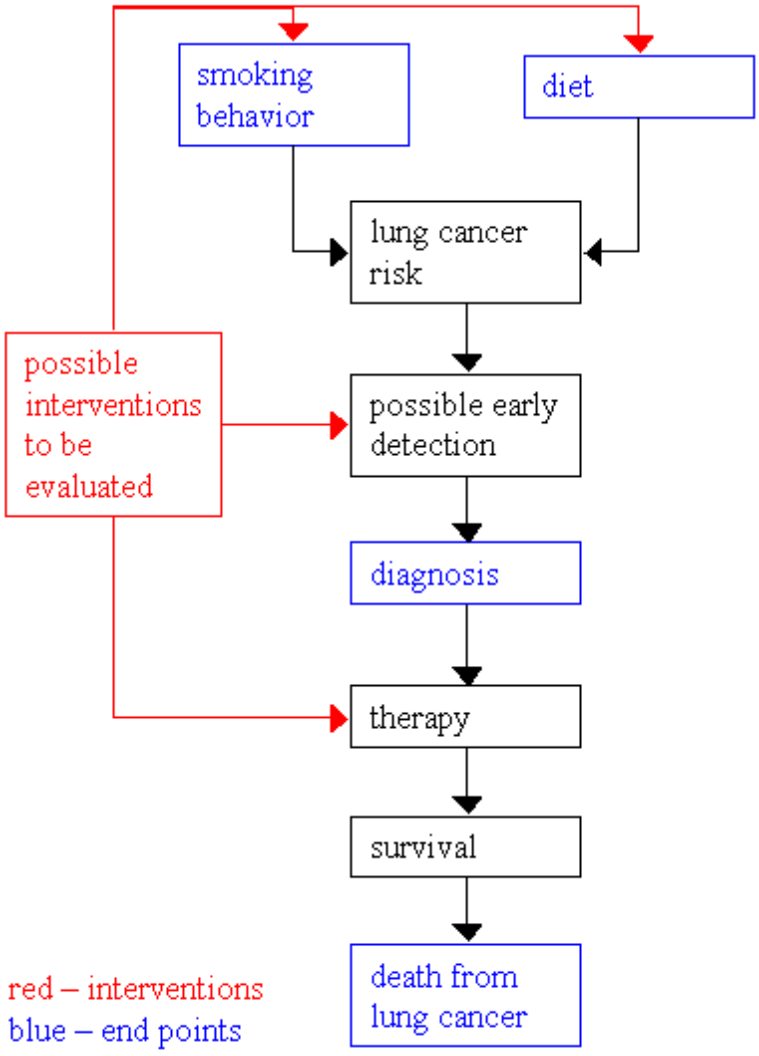
This document provides an overview of purpose and background of the MISCAN-lung model for lung cancer surveillance and offers a brief description of the model.

PURPOSE

The MISCAN-lung model is primarily intended for surveillance of population trends in lung cancer and secondarily for evaluation of interventions, particularly concerning smoking and screening.

MISCAN-lung includes the complete lung cancer chain of events, from behavior- and diet-related risk factors to death from lung cancer (see Figure 1) in order to facilitate evaluation of the influences on population trends and of effects of interventions.

Figure 1. Chain of events to be evaluated for surveillance and interventions that could contribute to the elimination of lung cancer.



Causal chain

Background

Information on aspects of lung cancer that make modeling important is given in Background Information Model Overview. This includes the influence of smoking as the main risk factor for the disease, the process of carcinogenesis, and the importance of screening as even new therapies still seem to have limited effect on advanced stage disease commonly present at diagnosis.

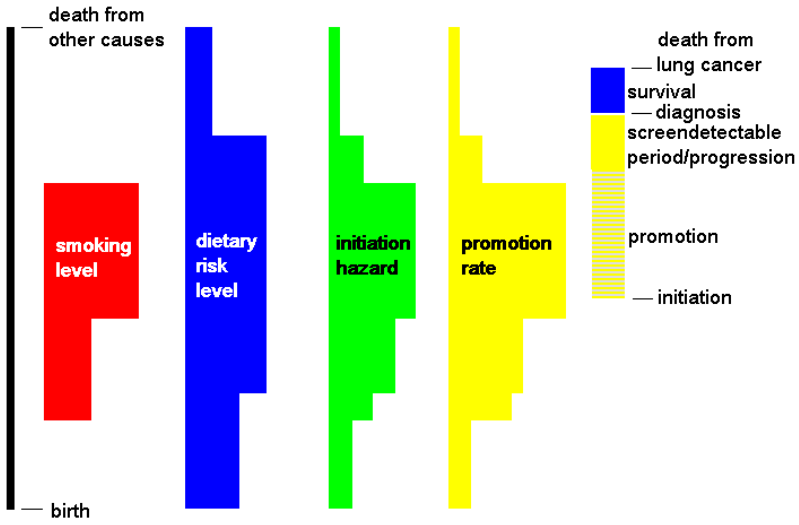
Model Description

The system that is modeled

MISCAN-lung models a human population consisting of individual life histories, in which lung cancer may develop.
The life histories consist of (see Figure 2):

- Exposure to risk factors, particularly smoke and diet.
- Carcinogenesis as influenced by risk factors.
- Timing of diagnosis of lung cancer.
- Net survival from lung cancer.
- Influence of screening on time of diagnosis and death.

Figure 2. Risk model – sample life history



Risk model

Figure 2: Individual life histories are created with events drawn from probability distributions. A life history starts with a birth date and assigning gender/ethnicity. Age of death from causes other than lung cancer is determined from a life table without this specific disease. Then the birth date and gender/ethnicity dependent smoking history is generated, i.e. starting age, stopping age and smoking intensity in between those time points. Likewise, time points are determined at which dietary risk changes. The steps in risk level are combined. The corresponding age-specific risk factors are applied to modify the initiation hazard and promotion and malignant transfer rates in the multistage carcinogenesis model. As a result, a malignant nodule may appear which, after progression to a clinically diagnosed lung cancer (see Natural History Component), leads to lung cancer death. (In this sample, before the projected time of death from other causes.) During progression, the tumor is assumed to be screen-detectable.

General modeling methodologies

We apply the technique of microsimulation of individual life histories in order to constitute a relevant population. The life histories are constructed stochastically by drawing events and development rates from probability distributions in order to reproduce distributions of personal characteristics over the population.

We developed two modules of MISCAN-lung model: smoking history generator module for Smoking Base Case (SBC) calculations and Mayo Lung Project (MLP) module, which is also used to analyze data collected from the Mayo CT screening trial (MCT). The first module uses the smoking history generator provided by NCI to determine probability of exposure to risk factors and subsequently, its impact on lung cancer development, survival, and mortality from other causes. The second module uses the smoking history data and screening data derived from the MLP or MCT to determine the effect of smoking and screening on lung cancer development and survival. The model profile applies to both modules in general. However, descriptions regarding screening only apply to the MLP module. At places where the two modules differ, apart from screening, descriptions of both modules are provided.

The primary unit of analysis

MISCAN-lung based estimates and comparisons with observations will be primarily on population level. By simulating large numbers of individual life histories accurate modeled statistics are generated for the population that those individuals represent.

Major components of the model

An overview of the major components of the model and their relations is shown in Component Overview.

Inputs (see also Parameter Overview)

Probability distributions for

- Time of birth.
- Exposure to risk factors;
- Influenced by exposure to risk factors:

  - Age of death from causes other than the disease of interest;
  - Initiation of carcinogenesis;
  - Promotion or clonal expansion;
  - Malignant transformation.
- Cell type of lung cancer;
- Influenced by cell type and early detection:
  - Progression to clinical lung cancer;
  - Dwelling times by preclinical stage of lung cancer;
  - Stage distribution at diagnosis;
  - Net survival from lung cancer, also influenced by stage at diagnosis.
- Compliance with screening.
- Sensitivity of screening test for detection of preclinical cancer.
- Consequences of screen-detection.

Outputs (see also Output Overview)

- Incidence of lung cancer by stage and cell type, mortality from lung cancer and from other causes, life years in disease states.
- Exposure to screening-tests and test results.

Important limitations of the model

- The most obvious limitation of the model concerns the limitations in knowledge that inform the model, which regards uncertainty in parameter estimates as well as in correct interpretation and structural composition of the model.
- The model has a limited amount of detail, particularly when modeling cancer characteristics.
- The chain of events modeled concentrates on biology and medical interventions in the disease; behavioral interventions e.g. to reduce exposure to tobacco smoke are only included in the model as their effect on exposure.

CONTRIBUTORS

The MISCAN-lung model is an extension of the MISCAN model, which has been developed at the Department of Public Health of Erasmus MC, Rotterdam, the Netherlands with contributions of several people.

The MISCAN-lung model includes the influence of risk factors. It has been developed by:
Rob Boer with contributions of:
Shin-Yi Wu, Haijun Tian, Lu Shi, Marjolein van Ballegooijen, Bill McCarthy, Barbara Berman
programmer of MISCAN extensions:
Floris van Maanen
consultants:
Robert Figlin, Jennifer Malin (until 2004)

ASSUMPTION OVERVIEW

SUMMARY

This document describes the assumptions used in the MISCAN-lung model.

BACKGROUND

A comprehensive model for lung cancer surveillance requires assumptions concerning the following aspects:

- Demography.
- Risk factor exposure.
- Risk factor exposure-effect relationships.
- Preclinical lung cancer.
- Screening.
- Clinical lung cancer.
- Lung cancer survival.
- Mortality from causes other than lung cancer.

The model aspects that are directly observable require the least model specific assumptions. Demography and mortality from other causes are generally observed quite accurately. However, when mortality from other causes needs to be distinguished by exposure to risk factors, e.g. smoking status, which is not directly observed in the U.S. population, there is already need for specific model assumptions.

Clinical lung cancer and its survival can be observed directly. However, often direct observations of survival as well as exposure to risk factors are not available. Therefore relevant assumptions are needed.

Exposure to risk factors is directly observable, but observation of the long term effects of such exposure is much more difficult. Therefore, surveillance of trends in recent history still requires data on exposure to risk factors of several decades earlier when observations were not made. This void can only be resolved by making some extrapolations from more recent observations.

The most typical assumptions for our model concern the aspects: Risk factor exposure-effect relationships, preclinical lung cancer, and screening where interpretation of observations takes place through (formal or informal) model assumptions.

ASSUMPTION LISTING

Demography

Demography assumptions focus on the population characteristics including gender, race/ethnicity, and age distribution. (see also Population Component)

Our model does not assume any entry or exit from the population due to migration. This is generally not a problem when studying a research cohort but to some extent it is

when studying a geographically defined dynamic population.

The primary purpose of our model is to study population trends in a geographically defined dynamic population, viz. the U.S. population. As long as we study a limited time period during which the population does not change substantially due to migration, and when risk factor exposure is measured retrospectively, e.g. by contemporary survey on smoking history, then the problem of migration is very limited: We simulate a births distribution that (in conjunction with modeled mortality) reproduces the demography during the period of the study. The simulated births will not accurately represent actual birth statistics in the U.S. but rather concern births of people who are alive in the U.S. during the study period. In the model, at times long before the period of interest in the study, the simulated population will then be larger than the actual U.S. population at those times. In other words, immigrants are modeled as being born in the U.S.

Risk factor exposure

In general we rely on self-reported exposure to risk factors, which tends to be not very accurate, and on a limited number of questions to characterize exposure history. Therefore, we should assume that there is a substantial amount of uncertainty concerning risk factor exposure. In a non-linear system, such as the effects of risk factors on lung cancer risk, this uncertainty may lead to incorrect estimates but an investigation into this potential problem did not show inaccuracies that would lead to faulty inference in our project.[1]

Risk factor exposure-effect relationships

We adapted the Moolgavkar model on multistage carcinogenesis[2,3] for use within our existing MISCAN microsimulation model (see also Risk Factors Component). Validation of the original Moolgavkar model for smoking and lung cancer is described elsewhere.[3]

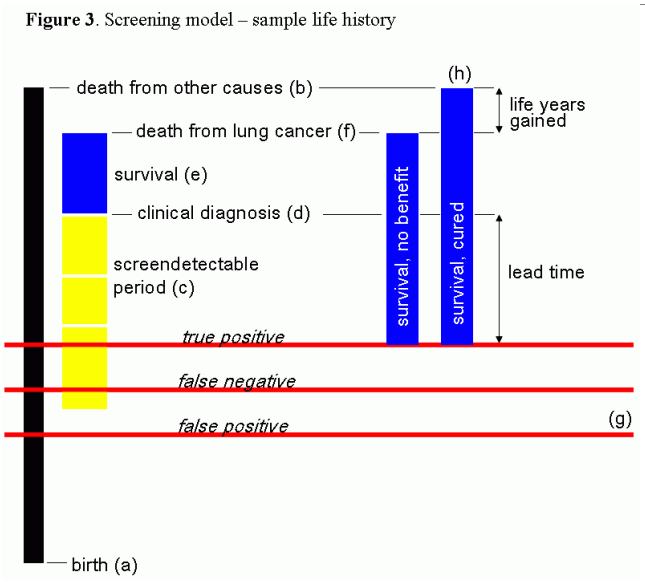
Our adaptation for microsimulation concerns the very early stage: The Moolgavkar model assumes that , after initiation of a stem cell, there is a stochastic process where an initiated cell can form an additional initiated cell, differentiate, or die. The vast majority of initiations does not lead to a clone of initiated cells with any slight chance of malignant transformation. In a microsimulation model this would imply the need to simulate many initiations that only die out within a quite limited period of time, adding a lot of computing time with no effect on risk projections. Therefore, our model only simulates initiations that grow out to a clone of initiated cells that is large enough to have surpassed the stage in which stochastic death or differentiation of individual cells can lead to the end of the whole clone.

Preclinical lung cancer

Although risk factors can influence all stages of the TSCE model, it is generally assumed that the stage of progression from malignant transformation to clinical cancer is not influenced as strongly by risk factors as the earlier stages of carcinogenesis.
We replaced the progression part of the TSCE model with a natural history model of lung cancer. (see Natural History Component)
We assumed a model structure for preclinical lung cancer that is similar to model structures that we have used for the evaluation of screening of other cancers.[4,5,6,7,8,9]
The model assumes that (at least during the screen-detectable period) lung cancer is one of three cell types: squamous cell, adeno/large cell, or small cell carcinoma and that it progresses from preclinical stage I-II to clinical diagnosis in stage I-II ,or to preclinical stage III-IV and then to clinical diagnosis in stage III-IV.

Of the four main cell types in which lung cancer is generally categorized, we joined adenocarcinoma and large cell carcinoma because we considered that there may be a nonnegligible probability of adenocarcinoma developing into large cell carcinoma during the
screen-detectable period.

Screening



Screening Model

(a) Distribution of births over calendar time. (b) Distribution of death from other causes over age. (c) Distribution of start of screen-detectable preclinical period over age. Dwelling time distributions of each preclinical disease state. Transition probabilities to diagnosis versus progression to next preclinical disease state. (d) Age specific incidence depends on distribution of start of screen-detectable preclinical period and dwelling time distributions of preclinical disease states. Stage distribution depends on transition probabilities to diagnosis versus progression to next preclinical disease state. (e) Net survival distribution from diagnosis to death from lung cancer. (f) Mortality from lung cancer depends on incidence, survival and mortality from other causes. (g) Screening defined by times of screening, compliance (two possible mechanisms for timing/compliance), and sensitivity and specificity of screening test. (h) Effect of early detection can be defined by several mechanisms such as probability by screen-detected stage of extending life from death of lung cancer to death of another cause.

We assume a preclinical lung cancer may be detected by screening, depending on the screening-test (see Screening Component). The screening-test is assumed to have a probability of systematic error, in which a preclinical cancer will always be missed due to personal, lesion, or test moment factors. If a preclinical cancer is not missed by systematic error, then it has a probability to be detected depending on the sensitivity of the screening-test. The sensitivity of a screen-test varies by the stage (I-II or III-IV) and type (squamous cell, adeno/large cell, or small cell) of the cancer development.
When a screening is offered, some people will accept it while others will not. In our model we assume that the reach of screening is determined by two factors: 1) whether a screening is a first screening or a repeated screening in a trial, and 2) whether the person attended or missed the previous screening.
There are several possibilities to simulate the consequences of early detection of lung cancer by screening, the most important of which are: no change in time of death; cure from lung cancer (defined by dying at the time originally simulated for death from causes other than
lung cancer); or a new survival distribution. As mentioned in section "Lung cancer survival" below, we currently have two alternative assumptions for MLP and MCT: maintain the cell type and stage specific survival curves for screen-detected cases so that any improvement just results from detection in an earlier stage, or assume 40% cure of screen-detected stage II cancer.

Clinical lung cancer
According to the multistage carcinogenesis model, after malignant conversion occurs, cellular growth is further deregulated and proceeds uncontrolled. This period in

carcinogenesis development can further be divided into preclinical lung cancer, including invasive stages, and clinical lung cancer (Natural History Component). As previously described, preclinical lung cancer can only be screen-detected, but clinical lung cancer can be both screen-detected and clinically detected. Our model structure for clinical lung cancer is similar to the model structures described in section "Preclinical lung cancer" above in order to correspond with the SEER registry; that is, the model assumes that lung cancer is one of three cell types: squamous cell, adeno/ large cell, or small cell carcinoma and that it progresses from preclinical stage I-II to clinical stage I-II, or to preclinical stage III-IV and then to clinical stage III-IV. The clinical detectability is determined by the carcinogenesis model, and the lung cancer survival is described below in section "Lung cancer survival".

Lung cancer survival

After clinical diagnosis we assume a net survival distribution based on SEER by stage category and cell type. In reality the studied population (e.g. the U.S. population) may have a different survival distribution than observed in SEER but we expect that any such differences have a relatively small effect on mortality.

Survival after screen-detection can be modeled in different ways. For the MLP module we made two different sets of tentative assumptions:

- In case of screen-detection a new survival distribution is started that follows the same survival distribution by cell type and by stage category as when clinically diagnosed but the survival curve, starting from the date of early diagnosis, possibly concerns an earlier stage category with a more favorable survival.
- Persons with a screen-detected lung cancer in stage I-II that would be fatal in the situation without screening, receive a probability of 40% of not dying from lung cancer (therefore dying from other causes at a later time), and all other cases die at the same time and from the same cause as in the situation without screening. This assumption gives a reasonably close similarity of observed and modeled survival in MLP.

Mortality from causes other than lung cancer

We assume that death from lung cancer and death from other causes are independent, and modeled two ways of mortality from other causes. In our MLP module, mortality from causes other than lung cancer is assumed to depend on exposure to risk factors. We assume a Gompertz distribution in case of constant exposure to risk factors where both the exponential growth rate in the Gompertz hazard model and the immediate hazard itself can depend on concurrent exposure to risk factors. By this mechanism smokers, for example, can have a higher relative risk of mortality from other causes, including a gradual moderation of the increased relative risk when quitting smoking. Alternatively, in the smoking history generator module (Smoking Generator Component), mortality from other causes is governed by the smoking history generator (developed by NCI), which gives the probability of death from other causes for each year and each birth cohort by smoking status.

REFERENCES:

[1] Heidenreich, W.F., Luebeck, E.G., Moolgavkar, S.H. "Effects of exposure uncertainties in the TSCE model and application to the Colorado miners data. " in Radiat Res 2004; 161: 1: 72-81

[2] Moolgavkar, S.H., Knudson, A.G. "Mutation and cancer: A model for human carcinogenesis." in J Natl Cancer Inst 1981; 66: : 1037-1052

[3] Hazelton, W.D., Clements, M.S., Moolgavkar, S.H. "Multistage carcinogenesis and lung cancer mortality in three cohorts." in Cancer Epidemiol Biomarkers Prev. 2005; 14: 5: 1171-81

[4] Habbema, J.D., van Oortmarssen, G.J., Lubbe, J.T., van der Maas, P.J., "The MISCAN simulation program for the evaluation of screening for disease." in Comput Methods Programs Biomed 1985; 20: 1: 79-93

[5] Loeve, F., Boer, R., van Oortmarssen, G.J., van Ballegooijen, M., Habbema, J.D. "The MISCAN-COLON simulation model for the evaluation of colorectal cancer screening." in Comput Biomed Res 1999; 32: 1: 13-33

[6] de Koning, H.J., Boer, R., Warmerdam, P.G., Beemsterboer, P.M., van der Maas, P.J. "Quantitative interpretation of age-specific mortality reductions from the Swedish breast cancer-screening trials." in J Natl Cancer Inst 1995; 87: 16: 1217-23

[7] van den Akker-van Marle, M.E., van Ballegooijen, M., van Oortmarssen, G.J., Boer, R., Habbema, J.D. "Cost-effectiveness of cervical cancer screening: comparison of screening policies." in J Natl Cancer Inst 2002; 94: 3: 193-204

[8] Loeve, F., Brown, M.L., Boer, R., van Ballegooijen, M., van Oortmarssen, G.J., Habbema, J.D., "Endoscopic colorectal cancer screening: a cost-saving analysis." in J Natl Cancer Inst 2000; 92: 7: 557-63

[9] Draisma, G., Boer, R., Otto, S.J., van der Cruijsen, I.W., Damhuis, R.A., Schroder, F.H., de Koning, H.J. "Lead times and overdetection due to prostate-specific antigen screening: estimates from the European Randomized Study of Screening for Prostate Cancer. " in J Natl Cancer Inst 2003; 95: 12: 868-78

PARAMETER OVERVIEW

SUMMARY

This document provides an overview of the parameters used to quantify the MISCAN-lung model for lung cancer surveillance.

BACKGROUND

The MISCAN-lung model uses four types of parameters:

- Demography parameters.
- Risk factors parameters.
- Natural history parameters.
- Screening parameters.

Currently, treatment parameters have not been modeled but they will be considered in the near future.

PARAMETER LISTING OVERVIEW

Demography Parameters (see also Population Component)

- Births:
  a. Number of birth cohorts;
  b. Distribution of the population among the birth cohorts;
  c. For each birth cohort parameters of its birth table to give the period of dates of birth within the birth cohort;
  d. For each birth cohort the parameters of its life table.
- Stratification of birth cohorts by gender and race/ethnicity.
- Mortality from other causes, either as:
  a. Parameters for the exponential growth rate of the hazard and the baseline hazard for death from other causes, and the dose effect relationships of risk factors on that exponential growth rate and immediate hazard;
  b. When mortality from other causes is governed by the smoking history generator (Smoking Generator Component): the probability of death from other causes for each year and each birth cohort by smoking status.

Risk Factors Parameters (see also Risk Factors Component)

- Number of risk factors: smoking and/or diet, with option of additional risk factor.
- Number of risk levels for each risk factor.
- Risk exposure (e.g., smoking, diet risk) over time at different risk levels.
- Hazards of multistage carcinogenesis as the dose effect relationships of risk factors on initiation, promotion, and malignant transformation.

Natural History Parameters (see also Natural History Component)

- Parameters for the cell type distribution of lung cancer.
- Parameters for the stage distribution of lung cancer.
- Parameters for the initiation, promotion, and malignant transformation of lung cancer.
- Parameters for duration distribution of screen-detectable disease states after malignant transformation.
- Parameters for the transition probability from each stage.
- Parameters for net survival from lung cancer after clinical diagnosis by stage of the cancer.

Screening Test Parameters (see also Screening Component)

- Parameters for screening policy, e.g., timing and dissemination of screening (e.g. start age, end age, screening interval, adherence to screening).
- Sensitivity of a screening test.
- Systemic error of a screening test.
- Parameters for consequence of screening after screen-detected diagnosis:

  a. Probability of dying from other causes due to early detection by screening;
  b. Survival benefit due to early detection by screening.
- Parameters for an individual's screening behavior/adherence:

  a. Probability of screening acceptance by screening type (initial vs. repeat screening);
  b. Probability of screening acceptance by previous screening acceptance.

Summary table

The following table summarizes parameter name, validation criteria, and its use for lung cancer surveillance in the MISCAN-lung model.

Parameters	Validation	Surveillance
	DEMOGRAPHY	
Births:	Reproducing age distribution of the study cohort.	Reproducing age distribution of U.S. population.
Stratification:	Distribution over gender and race/ethnicity as in study cohort.	Distribution over gender and race/ethnicity of U.S. population.
Mortality from other causes:	Correlation between durations, rate ratios of concurrent smoking for mortality hazard and increase rate of hazard in Gompertz distribution reproducing rate ratios of CPS.	--
	RISK FACTORS	
Smoking exposure over time:	Reconstructed from smoking history at enrollment in study.	Reconstructed from U.S. survey data.
Dietary risk over time:	Assuming average dietary risk distribution unless dietary history is available in study data.	Reconstructed from U.S. survey data.
Hazards of multistage carcinogenesis - Dose effect relationships:	Reproducing dose effect relationships estimated with the Moolgavkar model from CPS I and II, British Doctors Cohort and Nurses' Health Study.	--
	NATURAL HISTORY	
Cell-type distribution of lung cancer:	Reproducing distribution over squamous cell; adeno + large cell; and small cell carcinoma in the study cohort.	Reproducing distribution over squamous cell; adeno + large cell; and small cell carcinoma in SEER.
Stages of lung cancer:	Reproducing distribution over stage II or earlier and stage III or later in the study cohort, where available in control group.	Reproducing distribution over stage II or earlier and stage III or later in SEER.
Duration distribution of screen-detectable disease states:	Based on earlier model based estimates from screening studies including the Mayo Clinic Trial.	--
Transitions from each stage:	Based on stage distribution of lung cancer in unscreened study cohort after preclinical phase.	Based on clinical stage distribution of lung cancer in SEER.
Net survival from lung cancer:	Based on SEER disease-specific survival.	--
	SCREENING	
Test sensitivity:	Based on earlier model based estimates from screening studies including the Mayo Clinic Trial.	--
Timing of screening:	As reported from the study.	As reported from surveys.
Consequence of screening:	Based on expert opinion and a range based on confidence limits for improvement of prognosis model parameters for screening trial data.	--

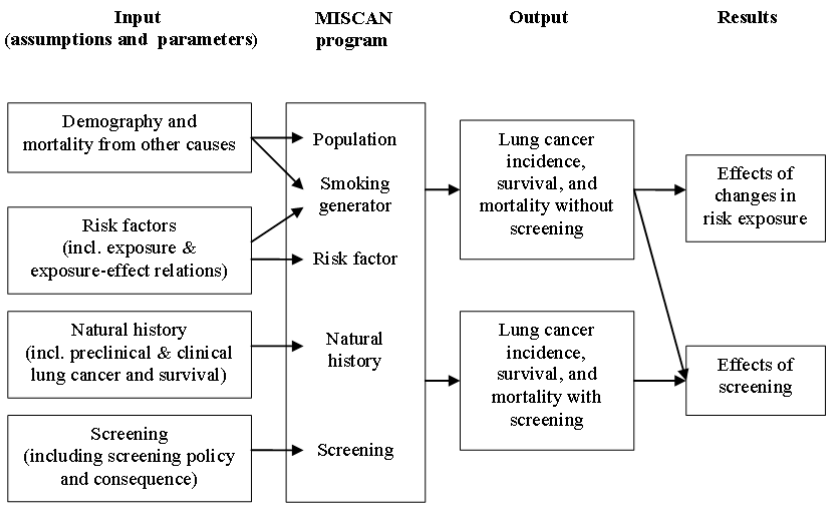
COMPONENT OVERVIEW

SUMMARY

An overview of the five major components in the MISCAN-lung model for lung cancer surveillance.

OVERVIEW

The MISCAN-lung model contains five primary components: population, risk factors, smoking generator, natural history, and screening.



Components overview

COMPONENT LISTING

The MISCAN-lung model consists of five major components.

1. **Population Component**: This component simulates a population of individual life histories, according to the demography and mortality from other causes assumptions and their parameters. Each individual in the population consists of a date of birth and an age of death.
2. **Risk Factors Component**: This component simulates how risk factors (such as smoking and diet) influence the hazard growth rate of lung cancer according to the exposure and exposure-effect relationships assumptions (see Assumption Overview) and their parameters (see Parameter Overview).
3. **Smoking Generator Component**: This component takes data from NCI's smoking generator and simulates the smoking history of an individual and deaths from other causes. In case this component is activated, it replaces the first risk factor in the *risk factors component* and the simulation of deaths from other causes in the *population component*.

4. **Natural History Component**: Subsequently, the *natural history* part of MISCAN-lung simulates separate lung cancer histories (natural histories) for each individual life history. The initiation, promotion, malignant transformation, and progression of lung cancer are generated according to an individual's exposure to risk factors (smoking and diet). The development of lung cancer into different cell types and stages is governed by the natural history assumptions (see Assumption Overview) and their parameters (see Parameter Overview). The survival of a person, once a preclinical lesion has developed into clinical lung cancer, depends on the cancer cell type and stage of disease. The life history of each person is altered according to the natural history that is simulated for that person. If he or she dies from lung cancer before he or she dies from other causes, his/her death age is adjusted accordingly.
5. **Screening Component**: After simulating the natural history if screening were absent, the *screening component* makes detection of preclinical lung cancer possible. Timing of screenings can follow an invitational schedule or an opportunistic pattern. Screening in the model potentially affects early stages (I and II) of all preclinical lung cancer, resulting in either a cure or a new survival upon screen-detection. The effectiveness of screening depends on the screening assumptions (see Assumption Overview) and their parameters (see Parameter Overview).

SMOKING HISTORY GENERATOR COMPONENT

SUMMARY

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 lung cancer consortium.

OVERVIEW

The core SHG software was parameterized using three tobacco control scenarios to 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-

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**

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**	**CTC**
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 (1890-1984)	User Defined		
Gender (Male/Female)	User Defined		
Race (All race)	User Defined		

[1] Cigarettes smoked per day,[2]Other 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

Parameter	Valid Values
Seed value for PRNG used for Initiation, Cessation, OCD[1], Smoking intensity quintile	Integer from -1 to 2147483647 (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, Smoking intensity quintile and CPD[4] data files	As derived from NHIS depending on the scenario

[1]Other 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. [3]Key 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. [4]Cigarettes 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 History	Smoking intensity quintile (5 quintiles ranging from light to heavy smoking), Yearly smoking dose (CPD[2])

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

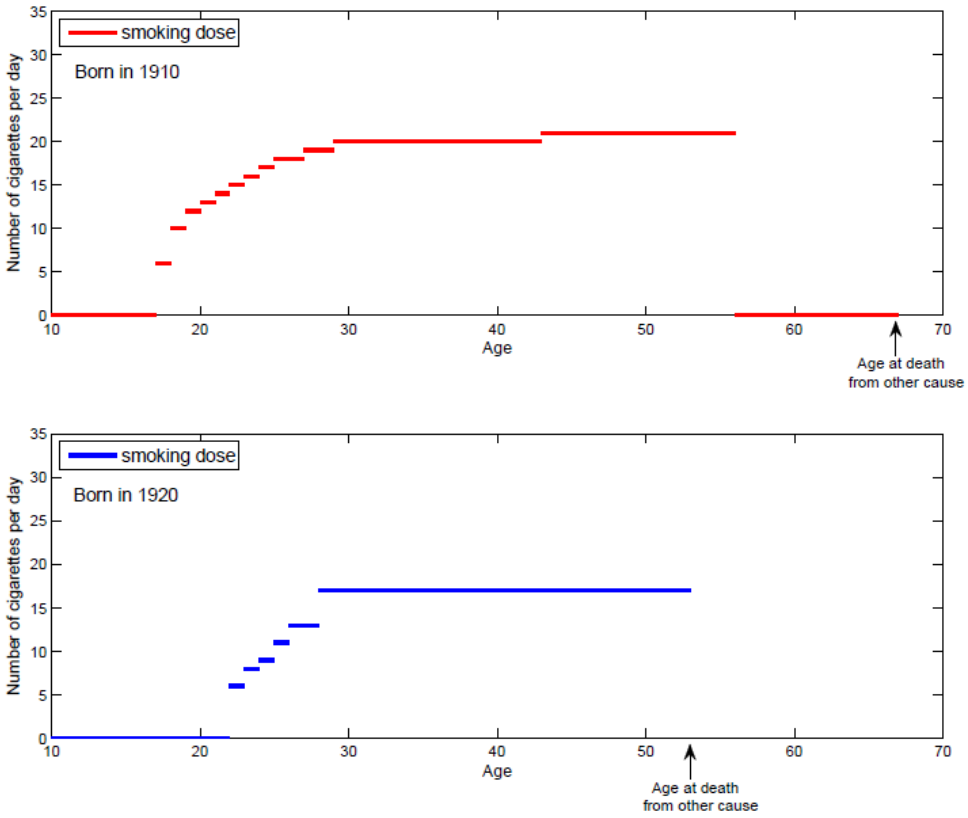


Figure SHG-1: Examples of the SHG-Generated Events

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

POPULATION COMPONENT

SUMMARY

This document gives a description of how the simulated population is modeled.

OVERVIEW

The model simulates a dynamic population by generating births according to a distribution over calendar time, e.g. the relative sizes of birth year bins of persons in a specified study.
Stratification on the basis of specific subpopulation characteristics is possible.
The general population model includes mortality from other causes that can depend on exposure to risk factors.
However, in the Smoking Base Case, mortality from other causes is determined by the *smoking history generator* (see Smoking Generator Component).
Thus, a population consists of individuals whose life histories in the absence of lung cancer begin with a date of birth and end on a date of death from other causes.

QUANTITATIVE DESCRIPTION

The MISCAN-lung code is an individual-based microsimulator. Life histories are determined by random draws from probability distributions (Monte Carlo simulation) for allocation to categories (e.g. birth year cohort) or selecting time to event (e.g. death from other causes).

POPULATION DYNAMICS

The MISCAN-lung code models the population by simulating individuals from birth to death from disease or death from other causes. Distribution across birth year bins must be defined at input. Age distributions in specific calendar years can be computed at output.

RECURRENCE

During the life history of an individual person multiple clones of malignant cells may be created, which each will progress through various states of disease from preclinical to clinical lung cancer. (Also see Risk Factors Component and Natural History Component.)

DISEASE DISTRIBUTION

In the MISCAN-lung model the lung cancer stages I and II are combined into stage 2-, and the stages III and IV into 3+.
Three lung cancer cell types are distinguished, i.e. small cell carcinoma, squamous cell carcinoma, and combined adeno / large cell carcinoma.
Upon creation of a malignant nodule (clone of malignant cells) the decision about cell type is made according to clinically observed fractions. (Also see Natural History Component.)

RELEVANT ASSUMPTIONS

See sections

- Demography;
- Preclinical lung cancer;
- Clinical lung cancer;
- Lung cancer survival;
- Mortality from causes other than lung cancer;

in Assumption Overview.

Relevant Parameters

See **Demography** and **Natural History** parameters in Parameter Overview.

RISK FACTORS COMPONENT

SUMMARY

This document describes how MISCAN-lung models carcinogenesis and the influence of risk factors on this process.

OVERVIEW

Multistage carcinogenesis

The multistage carcinogenesis model as developed by Moolgavkar et al.[1,2] consists of the following stages:

- Initiation of stem cells: One or more mutations result in an initiated cell that partially escapes growth control.
- Clonal expansion of initiated cells: The single initiated cell develops into a clone of initiated cells.
- Malignant transformation: Each of the initiated cells in an expanding clone can acquire further mutational changes leading to a malignant cell.
- Progression to diagnosis: Malignant cells develop into a symptomatic cancer.

An initiated cell multiplies at a fairly high rate, but there is an almost as high rate of cell death or differentiation. The latter implies, similarly to cell death, the end of the malignant potential of the cell. The model assumes that multiplication and death/differentiation are stochastic processes; therefore, the large majority of clones die out because there happened to be one more cell death/differentiation than reproduction of initiated cells. These clones contribute practically nothing to the cancer risk. A relatively small number of clones of initiated cells succeeds in growing to a substantial size, by which it is unlikely to die out. These clones follow a slow but sure path to containing a large number of initiated cells, making it more likely that one of the initiated cells undergoes malignant transformation.

Each cell in a clone of initiated cells has a hazard of malignant transformation. Because the clones increase in numbers of initiated cells, the hazard of malignant transformation in a clone increases over time. At a certain moment the number of cells is so large that the stochastic element doesn't play a significant role any more. A constant rate of generation of new initiated cells implies that the growth of an expanding clone is exponential. Because they are the clones that produce more initiated cells than cell deaths or differentiations, the clones that lead to cancer grow faster during the earlier stage of development. That implies that the time distribution from initiation to malignancy has a clear mode, which is generally estimated to be at a distance of several decades.

After malignant transformation, there is a stage of progression of the cancer until it is developed far enough to be diagnosed.

Contrary to the original Moolgavkar model, MISCAN-lung disregards the clones of initiated cells that die out early in their development. The shape of the distribution of

the period of clonal expansion to malignant transformation determines that in the first several decades of their lives people rarely get cancer. There is a steep increase in risk of cancer by age that reflects mainly the steepness of dwelling time distribution of clonal expansion to malignant transformation. When most of the clones that were initiated very early in life have come through as cancer, the slope of increase in cancer risk levels off. Hence the slope of increase in cancer risk at higher ages reflects to large extent the increase in the rate of initiation.

Risk Factors

Risk factors can influence initiation, clonal expansion (promotion), malignant transformation, and progression.

If a risk factor influences the rate of initiation, it will take a very long time before there is a substantial influence on cancer incidence because the stage of clonal expansion must be passed through before becoming a cancer.

If e.g. the rate of clonal expansion is reduced by half, then the rate of malignant transformation starts decreasing immediately. The subsequent period of progression from malignant cell to diagnosed cancer makes the influence on cancer incidence somewhat less immediate.

A change in a risk factor that leads to e.g. a reduction by half of the rate of malignant transformation has a more immediate effect on malignant transformation, but in the end only postpones the appearance of cancer by one doubling time of a clone of initiated cells; if the doubling time of cancer incidence is less than the doubling time of a clone of initiated cells, the rate ratio for malignant transformation decreases over time since change of the risk factor.

It is generally assumed that the stage of progression from malignant transformation to clinical cancer is not influenced as strongly by risk factors as the earlier stages of carcinogenesis.

Implementation of the Risk Factors Model

Input parameters

1. Parameters for carcinogenesis in absence of risk factors

For each stratum the model specification consists of:

- The number of cells that start an initiated close. This is generally larger than 1 in order to adjust for the higher initial growth rate of surviving clones.
- Basic rate of initiation.
- Basic rate of proliferation of a clone of initiated cells, also called: promotion.
- Basic rate of malignant transformation.
- Basic growth rate of cancer, also called: progression.

2. Risk Factors

MISCAN-lung can model up to 5 different explicit risk factors, each with up to 10 levels of exposure. For each stratum and risk factor the model specification consists of:

- The probability of starting at a risk level (thus, up to 10 probabilities of exposure intensity).
- The probability matrix of transition from the current level of exposure to the next risk level.
- For each current risk level:
  - The dwelling time distribution to first change of risk level.
  - The dwelling time distribution to second change of risk level.
  - The dwelling time distribution to third or later change of risk level.

Alternatively, the model for exposure to the first risk factor (= smoking) can be replaced by the smoking history generator (see Smoking Generator Component).

3. Dose effect relationships

For each risk factor and each level of exposure the model specification consists of the factor by which the rates of initiation, promotion, malignant transformation, and progression are adjusted at the time of exposure to the given level of the given risk factor.

Computation

The life history is split up into segments during which there is no change in the level of exposure to any of the risk factors. For each of these segments, the adjustment factor resulting from exposure to all of the risk factors is determined and applied to the rates of initiation, promotion, malignant transformation, and progression.
The rate of initiation of new clones is constant during each segment of constant risk factor exposure and is adjusted when exposure changes.
The current clone size, c is initiated with the specified number of cells that start an initiated clone.
The time to malignant transformation is determined iteratively as follows:
The proliferation rate of the clone of initiated cells, given current exposure to risk factors, is p; and the malignant transformation rate per initiated cell is m.
Let u be a draw from the standard uniform distribution.
Then it is determined if the following period of time is shorter than the length, l of the current segment of constant exposure to risk factors:

$$\{ln[c \times m - ln(1-u) \times ln(p)] - ln(c) - ln(m)\}/ln(p)$$

If so, then the time of malignant transformation is reached; if not, the clone size is updated to the value at the end of the current segment of constant exposure to risk factors: c becomes $c \times p^l$

The iterations are repeated for subsequent segments of constant exposure to risk factors until the moment of malignant transformation or the maximum life span is reached.

Progression from malignant transformation to clinical diagnosis and time from birth to death from causes other than lung cancer are determined similarly.

REFERENCES:

[1] Moolgavkar, S.H., Knudson, A.G. "Mutation and cancer: A model for human carcinogenesis." in J Natl Cancer Inst 1981; 66: : 1037-1052

[2] Hazelton, W.D., Clements, M.S., Moolgavkar, S.H. "Multistage carcinogenesis and lung cancer mortality in three cohorts." in Cancer Epidemiol Biomarkers Prev. 2005; 14: 5: 1171-81

SMOKING GENERATOR COMPONENT

SUMMARY

This document describes how data from the NCI's Smoking History Generator Application is used within the MISCAN-lung model.

OVERVIEW

The Smoking History Generator Application has been developed by NCI staff for the CISNET program, based on NHIS data on the U.S. population.
The original MISCAN-lung model structure for determining exposure to the first risk factor (i.e. smoking) was replaced to accommodate the optional use of data from the Smoking History Generator. Those data are provided in tables, which can be read by the MISCAN-lung code to produce appropriate random smoking histories for the individual persons simulated.

The onset of smoking is determined by a table of probability to start smoking by single year of age, 5-year birth cohort, gender and race (i.e. whites or all races). Cessation of smoking is determined by a similar table.
The smoking intensity is modeled as cigarettes per day (cpd). First, placement in one of five smoking intensity categories is determined by a table with a probability distribution over 5 categories (from light to heavy smoker) by age of initiation. Subsequently, the number of cigarettes smoked per day is determined using a table of cpd by year of birth, age, race and gender, and smoking intensity category.

It is assumed that, once a smoker has been assigned to an intensity class, this level of exposure will remain constant unless the person quits smoking altogether.

For *never* and *current* smokers, the time of *death from causes other* than lung cancer is determined from a table of other cause death probabilities by race and gender, year of birth, age, and smoking status (never or current) and intensity (5 categories for current smokers).
For *former* smokers, the difference between the 'current' and 'never' probability for the person is multiplied by the following excess risk formula,

$$exp[(-0.1711+0.00102\times cpd+0.00171\times QuitAge)\times(YearsQuit)^{1.08}]$$

and added to the 'never' probability to obtain the 'former' probability.
A thus generated smoking history can be treated as usual input to the risk factors model (see Risk Factors Component) in order to continue simulation of the development of lung cancer.

ADDITIONAL REMARKS

For the Smoking Base Case, the functionality of the Smoking History Generator application has been extended in 2009 by NCI staff.

- Birth years 1890-1900 were added to the tables.

- Table data were supplied for the counterfactual (No Tobacco Control) scenario, next to those for the actual (Tobacco Control) scenario.

The additional tables were used just like the original ones in MISCAN-lung.

NATURAL HISTORY COMPONENT

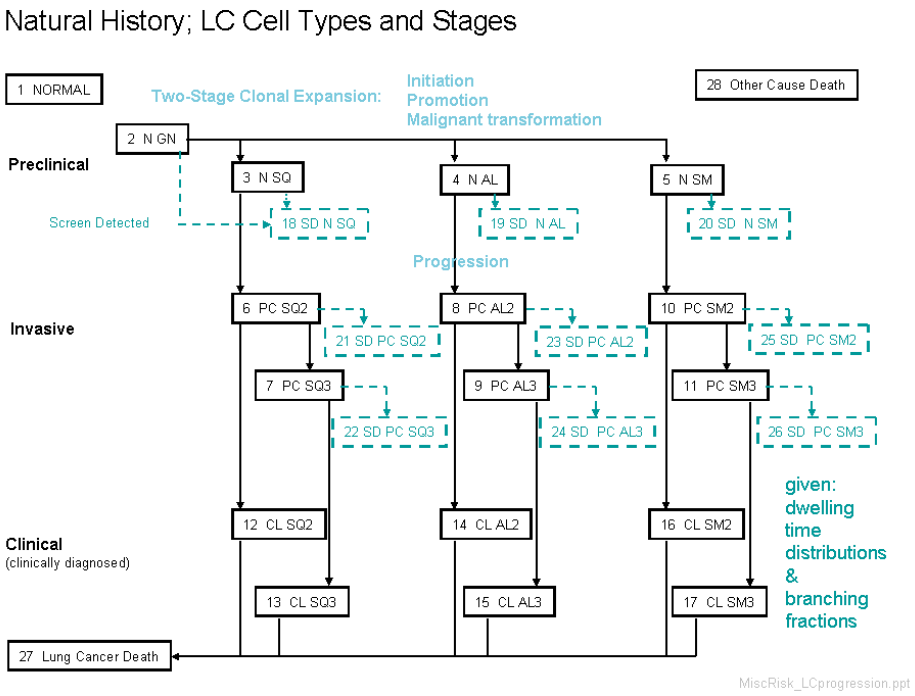
SUMMARY

This document gives a description of the model processes responsible for generating the natural history of disease.

OVERVIEW

The model simulates a network of disease states (Figure 4) categorized by the following dimensions: cell type (squamous cell, adeno/large cell, and small cell carcinoma); stage (stage I-II and stage III-IV); and clinical status (preclinical, clinically diagnosed, and screen-detected).

FIGURE 4:



Natural History

Natural History Model

(Figure 4) After initiation of cells in normal state (1) and promotion and malignant transformation of generic nodules (2), nodular fractions of squamous cell carcinoma (SQ), adeno plus large cell carcinoma (AL) and small cell carcinoma (SM) appear in states (3), (4) and (5), respectively. This process is governed by the parameters of the multistage carcinogenesis model. Lung cancer cell type distribution corresponds with clinically observed fractions.

Further progression through preclinical states (6-11) occurs, where stage 2- (= I-II) and stage 3+ (= III-IV) cancers may develop. Progression continues to clinically detectable cancers in states (12-17), which may result in the person's death from lung cancer upon entering state (27).

Branching and dwelling time

Branching fractions and dwelling time distributions determine the time-course of the state of the progression model. Three types of dwelling time distributions are used to describe the duration of the stay in one compartment until transition to the next compartment. They are:

- Weibull distribution, characterized by a mean value and a shape parameter;
- Piecewise linear distribution, consisting of a set of (time point, probability of transition before this time point) data;
- Fixed duration, i.e. transition after a fixed period of time.

Death from other causes

The simulated person may die from causes other than lung cancer, i.e. entering state (28), if this event –at a projected time point which was determined during an earlier step in the model– occurs before progression through the natural history has finished.

Screening

Screening, if performed, may detect cancers in preclinical states (2-11), which means transfer to corresponding screen-detected states (18-26). No further transition is modeled.

Disease Stages

The disease stages 2- and 3+ are distinguished as described in the above section "Natural History Model". As is shown in Figure 4, a stage 2- tumor may become a stage 3+ tumor. At transition time it is decided whether the nodule continues in the next model state as a stage 2- or stage 3+ tumor. Such branching takes place according to a fixed fraction parameter (model input).
Time until transition from one state to the next one in the model is determined from random draws from given dwelling time distributions.

Disease growth

Tumor size is NOT a quantity monitored in MISCAN-lung. Once malignant transformation has occurred, by which a generic nodule appears in state 2 (Figure 4), tumor progression continues by stochastic state transitions according to given dwelling time distributions and branching fractions.

It is assumed that the tumor is detected (clinical diagnosis) the moment the nodule, now identified as either squamous cell or small cell or adeno/large cell carcinoma, enters one of the clinical states (12 through 17 in Figure 4).

Disease evolution

See section **Disease growth** above.

Regression

So far, the possibility of tumor regression has NOT been modeled in MISCAN-lung. Only irreversible progression is modeled, which can be influenced by adapting the dwelling time distributions in the various states.

Relevant Assumptions

See **Assumption Listing** in Assumption Overview.

Relevant Parameters

See **Natural History** parameters in Parameter Overview.

SCREENING COMPONENT

SUMMARY

This document describes the processes in the model that are responsible for generating screening dissemination and detection of disease.

OVERVIEW

The screening component simulates the screening program for lung cancer and its effects.

DISEASE DETECTION MECHANISM

Preclinical lung cancer can pass through a number of disease states before clinical diagnosis. Each of these states has a dwelling time distribution. If a screening takes place during the phase of preclinical lung cancer, there is a probability of detection of the cancer by the screening test that depends on stage and cell type. This probability is called test sensitivity.

A screening examination may consist of more than one (up to three) screening-tests. In case of simple model assumptions, the probability of a positive test result is taken to be independent of the results of the same tests in previous screens, and also independent of the results of other tests applied in the same or in previous screenings.

Systematic errors from screening-tests can occur for any of the following reasons:

- Person: For example, it is possible that a person has always had a positive sputum test result in lung cancer screening.
- Lesion: For example, a lesion can be missed systematically because the screening-test is less sensitive for some lesions that for others.
- Test moment: For example, in lung cancer screening it is possible that a particular sputum cytology test yields a negative result because no material from any of the malignant lesions was present in the sputum at the moment of the test.

It should be noted that both dwelling time distributions and sensitivity are generally estimated from screening data. Therefore, the dwelling time distribution for lung cancer states concerns disease that is in principle screen-detectable and does not start at the time one single cell or an arbitrary low tumor size is present.

Screening Dissemination

Each stratum (*) may have its own definition of exposure to screening, which can be used to specify a relation between e.g. lung cancer risk and uptake of screening.

A typical screening policy is defined by the ages at which persons will be invited for screening and the year from which this policy is implemented. At first invitation a simulated person attends with a given probability. The age at first invitation is not always the first invitation age of the program because the person may be older when the program starts. At the subsequent invitations for screening the probability of a person attending depends on attendance to the previous screening. In general we have observed that the percentage of people who accepted the previous invitation and is again attending a subsequent screening is around 60 higher than that of people who did not show up at the previous invitation.

Alternatively, timing of screenings can be defined as the age distribution of receiving the first screening; a probability to receive a second screening and the interval distribution to that second screening; and the interval distribution to subsequent screenings depending on the length of the previous interval.

(*) A stratum as a subset of the modeled population that can have a different birth table (to define cohorts), life table, exposure to risk factors, risk-effect relationships, and screening participation.

Type / Detection Interaction

Test sensitivity can depend on tumor cell type.

Stage / Detection Interaction

Test sensitivity can depend on disease stage.

Length Bias

When sensitivity is constant during the preclinical cancer period, at first screening, screen-detected cancers will be found on average halfway their preclinical period. Therefore, the lead-time is on average half of their preclinical period. This implies that cancers with a long preclinical period tend to have a longer lead-time. Because of the longer duration of possible detection by first screenings, first-screening-detected cases tend to have longer dwelling times than the average cancer. When the dwelling time distribution is exponential, then the average lead-time of cases detected at first screening will be the same as the average dwelling time of the average cancer, despite the fact that among the screen-detected cases the average lead-time is only half of their dwelling time.
This phenomenon causes an extra long lead-time effect on survival from screen-detected cases.
At repeat screenings (unless after a very long interval) there will be relatively fewer cancers detected with long dwelling times but the average lead-time will be longer than half of their dwelling time.

Another possible length-time effect concerns a possible correlation between preclinical dwelling time and survival from clinical diagnosis. MISCAN-lung can explicitly model such an effect but our current models do not do this.

Detail

For each simulated initiation, an anatomical site may be generated, for instance central versus peripheral location in the lung.

Positive test results can change the course that the disease would take without screening. There are two ways of specifying the consequences of screen-detection: as modifications relative to the original course of the disease, or as a new course independent of the original course.

When defining consequences as modification relative to the original course of the disease, the model accounts for the effects of lead-time due to early detection of cancer and for diagnoses of cancer that would not have occurred without screening (often called overdiagnosis or extra incidence). The moment of death from disease can be delayed, and the probability distribution of the length of the delay should be specified. Important special cases of delay are complete cure (infinite delay) and no change (zero delay).

Defining consequences as an independent further course of the disease consists of specifying a new survival distribution from time of screen-detection.

Another possible consequence of screen-detection is a probability of (surgery) mortality at the time of diagnosis and treatment.

Relevant Assumptions

See section **Screening** under **Assumption Listing** in Assumption Overview.

RELEVANT PARAMETERS

See **Screening Test** parameters in Parameter Overview.

OUTPUT OVERVIEW

SUMMARY

This section describes the outputs generated by the MISCAN-lung model for lung cancer.

OVERVIEW

The output of the MISCAN-lung program consists of the simulated events (e.g., the number of cases diagnosed, number of cases missed by screening, and mortality from the disease and from other causes) and person time (e.g., the life years lost due to the disease and life years with the disease). Most output is given by calendar year and disease state. The output is stored in data files to enable further calculations with the simulated results.

The MISCAN-lung model simulates among others the outputs for the Smoking Generator and Screening Base Cases. The output on screening effects are limited to the case in which the Mayo Lung Project (MLP) module is activated.

OUTPUT LISTING

The MISCAN-lung model produces output data in files, which can be processed further to yield the final outputs of the model. (E.g. using Microsoft Excel or a statistical package like SAS or SPSS.)

The main outputs of MISCAN-lung are:

1. Lung cancer (LC) incidence;
2. Mortality (Lung Cancer and Other Causes);
3. Survival/life years in disease states;

by time, stage, cell type, and demographics.

Because our model is a microsimulation model, we can also produce the following outputs:

4. Age groups required in the output;
5. Lead time;
6. Overdiagnosis;
7. Individual life history;
8. Simulated screening tests and test results.

The outputs in the base case analyses include:
Smoking Base Case

1. Prevalence of lung cancer in 1986 by age groups in the range 30-84 y.
2. Age-adjusted lung cancer incidence rate by calendar year (1975-2000).
3. Age-adjusted lung cancer mortality rate by calendar year (1975-2000) and by smoking status.

4. Smoking prevalence by calendar year (1975-2000).
5. Smoking attributable lung cancer mortality.

Screening Base Case

6. Number of invitations for screen-tests and opportunistic screen-tests for each year.
7. Number of positive and negative test results per preclinical state and per year.
8. Total number of life years, life years lost due to cancer, number of specific deaths and non specific deaths.
9. Number of screenings that prevented cancer by year of screening.
10. Number of screenings that detected cancer early by year of screening.
11. Number of life years gained due to screening by year of screening.

During development of the MISCAN-lung model calibrations were performed for CISNET Base Case analyses. The model has been validated by simulating the Mayo Lung Project (flat screen X-ray screening) and ELCAP (CT screening). See Results Overview.

RESULTS OVERVIEW

SUMMARY

Describes the general results for lung cancer obtained from MISCAN and MISCAN-lung model output. (Before 2008.)

OVERVIEW

This document describes results of MISCAN & MISCAN-lung on calibration and validation, and results concerning Base Case analyses. (Before 2008.)

(The MISCAN model does NOT include the Risk Factors Component of MISCAN-lung.)

RESULTS LIST

1. Calibration

For the Base Case analyses we calibrated the MISCAN model to common inputs on:

- Cancer incidence per 100,000 (SEER 1975-1979) by age group;
- Cell type prevalence (SEER) by age group;
- Stage distribution (SEER) by age group;
- Relative survival by stage, cell type, and age group;

Calibration was done on cancer incidence, prevalence and stage and cell type distribution of cancers. MISCAN reproduces the base case inputs well. Only lung cancer incidence, prevalence, and mortality in the older age groups (> 70 years) differ significantly.

2. Model Validation by Simulation of Mayo Lung Project - flat screen X-ray screening

We have tried to estimate a model of screening for lung cancer from the Mayo Clinic randomized trial on lung cancer screening that started around 1975. The Mayo Lung Project[1,2,3,4,5,6] was a randomized controlled trial designed to detect lung cancer at a curable stage. Screening tests included chest X-rays, 3-day pooled sputum cytology studies, and lung-health questionnaires. These tests were given to a study population of 9,211 male outpatients with a negative first screening for lung cancer and high risk for the disease. Both trial arms (intervention and control) received a first screening and the intervention arm continued to receive screenings every four months for six years. Lung cancer diagnoses were followed up to 30 June 1983 and mortality was followed up to 31 December 1996. The trial was successful in detection of early lung cancer but not in prevention of lung cancer mortality.

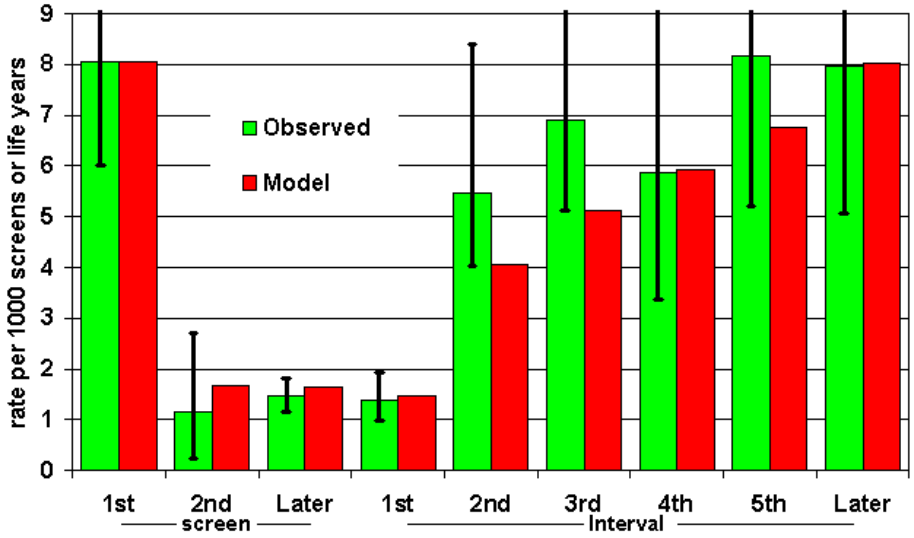
Four models (Models A, B, C, D) were developed before we had access to the data set of the Mayo Clinic trial. Model A assumes that screening test sensitivity is 100% and the sojourn time has an exponential distribution from the time of becoming screen-detectable to the time of clinical diagnosis without screening. Model B adjusted model A by assuming three times longer sojourn times and fitted the test sensitivity to the detection rate at first screening again. Model C adjusted model B by assuming the possibility of a systematic negative screening result. Finally, we constructed Model D,

which is in agreement with observed interval cancer incidence.

We then applied an automatic fit procedure based on the Nelder and Mead method (or amoeba) by simultaneously adjusting model parameters until best agreement with observed data was reached[7]. We fitted the modeled onset of preclinical screen-detectable disease so that incidence in a situation without screening agrees with incidence observed in SEER. Subsequently, we fitted the screen-detectable sojourn times, test sensitivity and a relative lung cancer risk of the trial to the results from the Mayo Clinic trial. Starting from the best fit, we further investigated to what extent indolent cancers give a better explanation of observed data. In addition, we tested the design of the study to test the randomization of the study.

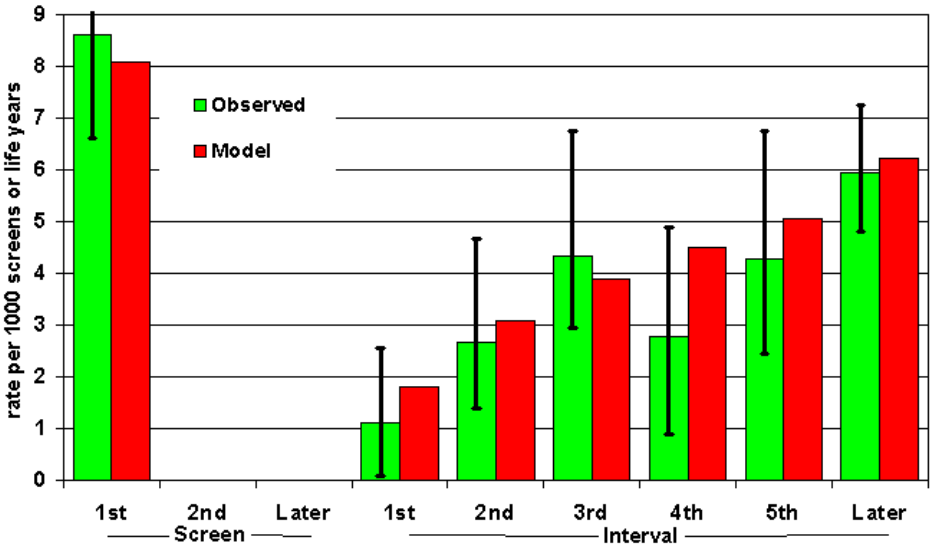
The best model fit so far has good agreement with observed data. Figure 8 compares screen-detected rates and interval cancer incidence of the intervention group and Figure 9 of the control group. Although the interval cancer incidence in the intervention group still looks low in comparison with observed data, this does not reach a threshold of statistical significance of 5% and therefore can be regarded as due to random noise. We found that even our best model fit predicts lower rates of interval cancer incidence of adeno carcinoma/large cell lung cancer. That appears to be consistent with an assumption of overdiagnosis of adeno carcinoma due to screening. The simulated results show that there are systematic missed lesions in either preclinical stage 2- and preclinical stage 3+. Our model also predicts that indolent cancers are not a serious issue in the Mayo Clinic trial. Finally, our model predicts higher cancer incidence in the study group compared to the control group, which provides evidence against the randomization of the trial.

Figure 8. Screen-detection rates per 1000 screens and interval cancer incidence per 1000 life years of the intervention group



cancers MLP intervention

Figure 9. Screen-detection rates per 1000 screens and interval cancer incidence per 1000 life years of the control group



cancers MLP control

The model results presented above are produced by the MISCAN model as *before* inclusion of the risk factors model based on the Moolgavkar model on multistage carcinogenesis[8,9,10,11,12]. We have designed a model for the Mayo Clinic trial that includes the risk factors model where the risk of lung cancer is predicted based on smoking history as reported by trial participants instead of based on a fit of the age effect on SEER data and an elevated risk due to high smoking prevalence in the trial population. This MISCAN-lung model is still too tentative to present its results here but we have concluded that the model predicts background incidence very well. Because of similarity of the screen-detectable phase of the disease in both models, we expect that this model will closely reproduce the results as presented above but it will provide additional opportunity to study screening results by smoking history.

3. Model Validation by Simulation of CT Screening

We simulated the Early Lung Cancer Action Project (ELCAP)[13,14,15,16,17,18] for validation concerning low dose CT screening by comparing observed data from ELCAP to the results of the MISCAN simulation. ELCAP is a non-comparative observational study that is designed to evaluate baseline and annual repeat screening by low dose CT in 1,000 individuals with higher risk of lung cancer. The baseline screening found that among the whole study population, a positive result (defined as 1-6 non-calcified nodules) was found three times more commonly on low-dose CT than on CRX (23% [95% CI 21-26] vs 7% [5-9]). In the whole study population, malignant tumors were found four times more frequently on low-dose CT than on CRX; and stage I tumors were detected six times more frequently on low-dose CT than on CRX (2.3% [1.5-3.3] vs 0.4% [0.1-0.9])[13,14,17]. The initial findings on repeat screening found that

annual repetition of CT screening is sufficient to minimize interval cancers[15,17].

We have begun to adapt the x-ray screening model for simulating the ELCAP. The distinct characteristic of the ELCAP is that there is detailed information on tumor size; thus, it is possible to study the relationship between tumor size and curability of lung cancers. Currently, ELCAP is limited because there are not enough cancer cases yet. Our simulation model is not limited by number of cancer cases and therefore is able to achieve the study goal even if there are not yet real data available. We take this into account in our development of the model by assuming more disease states, which influence the size of the tumor. Since ELCAP is designed to compare the screening by low dose CT and chest radiographs, we introduced two screening policies to represent these two kinds of tests.

4. Simulations of CT Screening in the Mayo CT Project


...UNDER CONSTRUCTION...

5. Smoking Base Case: Effects of Anti-Smoking Campaigns on lung cancer mortality


...UNDER CONSTRUCTION...

Also refer to the specific implementations:

- Smoking Base Case14Mar06
- Smoking Base Case16Feb09

which describe the assumptions used in two versions of our model for the Smoking Base Case.

REFERENCES:

[1] Fontana, R.S., Sanderson, D.R., Taylor, W.F., Woolner, L.B., Miller, W.E., Muhm, J.R., Uhlenhopp, M.A. “Early lung cancer detection: results of initial (prevalence) radiologic and cytologic screening in the Mayo Clinic study.” in Am Rev Respir Dis 1984; 130: : 561-565

[2] Fontana, R.S., Sanderson, D.R., Woolner, L.B., Taylor, W.F., Miller, W.E., Muhm, J.R. “Lung cancer screening: the Mayo program.” in J Occup Med 1986; 28: : 746-750

[3] Fontana, R.S., Sanderson, D.R., Woolner, L.B., Taylor, W.F., Miller, W.E., Muhm, J.R., Bernatz, P.E., Payne, W.S., Pairolero PC, Bergstralh EJ. “Screening for lung cancer. A critique of the Mayo Lung Project.” in Cancer 1991; 67: 4 Suppl: 1155-64

[4] Flehinger, B.J., Kimmel, M., Polyak, T., Melamed, M.R. “Screening for lung cancer. The Mayo Lung Project revisited.” in Cancer 1993; 72: : 1573-1580

[5] Marcus, P.M., Prorok, P.C. “Reanalysis of the Mayo Lung Project data: the impact of confounding and effect modification.” in J Med Screen 1999; 6: : 47-49

[6] Marcus, P.M., Bergstralh, E.J., Fagerstrom, R.M., Williams, D.E., Fontana, R.S., Taylor, W.F., Prorok, P.C. “Lung cancer mortality in the Mayo Lung Project: impact of extended follow-up.” in J Natl Cancer Inst 2000; 92: 1308-1316

[7] Barton, R.R., Ivey, J.S.J. “Nelder-Mead simplex modifications for simulation optimization.” in Management Science 1996; 42: : 954-973

[8] Hazelton, W.D., Luebeck, E.G., Heidenreich, W.F., Moolgavkar, S.H. “Analysis of a historical cohort of Chinese tin miners with arsenic, radon, cigarette smoke, and pipe smoke exposures using the biologically based two-stage clonal expansion model.” in Radiat Res 2001; 156: : 78-94

[9] Heidenreich, W.F., Luebeck, E.G., Moolgavkar, S.H. “Some properties of the hazard function of the two-stage clonal expansion model.” in Risk Anal 1997; 17: : 391-399

[10] Luebeck, E.G., Heidenreich, W.F., Hazelton, W.D., Paretzke, H.G., Moolgavkar, S.H. “Biologially based analysis of the data for the Colorado uranium miners cohort: age, dose and dose-rate effects.” in Radiat Res 1999; 152: 339-351

[11] Luebeck, E.G., Moolgavkar, S.H. “Multistage carcinogenesis and the incidence of colorectal cancer.” in Proc Natl Acad Sci 2002; 99: : 15095-100

[12] Moolgavkar, S.H., Knudson, A.G. “Mutation and cancer: A model for human carcinogenesis.” in J Natl Cancer Inst 1981; 66: : 1037-1052

[13] Henschke, C.I., McCauley, D.I., Yankelevitz, D.F., Naidich, D.P., McGuinness, G., Miettinen, O.S., Libby, D.M., Pasmantier, M.W., Koizumi, J., Altorki, N.K., Smith, J.P. “Early Lung Cancer Action Project: overall design and findings from baseline screening.” in The Lancet 1999; 354: 9173: 99-105

[14] Henschke, C.I. “Early lung cancer action project - Overall design and findings from baseline screening.” in Cancer 2000; 89: 11: 2474-2482

[15] Henschke, C.I., Naidich, D.P., Yankelevitz, D.F., McGuinness, G., McCauley, D.I., Smith, J.P., Libby, D., Pasmantier, M., Vazquez, M., Koizumi, J., Flieder, D., Altorki, N., Miettinen, O.S. “Early Lung Cancer Action Project - Initial findings on repeat screening.” in Cancer 2001; 92: 1: 153-159

[16] Henschke, C.I., Yankelevitz, D.F., Libby, D.M., McCauley, D., Pasmantier, M., Altorki, N.K., Smith, J.P., Miettinen, O.S. “Early Lung Cancer Action Project - Annual screening using single-slice helical CT.” in Annals NYAS 2001; 952: : 124-134

[17] Henschke, C.I., McCauley, D.I., Yankelevitz, D.F., Naidich, D.P., McGuinness, G., Miettinen, O.S., Libby, D., Pasmantier, M., Koizumi, J., Altorki, N., Smith, J.P. “Early Lung Cancer Action Project: A summary of the findings on baseline screening.” in Oncologist 2001; 6: 2: 147-152

[18] Wisnivesky, J.P., Mushlin, A.I., Sicherman, N., Henschke, C. “The cost-effectiveness of low-dose CT screening for lung cancer - Preliminary results of baseline screening.” in Chest 2003; 124: 2: 614-621

SMOKING BASE CASE14MAR06

Summary
This document describes the MISCAN-lung model assumptions for the Smoking Base Case, from which we submitted the results to the CISNET program on 14 March 2006.

The Smoking Base Case involved four models: white males and white females, and including and excluding smoking effects on lung cancer risk.

Demography
The birth table used describes the probability distribution of being born before the start of the calendar year:

Birth Table	
Calendar Year	Cumulative Probability
1901	0
1906	0.0514
1911	0.1081
1916	0.1678
1921	0.23
1926	0.2942
1931	0.3551
1936	0.4108
1941	0.468
1946	0.5367
1951	0.6214
1956	0.7163
1961	0.8173
1966	0.9157
1971	1

Mortality from causes other than lung cancer is governed by the Smoking History Generator provided for the Smoking Base Case by NCI staff (see Smoking Generator Component).

Disease States
The disease model includes the following disease states:

Disease States
Normal, No Known Lung Cancer
Preclinical
Squamous Cell, Stage II-
Squamous Cell, Stage III+
Adeno/Large Cell, Stage II-
Adeno/Large Cell, Stage III+
Small Cell, Stage II-
Small Cell, Stage III+
Clinical
Squamous Cell, Stage II-
Squamous Cell, Stage III+
Adeno/Large Cell, Stage II-
Adeno/Large Cell, Stage III+
Small Cell, Stage II-
Small Cell, Stage III+
Screen-Detected
Squamous Cell, Stage II-
Squamous Cell, Stage III+
Adeno/Large Cell, Stage II-
Adeno/Large Cell, Stage III+
Small Cell, Stage II-
Small Cell, Stage III+
End States
Death from Lung Cancer
Death from Other Causes

Risk Factors Model

The model includes exposure to one risk factor: cigarette smoking. The exposure to this risk factor is governed by the Smoking History Generator, provided for the Smoking Base Case by NCI staff (see Smoking Generator Component).

Based on the current smoking status over the course of the life history, the model assumes the following parameters for the development of lung cancer:

When not smoking: Clones of initiated cells start at a size of 200 cells; the rate of initiation is 0.007015 per year; the exponential growth rate of the clones of initiated cells is 0.0751; and the rate of malignant transformation to lung cancer is $1.403 \cdot 10^{-7}$. These base rates are derived from CPS I[1].

When smoking, the when-not-smoking rates are multiplied by the following factors, depending on the dose expressed in cigarettes per day:

Rate ratios by smoking dose relative to when not smoking			
Cigarettes per Day	Initiation	Promotion	Malignant Transformation
	4.532	1.323	4.532
	4.532	1.541	4.532
	4.532	1.688	4.532
	4.532	1.806	4.532
	4.532	1.907	4.532
	4.532	1.997	4.532
	4.532	2.078	4.532
	4.532	2.153	4.532
	4.532	2.223	4.532

The assumptions for the risk factor model are based on[1].

Preclinical lung cancer

Of the malignant transformations that are generated, 35.7% become squamous cell carcinoma, 44.0% adeno or large cell carcinoma, and 20.3% small cell carcinoma.

Dwelling times [in years] and stage distribution				
Cell type	Constant lag time from malignant transformation to screen-detectable preclinical cancer	Mean dwelling time in preclinical stage I-II	Mean dwelling time in preclinical stage III-IV	Percentage clinical diagnosis in stage I-II
Squamous	7.2 years	1.36	2.82	29.2%
Adeno/Large	8.2 years	1.36	2.82	30.0%
Small	5.9 years	0.39	1.11	9.4%

The assumptions on dwelling times for preclinical disease states are based on our model estimates of the Mayo Lung Project[2].

Survival from lung cancer

Survival from lung cancer is modeled as a probability of long term cause specific survival, and for the remaining cancers, a Weibull distribution for the time from clinical diagnosis to death from lung cancer. This cause specific cancer survival is superseded if death from causes other than lung cancer is earlier than death from lung cancer.

Survival from lung cancer				
			Weibull distribution for time from clinical diagnosis to lung cancer death	
Cell type	Stage at diagnosis	Long term survival	Mean	Shape
Squamous	I-II	0.180	2.419	0.573
Squamous	III-IV	0.060	0.752	0.641
Adeno/ Large	I-II	0.290	4.783	0.676
Adeno/ Large	III-IV	0.050	0.674	0.607
Small	I-II	0.080	1.049	0.727
Small	III-IV	0.010	0.507	0.738

Screening

The four models for the Smoking Base Case did NOT include any screening.

Model variants

The models for white males and white females are only different with respect to the exposure to cigarette smoking as determined by the Smoking History Generator (see Smoking Generator Component).

The models for no smoking effect are only different from the specification above with respect to the table "Rate ratios by smoking dose relative to when not smoking" where the rate ratios are all 1.

REFERENCES:

1 Hazelton, W.D., Clements, M.S., Moolgavkar, S.H. “Multistage carcinogenesis and lung cancer mortality in three cohorts.” in Cancer Epidemiol Biomarkers Prev. 2005; 14: 5: 1171-81

2 Fontana, R.S., Sanderson, D.R., Woolner, L.B., Taylor, W.F., Miller, W.E., Muhm, J.R., Bernatz, P.E., Payne, W.S., Pairolero PC, Bergstralh EJ. “Screening for lung cancer. A critique of the Mayo Lung Project.” in Cancer 1991; 67: 4 Suppl: 1155-64

SMOKING BASE CASE16FEB09

Summary

This document describes the MISCAN-lung model assumptions for the Smoking Base Case, from which we submitted the results to the CISNET program on 16 Feb 2009.

The Smoking Base Case involves four populations: white males, white females, all races males and all races females and considers the U.S. population aged 30-84 y in the calendar years 1975-2000. Three tobacco control scenarios are evaluated, i.e. actual tobacco control (TC), no tobacco control (NTC counterfactual) and complete tobacco control (CTC counterfactual, assuming everybody stopped smoking in 1965).

Demography

Birth tables used describe the probability distribution of being born before the start of a specific calendar year.

The original range of birth years comprised 1900-1970 in five-year bins. As this leads to an incomplete age range (30-84) in the calendar years 1975-1984, the range of birth years was later extended to include 1890-1900 (for all races only).

The birth tables are based on U.S. population data[1]. For the extended birth year range, small adjustments were made by trial and error to improve agreement between the age distributions as calculated by MISCAN-lung for the calendar years 1975, 1986 and 2000 and the age distributions observed in the U.S. all races male and female populations in those years[2].

Birth Table					
Calendar Year	Cumulative Probability				
	Original		Extended		
	Whites	All races	All races	male	female
1891	0	0	0	0	0
1896	0	0	0.0548	0.0449	0.0465
1901	0	0	0.1095	0.0969	0.0957
1906	0.0619	0.0615	0.1643	0.1517	0.1503
1911	0.1241	0.1233	0.2193	0.2122	0.2063
1916	0.1906	0.1887	0.2776	0.2704	0.2650
1921	0.2578	0.2545	0.3362	0.3290	0.3224
1926	0.3253	0.3212	0.3956	0.3866	0.3816
1931	0.3872	0.3824	0.4501	0.4400	0.4360
1936	0.4419	0.4368	0.4985	0.4884	0.4849
1941	0.4977	0.4924	0.5480	0.5379	0.5358
1946	0.5645	0.5585	0.6069	0.5979	0.5968
1951	0.6476	0.6406	0.6800	0.6710	0.6698
1956	0.7373	0.7305	0.7600	0.7510	0.7505
1961	0.8324	0.8267	0.8457	0.8367	0.8377
1966	0.9212	0.9187	0.9276	0.9218	0.9227
1971	1	1	1	1	1

Mortality from causes other than lung cancer is governed by the data tables of the Smoking History Generator Application (see Smoking Generator Component) provided for the Smoking Base Case by NCI staff.

Estimation of U.S. population size for the counterfactual scenarios of No Tobacco Control and Complete Tobacco Control.

For persons of a certain age, a, in a given year, yr, the expected number of lung cancer deaths, $D(a, yr)$, can be calculated from the model results (number of simulated lung cancer deaths, $Ds(a, yr)$; size of simulated population, $Ns(a, yr)$ and the actual U.S. population size, $N(a, yr)$:

$$D(a, yr) = \frac{Ds(a, yr)}{Ns(a, yr)} \times N(a, yr)$$

The latter quantity is known from observations for the actual tobacco control scenario in the U.S. but not for the two counterfactual scenarios as the latter never happened in reality.
Therefore, the population size in the case of for instance complete tobacco control, $N_{cc}(a, yr)$, is estimated as follows:

$$N_{cc}(a, yr) = \frac{Ns_{cc}(a, yr)}{Ns(a, yr)} \times N(a, yr)$$

where $Ns_{cc}(a, yr)$ is the population size resulting from the simulation of the complete tobacco control scenario.
Under this scenario, the estimated number of lung cancer deaths, $D_{cc}(a, yr)$, becomes:

$$D_{cc}(a, yr) = \frac{Ds_{cc}(a, yr)}{Ns_{cc}(a, yr)} \times N_{cc}(a, yr) = \frac{Ds_{cc}(a, yr)}{Ns(a, yr)} \times N(a, yr)$$

where $Ds_{cc}(a, yr)$ is the number of simulated lung cancer deaths under the complete tobacco control scenario.
Similar reasoning holds for the no tobacco control scenario.

Disease States
The disease model includes the following disease states:

Disease States
Normal, No Known Lung Cancer
Preclinical
Squamous Cell, Stage II-
Squamous Cell, Stage III+
Adeno/Large Cell, Stage II-
Adeno/Large Cell, Stage III+
Small Cell, Stage II-
Small Cell, Stage III+
Clinical
Squamous Cell, Stage II-
Squamous Cell, Stage III+
Adeno/Large Cell, Stage II-
Adeno/Large Cell, Stage III+
Small Cell, Stage II-
Small Cell, Stage III+
Screen-Detected
Squamous Cell, Stage II-
Squamous Cell, Stage III+
Adeno/Large Cell, Stage II-
Adeno/Large Cell, Stage III+
Small Cell, Stage II-
Small Cell, Stage III+
End States
Death from Lung Cancer
Death from Other Causes

Risk Factors Model

The model includes exposure to one risk factor: cigarette smoking. The exposure to this risk factor is governed by the data tables of the Smoking History Generator Application, provided for the Smoking Base Case by NCI staff (see Smoking Generator Component). MISCAN-lung reads those tables to produce appropriate random individual smoking histories for the simulated persons.

Based on the current smoking status over the course of the life history, the model assumes the following parameters for the development of lung cancer:

When NOT smoking: Clones of initiated cells start at a size of 80 or 30 cells for males or females, respectively; the rate of initiation is 0.024 or 0.036 per year; the exponential growth rate of the clones of initiated cells is 0.0973; and the rate of malignant transformation to lung cancer is $7.58 \cdot 10^{-8}$. These base rates are derived from HPFS or NHS for males or females, respectively[3].

When smoking, the when-not-smoking rates are multiplied by the following factors, depending on the dose expressed in cigarettes per day:

Rate ratios by smoking dose relative to when not smoking						
Cigarettes per Day	Males			Females		
	Initiation	Promotion	Malignant Transformation	Initiation	Promotion	Malignant Transformation
	1.0	1.1810	1.7804	1.0	1.2322	1.3026
	1.0	1.3208	2.2976	1.0	1.4116	1.5031
	1.0	1.4186	2.6437	1.0	1.5371	1.6373
	1.0	1.4987	2.9206	1.0	1.6400	1.7446
	1.0	1.5685	3.1575	1.0	1.7295	1.8365
	1.0	1.6311	3.3675	1.0	1.8099	1.9179
	1.0	1.6885	3.5578	1.0	1.8835	1.9917
	1.0	1.7418	3.7329	1.0	1.9519	2.0596
	1.0	1.8157	3.9735	1.0	2.0467	2.1528

The assumptions for the risk factors model are based on[4] and the newer data from[3].

Preclinical lung cancer

Of the malignant transformations that are generated, 35.7% become squamous cell carcinoma, 44.0% adeno or large cell carcinoma, and 20.3% small cell carcinoma.

Dwelling times [in years] and stage distribution				
Cell type	Constant lag time from malignant transformation to screen-detectable preclinical cancer	Mean dwelling time in preclinical stage I-II	Mean dwelling time in preclinical stage III-IV	Percentage clinical diagnosis in stage I-II
Squamous	0.01	1.36	2.82	29.2%
Adeno/Large	0.01	1.36	2.82	30.0%
Small	0.01	0.39	1.11	9.4%

The assumptions on dwelling times for preclinical disease states are based on our model estimates of the Mayo Lung Project[5].

Survival from lung cancer

Survival from lung cancer is modeled as a probability of long term cause specific survival, and for the remaining cancers, a Weibull distribution for the time from clinical diagnosis to death from lung cancer. This cause specific cancer survival is superseded if death from causes other than lung cancer is earlier than death from lung cancer.

Survival from lung cancer				
Weibull distribution for time from clinical diagnosis to lung cancer death				
Cell type	Stage at diagnosis	Long term survival (fraction)	Mean (y)	Shape
Squamous	I-II	0.180	2.419	0.573
Squamous	III-IV	0.060	0.752	0.641
Adeno/Large	I-II	0.290	4.783	0.676
Adeno/Large	III-IV	0.050	0.674	0.607
Small	I-II	0.080	1.049	0.727
Small	III-IV	0.010	0.507	0.738

Screening

The MISCAN-lung models for the Smoking Base Case do NOT include any screening.

Model variants

The models for the various population categories are only different with respect to the exposure to cigarette smoking as determined from the data tables of the Smoking History Generator Application (see Smoking Generator Component).The data tables provided cover the Tobacco Control and No Tobacco Control scenarios. The Smoking History Generator Application includes an option to calculate smoking histories in case of Complete Tobacco Control, when nobody smokes after the start of a given year (e.g. 1965).

As MISCAN-lung computes its own sets of smoking histories, for the scenario of Complete Tobacco Control we amended the data tables ourselves: the probabilities to start smoking in or after 1965 were set to zero; the probabilities to stop smoking in or after 1965 were set to one; the corresponding smoking intensities were set to zero cpd; and the probabilities of death from other causes in that time period were set equal to those for never smokers.

REFERENCES:

[1] Carter, S.B., et al., (eds.). "Historical Statistics of the United States. Millennial Edition. Vol 1. Cambridge University Press. " 2006;

[2] Gibson, T. "[General Use Data] Single Age Populations (0-84, 85+) for US (1969-2000) and SEER9 (1973-2000) by Sex and Race (White, Black, Other)." 2005;

[3] Meza, R., Hazelton, W.D., Colditz, G.A., Moolgavkar, S.H. "Analysis of lung cancer incidence in the nurses' health and the health professionals' follow-up studies using a multistage carcinogenesis model." in Cancer Causes Control 2008; 19: : 317-328

[4] Hazelton, W.D., Clements, M.S., Moolgavkar, S.H. "Multistage carcinogenesis and lung cancer mortality in three cohorts." in Cancer Epidemiol Biomarkers Prev. 2005; 14: 5: 1171-81

[5] Fontana, R.S., Sanderson, D.R., Woolner, L.B., Taylor, W.F., Miller, W.E., Muhm, J.R., Bernatz, P.E., Payne, W.S., Pairolero PC, Bergstralh EJ. "Screening for lung cancer. A critique of the Mayo Lung Project." in Cancer 1991; 67: 4 Suppl: 1155-64

BACKGROUND INFORMATION MODEL OVERVIEW

Background Information (Model Overview)

SUMMARY
This document provides background information on various aspects of lung cancer that are of relevance for modeling the disease.

BACKGROUND
Lung cancer is a major health problem in the United States, even though the age-adjusted incidence and mortality rates of lung cancer have been decreasing each year, since around 1992, by an average of 1.8% for men and 0.6% for women. It is generally assumed that most of this population trend is due to changes in smoking behavior but this assumption is so far not quantified and there are no good tools to predict future trends other than extrapolating observed trends.

Risk factors
Exposure to tobacco smoke is by far the most serious risk factor for cancer of the lung and bronchi. An estimated 87% of lung cancer deaths in 2003 are attributable to active smoking. Other risk factors include exposure to second-hand smoke (passive smoking); radon (a naturally occurring air pollutant); asbestos; and diet.
In addition to tobacco, poor dietary quality has been related to lung cancer. Expert consensus suggests that as much as 20-30% of lung cancers are attributable to a poor quality diet.

There is an intricate relationship between exposure to risk factors and resulting risk of lung cancer. Generally, this relationship is described by empirical studies as a relative or excess risk among those exposed to a risk factor with respect to those who were not exposed, or as a comparison between different levels of exposure. Very few report on possible mechanisms that explain the timing from exposure to expression. There is general agreement that exposure to tobacco smoke leads to a very strong increase in risk for lung cancer and that the time from exposure to lung cancer can be several decades.

Multistage Carcinogenesis
The concept of multistage carcinogenesis provides a possible explanation of the long duration from exposure to expression. Carcinogenesis proceeds through at least the following stages:

Initiation: Initiation is the process in which a single somatic cell undergoes non-lethal, but heritable, mutation. The initiated cell can escape cellular regulatory mechanisms.
Promotion: Promotion is the process in which the initiated cell is exposed to a tumor promoter that causes phenotypical clonal expansion. Tumor promoters are either external or internal stimuli and stimulate growth of initiated cells.
Malignant conversion: During malignant conversion or transformation, cellular growth is further deregulated. Like initiation, this step requires genetic alteration.
Progression: During this stage, cellular growth is further deregulated and proceeds

uncontrolled. Progression is probably the most complex stage, because both acquired genetic and phenotypic alterations occur, and cellular expansion is rapid.

The first quantitative mechanistic model concerning carcinogenesis was published by Armitage and Doll (1954). Subsequent models incorporate new knowledge concerning multistage carcinogenesis, such as the clonal expansion occurring during the promotion stage. Dr. Moolgavkar at the Fred Hutchinson Cancer Research Center is at the forefront of this model development.
The Moolgavkar model summarizes the promotion stage in a single step and agrees very well with observed epidemiological evidence. However, current versions of the Moolgavkar model do not account for the progression stage in any detail. This stage is particularly important for evaluation of early detection of (lung) cancer.

Early detection
Since neither primary prevention nor treatment has had a satisfying impact on lung cancer incidence or mortality, secondary prevention (screening of asymptomatic individuals) remains a topic of great interest. Because lung cancer is usually diagnosed based on symptoms, the disease is usually so far advanced that curative therapy is not possible. Screening has the potential to detect lung cancer at earlier stages, when survival rates are considerably higher.

It is anticipated that new technology for lung cancer screening, particularly CT screening, will make it possible to better detect aggressive cancers early enough to be curable. Improved screening has the potential to prevent thousands of lung cancer deaths annually.

Diagnosis and therapy
New lung cancer therapies unfortunately have not had a substantial impact on mortality so far. Most clinical trials did not show major improvements in survival, and population-based survival from a diagnosis of lung cancer has improved only slightly over the past few decades.

However, there is a striking variation in the treatment of lung cancer that raises concerns about disparities in the care of patients of different racial/ethnic groups and advanced age. Several population-based studies have found that black and Hispanic patients are less likely to undergo potentially curative surgical resection for early stage non-small cell lung cancer than white patients, even when controlling for differences in comorbid illness and age. In addition to these variations according to patients' racial/ethnic background, many studies have demonstrated a marked decline in the use of curative treatments with increasing patient age.

Go back to Model Overview

KEY REFERENCES

Barton, R.R., Ivey, J.S.J. (1996) Nelder-Mead simplex modifications for simulation optimization. in *Management Science* 42, p 954-973

Carter, S.B., et al., (eds.). (2006) Historical Statistics of the United States. Millennial Edition. Vol 1. Cambridge University Press.

,

Draisma, G., Boer, R., Otto, S.J., van der Cruijsen, I.W., Damhuis, R.A., Schroder, F.H., de Koning, H.J. (2003) Lead times and overdetection due to prostate-specific antigen screening: estimates from the European Randomized Study of Screening for Prostate Cancer.
in *J Natl Cancer Inst* 95:12, p 868-78

Flehinger, B.J., Kimmel, M., Polyak, T., Melamed, M.R. (1993) Screening for lung cancer. The Mayo Lung Project revisited. in *Cancer* 72, p 1573-1580

Fontana, R.S., Sanderson, D.R., Taylor, W.F., Woolner, L.B., Miller, W.E., Muhm, J.R., Uhlenhopp, M.A. (1984) Early lung cancer detection: results of initial (prevalence) radiologic and cytologic screening in the Mayo Clinic study. in *Am Rev Respir Dis* 130, p 561-565

Fontana, R.S., Sanderson, D.R., Woolner, L.B., Taylor, W.F., Miller, W.E., Muhm, J.R. (1986) Lung cancer screening: the Mayo program. in *J Occup Med* 28, p 746-750

Fontana, R.S., Sanderson, D.R., Woolner, L.B., Taylor, W.F., Miller, W.E., Muhm, J.R., Bernatz, P.E., Payne, W.S., Pairolero PC, Bergstralh EJ. (1991) Screening for lung cancer. A critique of the Mayo Lung Project. in *Cancer* 67:4 Suppl, p 1155-64

Gibson, T. (2005) [General Use Data] Single Age Populations (0-84, 85+) for US (1969-2000) and SEER9 (1973-2000) by Sex and Race (White, Black, Other).,

Habbema, J.D., van Oortmarssen, G.J., Lubbe, J.T., van der Maas, P.J. (1985) The MISCAN simulation program for the evaluation of screening for disease. in *Comput Methods Programs Biomed* 20:1, p 79-93

Hazelton, W.D., Clements, M.S., Moolgavkar, S.H. (2005) Multistage carcinogenesis and lung cancer mortality in three cohorts. in *Cancer Epidemiol Biomarkers Prev.* 14:5, p 1171-81

Hazelton, W.D., Luebeck, E.G., Heidenreich, W.F., Moolgavkar, S.H. (2001) Analysis of a historical cohort of Chinese tin miners with arsenic, radon, cigarette smoke, and pipe smoke exposures using the biologically based two-stage clonal expansion model. in *Radiat Res* 156, p 78-94

Heidenreich, W.F., Luebeck, E.G., Moolgavkar, S.H. (2004) Effects of exposure uncertainties in the TSCE model and application to the Colorado miners data. in *Radiat Res* 161:1, p 72-81

Heidenreich, W.F., Luebeck, E.G., Moolgavkar, S.H. (1997) Some properties of the hazard function of the two-stage clonal expansion model. in *Risk Anal* 17, p 391-399

Henschke, C.I. (2000) Early lung cancer action project - Overall design and findings from baseline screening. in *Cancer* 89:11, p 2474-2482

Henschke, C.I., McCauley, D.I., Yankelevitz, D.F., Naidich, D.P., McGuinness, G., Miettinen, O.S., Libby, D., Pasmantier, M., Koizumi, J., Altorki, N., Smith, J.P. (2001) Early Lung Cancer Action Project: A summary of the findings on baseline screening. in *Oncologist* 6:2, p 147-152

Henschke, C.I., McCauley, D.I., Yankelevitz, D.F., Naidich, D.P., McGuinness, G., Miettinen, O.S., Libby, D.M., Pasmantier, M.W., Koizumi, J., Altorki, N.K., Smith, J.P. (1999) Early Lung Cancer Action Project: overall design and findings from baseline screening. in *The Lancet* 354:9173, p 99-105

Henschke, C.I., Naidich, D.P., Yankelevitz, D.F., McGuinness, G., McCauley, D.I., Smith, J.P., Libby, D., Pasmantier, M., Vazquez, M., Koizumi, J., Flieder, D., Altorki, N., Miettinen, O.S. (2001) Early Lung Cancer Action Project - Initial findings on repeat screening. in *Cancer* 92:1, p 153-159

Henschke, C.I., Yankelevitz, D.F., Libby, D.M., McCauley, D., Pasmantier, M., Altorki, N.K., Smith, J.P., Miettinen, O.S. (2001) Early Lung Cancer Action Project - Annual screening using single-slice helical CT. in *Annals NYAS* 952, p 124-134

Loeve, F., Boer, R., van Oortmarssen, G.J., van Ballegooijen, M., Habbema, J.D. (1999) The MISCAN-COLON simulation model for the evaluation of colorectal cancer screening. in *Comput Biomed Res* 32:1, p 13-33

Loeve, F., Brown, M.L., Boer, R., van Ballegooijen, M., van Oortmarssen, G.J., Habbema, J.D. (2000) Endoscopic colorectal cancer screening: a cost-saving analysis. in *J Natl Cancer Inst* 92:7, p 557-63

Luebeck, E.G., Heidenreich, W.F., Hazelton, W.D., Paretzke, H.G., Moolgavkar, S.H. (1999) Biologially based analysis of the data for the Colorado uranium miners cohort: age, dose and dose-rate effects. in *Radiat Res*:152, p 339-351

Luebeck, E.G., Moolgavkar, S.H. (2002) Multistage carcinogenesis and the incidence of colorectal cancer. in *Proc Natl Acad Sci* 99, p 15095-100

Marcus, P.M., Bergstralh, E.J., Fagerstrom, R.M., Williams, D.E., Fontana, R.S., Taylor, W.F., Prorok, P.C. (2000) Lung cancer mortality in the Mayo Lung Project: impact of extended follow-up. in *J Natl Cancer Inst*:92, p 1308-1316

Marcus, P.M., Prorok, P.C. (1999) Reanalysis of the Mayo Lung Project data: the impact of confounding and effect modification. in *J Med Screen* 6, p 47-49

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

Meza, R., Hazelton, W.D., Colditz, G.A., Moolgavkar, S.H. (2008) Analysis of lung cancer incidence in the nurses' health and the health professionals' follow-up studies using a multistage carcinogenesis model. in *Cancer Causes Control* 19, p 317-328

Moolgavkar, S.H., Knudson, A.G. (1981) Mutation and cancer: A model for human carcinogenesis. in *J Natl Cancer Inst* 66, p 1037-1052

Wisnivesky, J.P., Mushlin, A.I., Sicherman, N., Henschke, C. (2003) The cost-effectiveness of low-dose CT screening for lung cancer - Preliminary results of baseline screening. in *Chest* 124:2, p 614-621

de Koning, H.J., Boer, R., Warmerdam, P.G., Beemsterboer, P.M., van der Maas, P.J. (1995) Quantitative interpretation of age-specific mortality reductions from the Swedish breast cancer-screening trials. in *J Natl Cancer Inst* 87:16, p 1217-23

van den Akker-van Marle, M.E., van Ballegooijen, M., van Oortmarssen, G.J., Boer, R., Habbema, J.D. (2002) Cost-effectiveness of cervical cancer screening: comparison of screening policies. in *J Natl Cancer Inst* 94:3, p 193-204

FLEXKB DOCUMENT
Version: HI.001.03302011.70013
Document generated: 03/30/2011

FRED HUTCHINSON CANCER RESEARCH CENTER (FHLUNG)

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.

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

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

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
- Population Component
- Natural History 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.

MODEL PURPOSE

SUMMARY

This document provides a brief overview of the Fred Hutchinson Cancer Research Center (FHCRC) lung cancer model. This model uses an underlying biologically based multistage model (with additional age, period and cohort effects) to represent the effects of smoking on the natural history of lung cancer. The FHCRC lung cancer model may be utilized to address questions about the impact of public health information on US lung cancer trends, and to predict the hypothetical impact of alternative tobacco control policies.

PURPOSE

The purpose of the FHCRC lung cancer model is to serve as an effective tool for evaluating lung cancer trends in the US population, and the effects of possible interventions. The FHCRC lung cancer model combines an underlying biologically based natural history model that is calibrated to individual smoking histories in substantial US lung cancer mortality cohorts with additional age, period, and birth cohort effects to improve the calibration to US lung cancer mortality data. Limitations include that it is not calibrated to US lung cancer incidence or CT screening cohort data.

Two distinct modeling projects contributed to the development of the FHCRC lung cancer model. The first project consisted of calibrating a biologically based natural history model to individual smoking histories in several substantial lung cancer cohorts. The calibrated model parameters (including background and smoking dose response parameters) were shared with other Cancer Intervention and Surveillance Modeling Network (CISNET) lung cancer modeling groups. The second project (called the Lung Smoking Base Case) consists of combining the biologically based natural history model with additional age, period, and birth cohort effects. This model is calibrated to lung cancer deaths in the US population by single years of age and calendar year. It should be of use in evaluating the effects of alternative tobacco control policies, and in making projections of future US lung cancer mortality.

Development of the FHCRC lung cancer model took place in two distinct projects:

Project One : Calibrating a natural history model to smoking cohort data
Project Two : Lung Smoking Base Case - Modeling US lung cancer mortality trends

MODEL OVERVIEW

SUMMARY

This document describes previous work leading to this model and model itself in general terms.

PURPOSE

We wish to understand the effects of smoking and other factors on US lung cancer mortality. We modeled the effects of smoking using the biologically based two-stage clonal expansion (TSCE) model (Moolgavkar et al. 1979, 1981, 1990; Heidenreich et al., 1997). The TSCE model relates smoking to biological rates for cell initiation, promotion, and malignant conversion processes. The effects of additional unknown factors that may have influenced US lung cancer mortality rates were modeled using period and cohort effects.

BACKGROUND

Lung cancer is the leading cause of cancer death in the US, and smoking is the most important risk factor for developing lung cancer. Thus in modeling lung cancer in the US, we felt it was important to use the best available methods to relate smoking to lung cancer risk. The biologically based TSCE model seemed best for this purpose.

The FHCRC lung cancer project began with calibration of the TSCE model to several large smoking cohorts, modeling individual smoking histories in relation to lung cancer incidence and mortality [Project One]. This was followed by the Lung Smoking Base Case [Project Two] in which we used the calibrated TSCE model to represent effects of smoking on lung cancer mortality in the US population. We also introduced additional corrections as a function of period and birth cohort to improve the fit to US lung cancer mortality.

The TSCE model was initially developed by Moolgavkar, Venzon, and Knudson. This model has been applied to analyze many types of cancer, including the effects of smoking and other exposures. Calibrating the TSCE model to cohort data consists of estimating dose-response relationships for these exposures as they affect cell initiation, promotion, and malignant conversion rates. Maximum likelihood methods allow optimization of the model to represent temporal patterns of risk associated with different exposure histories of individuals in the cohort.

The TSCE natural history model represents basic cellular processes, including cell division, apoptosis, and mutation, that contribute to three distinct phases in the carcinogenic process: initiation, promotion (birth minus death of initiated cells) and malignant conversion (TSCEModel Details). The TSCE model represents a significant simplification of the biological processes associate with lung cancer. The model ignores the possibility of multiple cancer pathways and disease subtypes. However, it does provide a rigorous mathematical representation of processes that are considered as the

rate-limiting events in carcinogenesis, and has provided excellent fits to individual and population data for many cancer types (Moolgavkar and Luebeck, 2003).

The FHCRC lung cancer model builds on previous analyses using the TSCE model to relate lung cancer risk to individual exposure patterns for smoking, radon, arsenic, fibers, and radiation (Castren et al., 1999; Hazelton et al., 2001, 2005, 2006; Haylock et al., 2004; Heidenreich et al, 2002; Kai et al, 1997; Little et al., 2002; Luebeck et al., 1999, 2000; Meza et al., 2008; Moolgavkar et al., 1989, 1993, 1998, 1999, 2000, 2001a, 2001b, 2001c; Stevens et al., 1979, 1984). These analyses consistently show that the most important lung cancer risk factor is tobacco smoke, with the risk increasing non-linearly with smoking duration (Hazelton et al., 2005; Meza et al., 2008).

The FHCRC lung cancer model was applied to the Smoking Base Case [Project Two], using the TSCE model to represent effects of smoking, and period and cohort effects to represent other unknown factors. Inputs included US population data and lung cancer deaths for males and females binned by single year of age and calendar year, and a smoking history generator developed by NCI to simulate smoking histories and other cause mortality for individuals in the US. Outputs are estimates of lung cancer deaths by gender, age, and calendar year given historical smoking patterns, and also counter-factual estimates for lung cancer deaths given alternative US smoking patterns.

Model Description

The FHCRC lung cancer model consists of a biologically based TSCE natural history model of the effects of smoking on lung cancer morality, along with period and birth cohort effects to represent lung cancer mortality in the US population (called the TSCE-PC model, representing the age effects given by the TSCE natural history model, along with period and birth cohort effects). A second model (called the TSCE-APC model) includes additional age effects to capture possible discrepancies between age effects in the TSCE calibration to US lung cancer mortality, and compensate for possible limitations of the TSCE model in representing the effects of tobacco smoke on lung cancer mortality.

For more details see: TSCEModel Details

Contributors

William D Hazelton
Jihyoun Jeon
Rafael Meza
Suresh Moolgavkar

ASSUMPTION OVERVIEW

SUMMARY

This document discusses assumptions underlying the model as well as some of their implications.

BACKGROUND

The TSCE natural history model relates individual non-smoking and smoking histories to cellular processes that contribute to the development of lung cancer. The TSCE model allows calculation of the time-dependent probability for lung cancer mortality at each age. The TSCE model includes an initial rate for mutation or epi-genetic change leading to initiated cells, a birth rate and death rate for the initiated cells, and a rate for second mutation or epi-genetic change that occurs during the cell division process of an initiated cell to generate a malignant cell as well as another initiated cell. A lag time or lag time distribution is used to represent the time from the first malignant cell to cancer incidence or mortality from cancer. Smoking is assumed to affect any or all of the rates through flexible dose-response functions.

To assess model reliability (Project One), we looked at consistency of model parameter estimates between the different lung cancer mortality cohorts. We also analyzed separately lung cancer incidence by subtype in the NHS and HPFS cohorts, where that data is available.

In working on the Lung Smoking Base Case - Project Two, the FHCRC group found that the TSCE natural history model, when calibrated to cohort data for lung cancer in relation to smoking, does not fully account for lung cancer mortality in the US population. Thus the FHCRC group found it necessary to include additional age, period, and birth cohort adjustments that may correct for limitations of the TSCE natural history model, for changing cigarette composition, and other exposures and environmental factors that contribute to lung cancer in the US population. The FHCRC lung cancer model combines the TSCE natural history model for smoking with these additional age, period, and birth cohort adjustments.

ASSUMPTION LISTING

1. The TSCE natural history model (used in the Project One and Project Two by the FHCRC group) assumes two stochastic rate-limiting mutation events, clonal expansion of initiated cells, and a lag time from the first occurrence of a malignant cell to the time of lung cancer death.
2. Calibration of the TSCE model to the HPFS and NHS lung cancer mortality data (see Project One and Calibration and Validation sections) and subsequent estimation of lung cancer deaths for the Lung Smoking Base Case included a model assumption that there are a fixed number (10^7) normal stem cells in lung. Clearly the number must increase during embryogenesis, and any trend throughout life has not been ascertained, nor has the total number, as lung stem cells are difficult to identify. However, the likelihood based modeling approach will estimate an initiation rate that will compensate any error on the assumed number of stem cells at risk for initiation.

3. The background rates for initiation and malignant conversion rates set equal to each other, $\nu = \mu$. This assures identifiability of the TSCE model parameters.
4. Normal stem cells may undergo faulty division to create an initiated cell.
5. Growth of the population of initiated cells (promotion) is modeled stochastically through cell birth and death process. This process can not be observed directly, but is consistently estimated between different cohorts as the most important mechanism whereby cigarettes influence the risk of lung cancer.
6. Malignant conversion in the TSCE model is assumed equivalent to first occurrence of a malignant cell arising through faulty division of an initiated cell.
7. A lag time or lag time distribution is assumed to represent the time between the occurrence of the first malignant cell and cancer death.
8. The effects of cigarette smoke is modeled as a constant dose rate during periods of smoking, and the smoking dose has separate non-linear (power-law) influences on the initiation, promotion, and malignant conversion rates in the TSCE model.
9. The smoking dose-response from the HPFS/NHS calibration is applicable to the simulated US population data, given subsequent adjustments for additional age, period, and birth cohort.
10. The SHG provided by NCI provides individual smoking histories consistent with the historical patterns of smoking in the US.
11. The other cause mortality input provided by the SHG reflects historical trends in the US population.
12. An additional age effect may correct for deficiencies of the TSCE model, and differences between age effects in the calibration cohort and the US population.
13. A period effect is adequate to account for historical changes in cigarette composition, demographic changes, and changes in health care that may influence lung cancer mortality.
14. A birth cohort effect is able to capture effects of environmental, nutritional, and other factors that influences the lifetime risk for lung cancer.

PARAMETER OVERVIEW

SUMMARY

This document provides an overview of the major parameters in the model, their sources, and general implications they have on model outputs.

BACKGROUND

Parameters related to natural history for the Lung Smoking Base Case (Project Two) were estimated from the TSCE model calibration to lung cancer mortality in smoking cohorts (Project One), focusing on the HPFS and NHS cohorts. Additional age, period and birth cohort parameters were estimated using US population and lung cancer mortality data. Additional demographic parameters are embedded in the SHG to reflect historical smoking trends and rates for other cause mortality.

PARAMETER LISTING OVERVIEW

The FHCRC lung cancer model parameters are categorized into:

1. Background parameters of the TSCE model (Natural History Component).
2. Dose-response parameters that relate current cigarette smoke exposure to the rates for initiation, promotion (birth minus death of initiated cells) and malignant conversion (Natural History Component).
3. Lag time parameters that describe the time lag or gamma lag time distribution from the first malignant cell to cancer death (See Natural History Component and Survival Mortality Component).
4. Additional demographic parameters used in the Lung Smoking Base Case (Project Two): additional age, period, and birth cohort parameters applied to the simulated US population (Population Component) to adjust the lung cancer mortality calculations to represent US lung cancer deaths (Survival Mortality Component).

The additional age, period, and birth cohort parameters are based on:

1. Smoothing parameters used to generate single year US population data from census data
2. Smoothing parameters used to generate annual US lung cancer mortality data

BACKGROUND PARAMETERS:

1. Background initiation, malignant conversion rate
2. Background initiated cell division rate
3. Background net initiated cell promotion rate

DOSE-RESPONSE PARAMETERS FOR FULL MODEL:

1. Coefficient multiplying dose response for initiation
2. Power of dose for initiation

3. Coefficient multiplying dose response for promotion
4. Power of dose for promotion
5. Coefficient multiplying dose response for malignant conversion
6. Power of dose for malignant conversion

NOTE 1. Typically only three or four dose-response parameters are required to model lung cancer incidence or mortality due to cigarettes - two for the dominant effect of promotion, and one or two describing the much smaller effect on initiation or malignant conversion.

LAG TIME PARAMETERS:

Mean and standard deviation for gamma distribution, or fixed lag time

Component Overview

Summary

This is a description of the basic computational building blocks/components of the model.

Overview

Several components are involved to construct the FHCRC lung cancer model for the Lung Smoking Base Case. A Population Component uses individual simulated smoking and other cause mortality histories generated by the Smoking History Generator to generate a simulated US population. A Natural History Component utilizes the TSCE model, previously calibrated to smoking cohort data in Project One, to estimate lung cancer deaths in the simulated US population based on the TSCE model. A Survival Mortality Component includes effects of the lag time from first malignant cell to lung cancer death in the TSCE model, and adjustments for additional age, period, and birth cohort to improve the fit to US lung cancer mortality.

Component Listing

The components used to construct the FHCRC lung cancer model include:

- Population Component
- Natural History Component
- Survival Mortality Component

SMOKING HISTORY GENERATOR COMPONENT

SUMMARY

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 lung cancer consortium.

OVERVIEW

The core SHG software was parameterized using three tobacco control scenarios to 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-

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**

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**	**CTC**
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 (1890-1984)	User Defined		
Gender (Male/Female)	User Defined		
Race (All race)	User Defined		

[1] Cigarettes smoked per day,[2]Other 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

Parameter	Valid Values
Seed value for PRNG used for Initiation, Cessation, OCD[1], Smoking intensity quintile	Integer from -1 to 2147483647 (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, Smoking intensity quintile and CPD[4] data files	As derived from NHIS depending on the scenario

[1]Other 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. [3]Key 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. [4]Cigarettes 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 History	Smoking intensity quintile (5 quintiles ranging from light to heavy smoking), Yearly smoking dose (CPD[2])

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

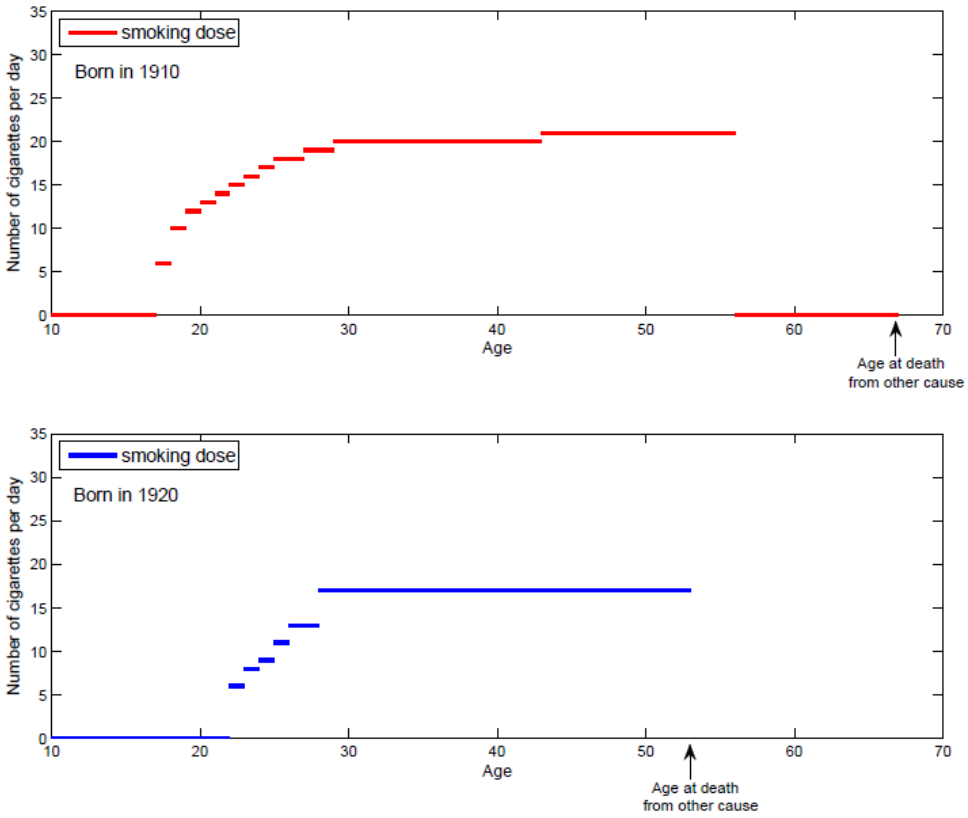


Figure SHG-1: Examples of the SHG-Generated Events

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

Population Component

Summary

We use individual level simulation of the full US population stratified by single year age and birth year, and gender.

Overview

The purpose of the population component was to model smoking and other cause mortality in the US. The Smoking History Generator provided simulated individual histories, including birth year, age and smoking intensity at start of smoking, and ages and intensities at subsequent ages when smoking habits changed or when smoking stopped, and projected age of other cause death.

Quantitative Description

We use individual-based microsimulation of the full US population stratified by single year age and birth year, and gender. These smoking histories were sampled and used to build up a synthetic population that matched the full US population binned by single year ages 30-84 and calendar years 1975-2000. The TSCE model (a Markov transition model at the cellular level) was applied to each of the time-dependent smoking histories to estimate lung cancer deaths. This was combined with a statistical model representing additional age, period, and birth cohort effects, based on the stratification of the US population by age and calendar year.

Population Dynamics

The TSCE model relates smoking to cell dynamics from birth to death for each individual in the population. However, changes in lung cancer over time in the US population were not fully captured by the TSCE model. Thus we made a second calibration by estimating additional period and birth cohort effects. These factors may adjust for model misspecification, effects of changing cigarette composition, changing demographics, and other exposures and environmental factors.

Recurrence

The FHCRC model represents lung cancer mortality, not detection or recurrence.

Natural History Component

Summary

The FHCRC lung model represents lung cancer development as a stochastic (two-stage) cellular process that is influenced by an individual's smoking history.

Overview

The natural history component relates the probability of lung cancer mortality to each individual's smoking history using the biologically-motivated TSCE model. In the TSCE model, smoking influences the carcinogenic process throughout life, with the model representing survival until occurrence of a first malignant cell. A lag time is used to represent time from first malignant cell to lung cancer death (Survival Mortality Component). The natural history model builds on the (Population Component) that simulates the full US population based on simulated individual smoking histories and date of other cause mortality. The natural history provides estimates of US lung cancer mortality by age, gender and calendar year. These estimates from the natural history component are further adjusted by calendar year and birth cohort (Survival Mortality Component).

Disease Stages

The TSCE model represents carcinogenesis as a lifeling process consisting of initiating mutations, clonal expansion of initiated cells, and malignant conversion. In the TSCE model, malignant conversion is defined as occurrence of the first malignant cell. The probability of lung cancer is related to the evolving joint probability distribution of these different cells throughout life. There is no explicit definition of disease stages in this model.

Disease growth

We model the probability distribution for discrete cells as described above, allowing almost continuous growth of the intermediate lesions that consist of initiated cells.

Stage Transition Trends

The cell birth, death, and mutation in the carcinogenic process naturally give rise to changing transition rates because any cell in the increasing mass of intermediate cells can mutate to give rise to cancer. Thus in general, the effective rates increase with age, even if the rates for a single cell are held constant.

Regression

The model allows for disease regression as pre-malignant clones may become extinct through random process of cell death.

Survival Mortality Component

Summary

This document describes how survival and mortality are modeled.

Overview

The TSCE model (See Natural History Component) represents cellular dynamics during carcinogenesis until occurrence of the first malignant cell. Thus it is necessary to include a lag time, or distribution of lag times, to represent the time from first malignant cell to lung cancer death. Further adjustments to lung cancer deaths in the US population are modeled as calendar year and birth cohort effects.

Survival Estimation Covariates

Survival and mortality depend on disease progression according to the TSCE natural history model, combined with a statistical model representing additional age, period, and birth cohort effects.

Other Cause Mortality

The Smoking History Generator was developed to reflect historical smoking trends and rates for other cause mortality in the US population. The FHCRC lung cancer model used the simulated individual smoking histories including the age of other cause death provided by the SHG.

OUTPUT OVERVIEW

SUMMARY

This document describes the types of outputs generated by the FHCRC lung cancer model.

OVERVIEW

The FHCRC lung cancer model provides estimates of lung cancer deaths in the US population by gender, age, and calendar year, given historical smoking patterns or alternative (counter-factual) smoking scenarios.

OUTPUT LISTING

1. **Project One :** Biological parameters and dose-response parameters related to smoking exposure in the TSCE model (See Table3)

- back ground initiation rate
- back ground division rate of an initiated cell
- back ground net cell proliferation rate of an initiated cell
- back ground malignant conversion rate
- two parameters regarding dose-response effect for the division rate and the net cell proliferation rate
- two parameters regarding dose-response effect for the malignant conversion rate
- lag time parameter for the progression (assumed 5 years)

These estimates for pre-malignant cellular kinetics relate directly to the commonly reported lead time from smoking exposure to lung cancer, and to the growth rate of pre-malignant lesions. Validation of these rates included comparison with calibration to other cohorts. These outputs were the basis for estimating the relation between smoking and lung cancer in the simulation of the US population (Project Two).

2. **Project Two :** Parameters for secular time trends in both TSCE-PC model and TSCE-APC model

- calendar year effects (calendar years from 1975 to 2000) (Figure2 and Figure3)
- birth cohort effects (birth cohorts 1890-1894,1895-1899,....,1960-1964,1965- for AC) (Figure4 and Figure5)
- additional age effects in the TSCE-APC model (Figure6)

3. **Lung Smoking Base Case:** Model outputs in both TSCE-PC model and TSCE-APC model

- Age-standardized lung cancer mortality rates (using the 2000 US standard population, Census P25-1130) by calendar year for US males and US females for three different smoking scenarios (ATC, NTC, CTC) (For ATC case, see Figure7 and Figure8. Figures for NTC and CTC cases will be available in the forthcoming paper)
- Avoided lung cancer deaths resulted from the actual tobacco control (# lung cancer deaths in NTC – # lung cancer deaths in ATC) (Figures will appear in the forthcoming paper)
- Avoidable lung cancer deaths by assuming all smoking stops in 1965 (# lung cancer deaths in ATC – # lung cancer deaths in CTC) (Figures will appear in the forthcoming paper).

These estimates for lung cancer deaths, stratified by gender, age and calendar year, may be converted to stratum specific lung cancer rates by dividing by population counts within each stratum. Estimates of avoided and avoidable deaths were compared against historical outcomes for lung cancer mortality in the US to provide some indication of the benefit of the US Surgeon General's warnings about the dangers of smoking, and also an upper limit to the potential benefits of additional efforts at reducing tobacco consumption.

RESULTS OVERVIEW

SUMMARY

This document summarizes results from calibration of the FHCRC lung cancer model to cohort data (Project One) and application to understand the effects of smoking on lung cancer mortality in the US population (Project Two).

OVERVIEW

The FHCRC lung cancer model used the TSCE model incorporating the mechanisms of initiation, promotion and malignant conversion in carcinogenesis to analyze lung cancer mortality in five large cohorts; the British Doctors's, CPS I and II, HPFS and NHS cohorts (Project One). The parameter estimates from calibration to lung cancer mortality using these cohorts are closely tied to the model purpose of understanding the underlying biological mechanisms that relate tobacco smoke to lung cancer in the US population (Project Two).

By fitting the model to these cohorts data, we estimated the biological parameters related to age-specific cancer rates and dose-response parameters related to smoking exposure. The key biological parameters are the rate of initiation, the rates of cell division and apoptosis/differentiation of initiated cells, and the rate of malignant conversion of initiated cells. Progression from the first appearance of malignancy to death from lung cancer is modeled as a constant lag time. And the effect of smoking habits on age-specific lung cancer mortality is modeled via a dose-response function on each of these parameters. Based on the likelihood approach, we chose the most parsimonious model which is consistent with these cohorts data, and summarized the list of parameters in the final model here.

We began the model calibration to each cohort assuming that smoking could influence promotion, initiation, or malignant conversion. This is called the full model. After optimization of the full model, we progressively eliminated parameters that did not significantly contribute to the likelihood. This is called the reduced model (refer to Table1).

We found that the contemporaneous CPS-I and British doctors cohorts could be fit with all but one parameter in common. The CPS-II cohort was followed about 20 years later, and represented individuals smoking newer cigarettes. The CPS-II cohort dose-

response differs from the earlier cohort by having no significant effect of cigarette smoking dose-response on initiation, but a larger dose-response on promotion than found in the CPS-I cohort (for details, refer to the paper by Hazelton et al. 2005, See Table2). The HPFS and NHS cohorts are in the similar period as the CPS-II cohort, and we found no significant effect of smoking dose-response on initiation, which is consistent with the CPS-II cohort. However, we found significant effect of smoking dose-response on promotion as well as malignant conversion in the HPFS and NHS lung cancer mortality cohorts (refer to Table3).

The FHCRC lung cancer model combines the TSCE model, which was calibrated to the lung cancer mortality in the HPFS cohort for males and the NHS cohort for females, with US population-based adjustments for secular time trends (period and birth cohort in the TSCE-PC model, and additional age, period and birth cohort in the TSCE-APC model). The NCI's SHG can simulate individual-level smoking histories and generate cohorts of individuals for different smoking scenarios. We use the FHCRC lung cancer model to predict lung cancer mortality for various smoking scenarios using the cohorts generated by the SHG.

Three smoking scenarios are explored in the Lung Smoking Base Case study.

1. Actual tobacco control (ATC): Due to the increasing information about the harmful effects of smoking on the public health, including the US Surgeon General's report about the risk of smoking around 1964, smoking habits have been changed in the US over last several decades starting in the early 1950s. The SHG generated individuals with smoking histories mimicking actual smoking trends in the US population.
2. No tobacco control (NTC): To investigate the effects of tobacco control on lung cancer mortality, individuals with comparative smoking histories by assuming that no tobacco control occurred were generated by the SHG.
3. Complete tobacco control (CTC): To explore the maximum potential benefits from tobacco control on lung cancer mortality, the SHG generated individual histories that all smokers quit smoking in 1965 and no individuals began smoking after that.

There are two separate cohorts generated by the SHG and we present the results for both cohorts.

1. Empirical Cohorts (EC): Starting with birth cohort 1900, based on empirical data
2. All Cohorts (AC): Starting with birth cohort 1890, involves extrapolated smoking history data

Results List

Project 1 -TSCE smoking natural history model output

During the calibration of the TSCE natural history model to different lung cancer mortality cohorts [See Project One], we found that smoking exposure tends to increase

rates of mutation, and more importantly, smoking significantly increases the growth (promotion) of initiated cells, leading to increased net cell proliferation rates of pre-malignant cells, and a subsequent rise in risk for lung cancer.

The FHCRC group chose to use the lung cancer mortality calibration of the TSCE model to the HPFS cohort for males and the NHS cohort for females that were performed as the first project [Project One]. This TSCE model calibration was used as the primary component in constructing the FHCRC lung cancer model. The HPFS (1986-2002) and NHS (1976-2000) cohorts calibration was chosen on the basis that these cohorts may take account for the risks from current cigarette compositions, and may well represent the contemporary US lung cancer trend and demographics for the Lung Smoking Base Case, which are ranged over the period 1975-2000. Furthermore these cohorts have information for former smokers as well as never and current smokers, with extensive cross tabulation of lung cancer deaths and population at risk by gender, race, age, duration of smoking, and smoking rate. Because of the explicit information for former smokers, these cohorts may be adequate to reflect the impact of quitting smoking on lung cancer mortality. To compensate for possible limitations of the cohort-calibrated TSCE model, the FHCRC group introduced additional age, period, and birth cohort effects when calibrating to lung cancer mortality in the US population.

Project 2 - Outputs of additional age, period, and birth cohort adjustments in Lung Smoking Base Case

For the Lung Smoking Base Case (see Project Two) the expected number of lung cancer deaths were calculated from the cohort-calibrated TSCE model, using the individual histories generated by the NCI's SHG that were selected to fill up the full US population table by single year of age from 30 to 84, and calendar years from 1975 to 2000. The outputs from the additional age, period, and birth cohort adjustment consist of a set of parameters (see Parameter Overview) to optimize the FHCRC lung cancer model projections to observed US population counts and numbers of lung cancer deaths by gender and single year of age from 30 to 84, and calendar years from 1975 to 2000. The Lung Smoking Base Case [Project Two] study utilizes calibration to a lung cancer mortality cohort [Project One] that includes background, dose-response, and lag time parameters relating to the effects of smoking to lung cancer mortality. These parameters are used in calibrating to US population and lung cancer mortality data. Outputs from this process include projections of lung cancer mortality by single years for ages 30-84, and calendar years 1975-2000. In the Lung Smoking Base Case, we predict the number of avoided deaths from lung cancer under the actual tobacco control scenario compared to no tobacco control scenario, and also the potential avoidable deaths from lung cancer if all smokers quit smoking in the year 1965, and no one started smoking after that. The explicit results from the Lung Smoking Base Case will appear in the forthcoming paper.

PROJECT ONE

The first project consisted of calibrating a natural history model to substantial US smoking cohorts to model the probability of lung cancer death at different ages in terms of an individual's smoking behavior up to that age. The two stage clonal expansion (TSCE) model was chosen to represent the natural history of lung cancer (Moolgavkar and Venzon, 1979; Moolgavkar and Knudson, 1981). The TSCE model is a stochastic model that represents the processes of initiation, promotion, and malignant conversion during carcinogenesis. The TSCE model parameters were calibrated to lung cancer mortality among never and current smokers in the American Cancer Society CPS-I and CPS-II cohorts and also the British Doctors cohort from the UK (Hazelton et al., 2005). We also calibrated the model to lung cancer incidence (Meza et al., 2008) and mortality in the prospective Health Professionals Follow-up Study (HPFS) and the Nurses' Health Study (NHS) cohorts. These calibrated natural history model parameters were shared with other CISNET lung cancer modeling groups for use in modeling US lung cancer mortality. The TSCE model provided excellent fits to the data representing smoking histories for individuals in the different cohorts. The FHCRC lung group's use of these natural history parameters is discussed in more detail below.

The natural history model calibrations to different cohorts allowed us to compare estimates for the effects of tobacco on lung cancer through analysis of incidence (HPFS and NHS) or mortality (CPS-I, British Doctors, CPS-II, HPFS and NHS). Follow-up for the CPS-I and British Doctors cohorts occurred about 20 years prior to CPS-II, HPFS and NHS, allowing us compare the effects of earlier versus later cigarette compositions. In general, all models indicated that the most important dose-response effect of tobacco smoke is on promoting the growth (increasing the clonal expansion rate) of pre-malignant cells. This promotion effect is slightly stronger in the CPS-II, HPFS, and NHS analysis with the more recent (lower tar and nicotine) cigarette compositions. Lung cancer risk is also slightly increased by an influence of smoking on initiation, but this effect is insignificant in the newer cigarettes.

Other factors outside the natural history model's domain influence the lung cancer mortality rates in the US population. First, individuals in the US smoking cohorts are not fully representative of the US population. Second, the composition of cigarettes has changed substantially over time. Third, other exposures and environmental factors may contribute to lung cancer risk. In addition, the TSCE model consists of only two stages, whereas the lung cancer process is biologically complex, and progresses along many pathways. Thus, the TSCE natural history model may not be capable of fully capturing the effects of smoking on the subsequent risk of lung cancer. Therefore, in addition to the biologically based smoking model, we introduce additional age, period, and birth cohort effects to adjust for these other factors when calibrating the FHCRC lung cancer model to US lung cancer mortality data.

See Also: Project One, Project Two

PROJECT TWO

Each CISNET lung cancer modeling group has developed age-specific models of US lung cancer mortality trends for ages 30-84 and calendar years 1975-2000. This comparative modeling effort of different groups is called the Lung Smoking Base Case. The FHCRC lung cancer modeling group chose to use the TSCE natural history model, calibrated to relate smoking histories to lung cancer mortality, as the central element in addressing the Lung Smoking Base Case. However, the FHCRC group found that additional age, period, and birth cohort effects were needed to account for changes in cigarette composition over time, other factors that contribute to lung cancer, and limitations of the TSCE natural history model. This combination of biologically based modeling, combined with statistical adjustments for additional age, period, and birth cohort, constitutes the FHCRC lung cancer model. The FHCRC lung cancer model was subsequently used to evaluate the impact of public health messages on lung cancer trends and life years lost (or gained) under alternative tobacco consumption scenarios.

Smoking is the most significant risk factor in modeling US lung cancer trends. The Smoking History Generator (SHG) provided by the National Cancer Institute (NCI) was used to generate individuals with simulated smoking histories. We developed software to sample and combine these simulated individuals to imitate the full US population table for each calendar year (ranging from 1975 to 2000) and age bin (ranging from 30 to 84) separately for males and females. Then the cohort-calibrated TSCE natural history model was applied to each of these simulated individuals with smoking histories to calculate the expected lung cancer deaths in each cell of the US population table for males and females. However, comparison of the number of expected lung cancer deaths against the number of observed lung cancer deaths in each cell revealed discrepancies and apparent age, period, and birth cohort trends are not fully accounted for by the model.

Factors other than smoking influence US lung cancer mortality trends. Although the TSCE natural history model accounted for most lung cancer deaths observed in the US, additional age, period, and birth cohort factors were required to accurately represent the detailed lung cancer mortality outcomes in the US population. Thus, we did further calibration by applying additional age, period, and birth cohort effects to the modeled US lung cancer mortality rates by the TSCE natural history model to match the observed US lung cancer mortality rates.

IMPACT OF PUBLIC HEALTH MESSAGES ON US LUNG CANCER MORTALITY

It is generally thought that increasing US lung cancer trends slowed, and even decreased, in response to increasing public awareness during the 1960's and later about the dangers of smoking (Irvine et al., 2006; Musk et al., 2003). This public knowledge came from many sources, including statements of the US Surgeon General about the risk of smoking (Parascandola et al., 2001, 2006).

See Also: Project One, Project Two

TSCEMODEL DETAILS

The TSCE model is a mathematical model that represents the carcinogenic process by tracking the probability distribution for the number of cells on the pathway to cancer. The model assumes that any of the normal stem cells in lung may undergo, at random, a first mutation step (called initiation) at rate $\nu(t)$ during the course of cell division to create an initiated cell. Each initiated cell may undergo cell division at rate $\alpha(t)$ or cell death at rate $\beta(t)$. A random second mutation event may occur at rate $\mu(t)$ as any of the initiated cells undergo cell division, producing a malignant cell. After occurrence of the first malignant cell, a lag time is used to represent the time between the appearance of the first malignant cell and lung cancer mortality (See Figure1).

Let $d(t)$ be the exposure dose to smoking at age t. Then we assume that initiation, promotion, and malignant conversion rates may be altered during periods of exposures through flexible dose-response relationships:

$$\theta(d(t)) = \theta_0(1 + \theta_c d(t)^{\theta_e}),$$

where θ represents a biological parameter in the model, θ_0 is the background parameter, and θ_c and θ_e are the dose-response coefficients corresponding to smoking. Closed form expressions for the hazard and survival function of the TSCE model are known in the case of piecewise constant parameters (Heidenreich et al., 1997).

$$S_2(t) = \exp\left\{\sum_{j=1}^{n} \frac{\nu_j X}{\alpha_j} \ln\left(\frac{q_j - p_j}{f_j(t_{j-1}, t_n)}\right)\right\},$$

$$h_2(t) = \sum_{j=1}^{n} \frac{\nu_j X}{\alpha_j} \frac{\partial}{\partial t_n} \ln\left(f_j(t_{j-1}, t_n)\right) = \sum_{j=1}^{n} \frac{\nu_j X}{\alpha_j} \frac{1}{f_j(t_{j-1}, t_n)} \frac{\partial}{\partial t_n} f_j(t_{j-1}, t_n),$$

where X is the number of normal stem cells, n is the number of age-periods with different parameter values before age $t_n \equiv t$; $[t_{j-1}, t_j], j = 1, ..., n$ denote the end-points of the $j-th$ age-period, d_j is the smoking-dose during the $j-th$ age-period, $t_0 = 0$, and ν_j, α_j, g_j, μ_j denote the parameter values during the $j-th$ age-period, and

$$g_j = g(1 + g_c d_j^{g_e}), \qquad \alpha_j = \alpha(1 + g_c d_j^{g_e}),$$

$$\nu_j = \nu(1 + \nu_c d_j^{\nu_e}), \qquad \mu_j = \mu(1 + \mu_c d_j^{\mu_e}),$$

$$p_j, q_j = \frac{1}{2}\left(-g_j \mp \sqrt{g_j^2 + 4\alpha_j \mu_j}\right).$$

$$\tilde{y}_n = 0, \qquad \tilde{y}_{j-1} = \frac{\alpha_{j-1}}{\alpha_j} \frac{(\tilde{y}_j - p_j) q_j e^{q_j(t_{j-1} - t_j)} + (q_j - \tilde{y}_j) p_j e^{p_j(t_{j-1} - t_j)}}{f_j(t_{j-1}, t_n)},$$

$$f_j(t_{j-1},t_n)=(\tilde{y}_j-p_j)\exp\{q_j(t_{j-1}-t_j)\}+(q_j-\tilde{y}_j)\exp\{p_j(t_{j-1}-t_j)\},$$

$$\frac{\partial}{\partial t_n}f_n(t_{n-1},t_n)=[\exp\{q_n(t_{n-1}-t_n)\}-\exp\{p_n(t_{n-1}-t_n)\}]p_nq_n,$$

$$\frac{\partial}{\partial t_n}f_j(t_{j-1},t_n)=[\exp\{q_j(t_{j-1}-t_j)\}-\exp\{p_j(t_{j-1}-t_j)\}]\frac{\partial}{\partial t_n}\tilde{y}_j,$$

$$\frac{\partial}{\partial t_n}\tilde{y}_{n-1}=\frac{\alpha_{n-1}}{\alpha_n}\frac{(q_n-p_n)^2e^{(p_n+q_n)(t_{n-1}-t_n)}}{(f_n(t_{n-1},t_n))^2}p_nq_n,$$

$$\frac{\partial}{\partial t_n}\tilde{y}_{j-1}=\frac{\alpha_{j-1}}{\alpha_j}\frac{(q_j-p_j)^2e^{(p_j+q_j)(t_{j-1}-t_j)}}{(f_j(t_{j-1},t_n))^2}\frac{\partial}{\partial t_n}\tilde{y}_j.$$

These equations may be used to calculate the TSCE model survival $S_2(u)$ and hazard $h_2(u)$ at any time u.

If we assume a constant or gamma lag time between the appearance of the first malignant cell and lung cancer death, the survival probability at age t of an individual with smoking history d_j, $S(t;\bar{\theta}(d_j))$, is given by

$$S(t;\bar{\theta}(d_j))=\begin{cases}S_2(t-t_{lag};\bar{\theta}(d_j)) & \text{if lag time is constant,}\\ 1-\int_0^t(1-S_2(u;\bar{\theta}(d_j)))g(t-u)du & \text{if lag time is gamma distributed,}\end{cases}$$

where $\bar{\theta}(d_j)$ denotes the vector of identifiable model parameters given the smoking history d_j, and $g(\cdot)$ is the gamma density. We calculate the probability for lung cancer mortality in each single year of age for the individual, given the gender and the full smoking history of the individual. During calibration to smoking cohort data, the study follow-up times and known outcome for each individual (death from lung cancer, or study censoring) were combined with the model probabilities for individual death from lung cancer at each age to form an individual likelihood.

The individual likelihood $\mathcal{L}_j=\mathcal{L}_j(t_j,s_j;\bar{\theta}(d_j))$ depends on time of entry into the study s_j , censoring or failure time t_j, and on detailed smoking exposure histories in conjunction with general dose-response models for the biological parameters in the TSCE model, and on the lag time or lag time distribution. We assume that each individual is lung cancer free at the beginning of the study, s_j. The individual likelihoods for cases and survivors, including left truncation, are given by

$$\mathcal{L}_j(t_j,s_j;\bar{\theta}(d_j))=\begin{cases}-S'(t_j;\bar{\theta}(d_j))/S(s_j;\bar{\theta}(d_j)) & \text{if death from lung cancer,}\\ S(t_j;\bar{\theta}(d_j))/S(s_j;\bar{\theta}(d_j)) & \text{otherwise,}\end{cases}$$

where the prime denotes derivative with respect to t.

Assuming independence between individuals, the cohort likelihood is the product of individual likelihoods over all subjects j,

$$\mathcal{L}=\prod\mathcal{L}_j(t_j,s_j;\bar{\theta}(d_j)).$$

Gradient search methods (Bhat FORTRAN software library, Luebeck 2009) were then used to maximize the likelihood, leading to a set of model parameters that relate individual histories to the probability of death from lung cancer at each age.

TSCE SMOKING NATURAL HISTORY MODEL INPUTS

Individual histories include gender, age at start smoking (if a smoker), beginning smoking rate (number of cigarettes per day), age at each change in smoking habit, and smoking rate during each of these periods, the age at quit smoking (if that occurs), age at entry into the study, age at lung cancer death or end of study follow-up. These individual history inputs were input from the cohort records during calibration to smoking cohort data [see Project One].

The TSCE-PC and TSCE-APC model calibrations utilized histories generated by the NCI provided Smoking History Generator (SHG) to fill up the full US population tables by gender and single year of age from 30 to 84, and calendar years from 1975 to 2000. The TSCE natural history model used the simulated smoking history inputs for individuals that contributed to each cell of the simulated US population table to calculate the expected number of lung cancer deaths by single year of age and calendar year.

INPUTS FOR MODEL CALIBRATION IN LUNG SMOKING BASE CASE

Assume lung cancer mortality data is in tabular form for $i = 1, \cdots, I$ age groups and $j = 1, \cdots, J$ calendar years. For age group i, the number of lung cancer deaths during calendar year j can be assumed to follow a Poisson distribution with mean $\Lambda_{i,j}$.

We consider two calibration approaches:

1. Age-Period-Cohort model (TSCE-PC model):

$$\Lambda_{i,j} = PY_{i,j} h(a_i) b_{i,j} c_j,$$

where a_i is the mean of the $i-th$ age group, $b_{i,j}$ and c_j are coefficients that adjust for birth cohort and calendar year (period) effects, respectively, $PY_{i,j}$ is the person years at risk, and $h(a_i)$ represents the hazard function of the TSCE natural history model with lag time evaluated at age a_i.

2. Age-Age-Period-Cohort model (TSCE-APC model):

$$\Lambda_{i,j} = PY_{i,j} h(a_i) e_i b_{i,j} c_j,$$

where e_i is the coefficient that adjust for additional age effect for the $i-th$ age group, and other variables are the same as above.
The overall likelihood $\mathcal{L}$ for the observed lung cancer mortality in all age-calendar year groups is given by

$$\mathcal{L} = \prod_{i,j} \frac{\Lambda_{i,j}^{O_{i,j}} e^{-\Lambda_{i,j}}}{O_{i,j}!},$$

where $O_{i,j}$ is the number of lung cancer deaths in the $i-th$ age group during calendar year j.

The parameters in the TSCE natural history model were estimated by calibrating to lung cancer mortality cohorts [refer to the Project One]. And these estimated parameters are used in the calculation of $h(a_i)$. Thus the above likelihood is used to estimate the secular terms: period and birth cohort effects in the TSCE-PC model, and additional age, period and birth cohort effects in the TSCE-APC model. Note in these two models, `TSCE' refers the age effect calculated from the TSCE model.

In the Lung Smoking Base Case study, the inputs for model calibration consist of US population counts and numbers of lung cancer deaths by gender and single year of age from 30 to 84, and calendar years from 1975 to 2000.

TABLE3

Table 2: Reduced model estimates for the HPFS and NHS cohorts

Parameter	p_1	p_2	p_3	p_5	p_6	p_{10}	p_{11}
Cohort	Background cell division rate (per cell per year)	Background net cell proliferation rate (per cell per year)	Background initiation and malignant conversion rate (per cell per year)	Tobacco promotion rate coefficient	Tobacco promotion rate power	Tobacco malignant conversion rate coefficient	Tobacco malignant conversion rate power
HPFS	3.000	0.106	7.311e-8	0.134	0.474	0.175	0.552
NHS	3.000	0.083	9.370e-8	0.189	0.491	0.087	0.711

FIGURE2

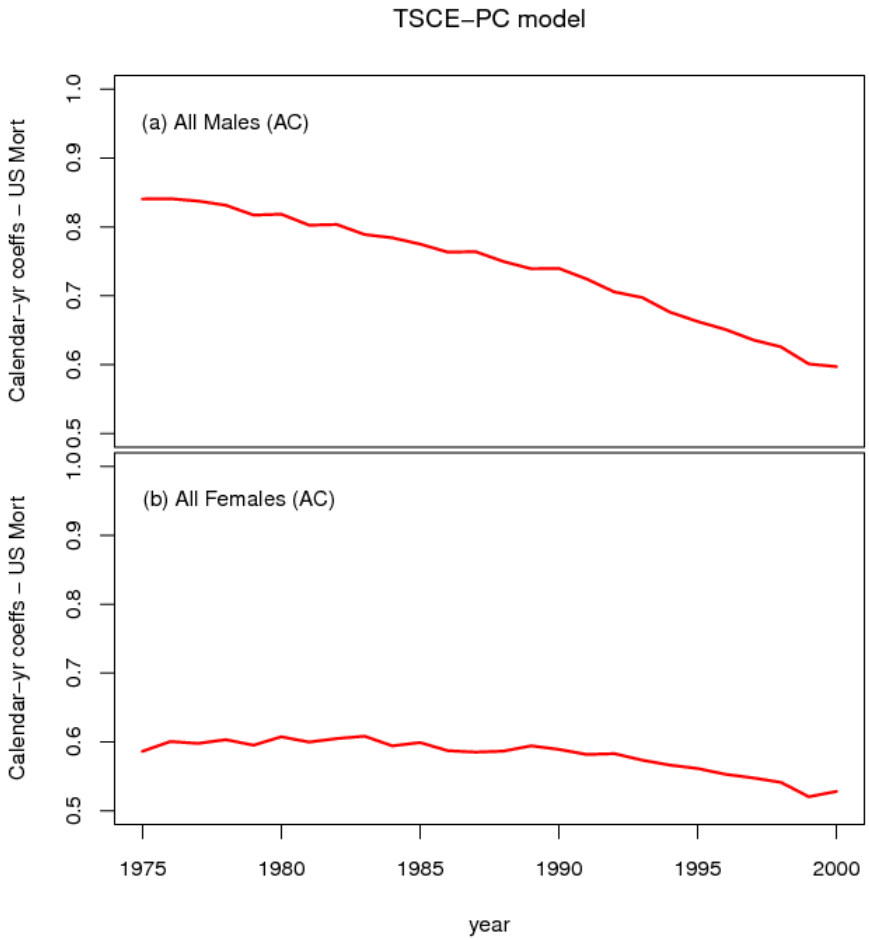


Figure 2: Calendar year effects (1975-2000). AC: all cohorts starting with birth cohort 1890, involves extrapolated smoking history data. Age-Period-Cohort model is used.

FIGURE3

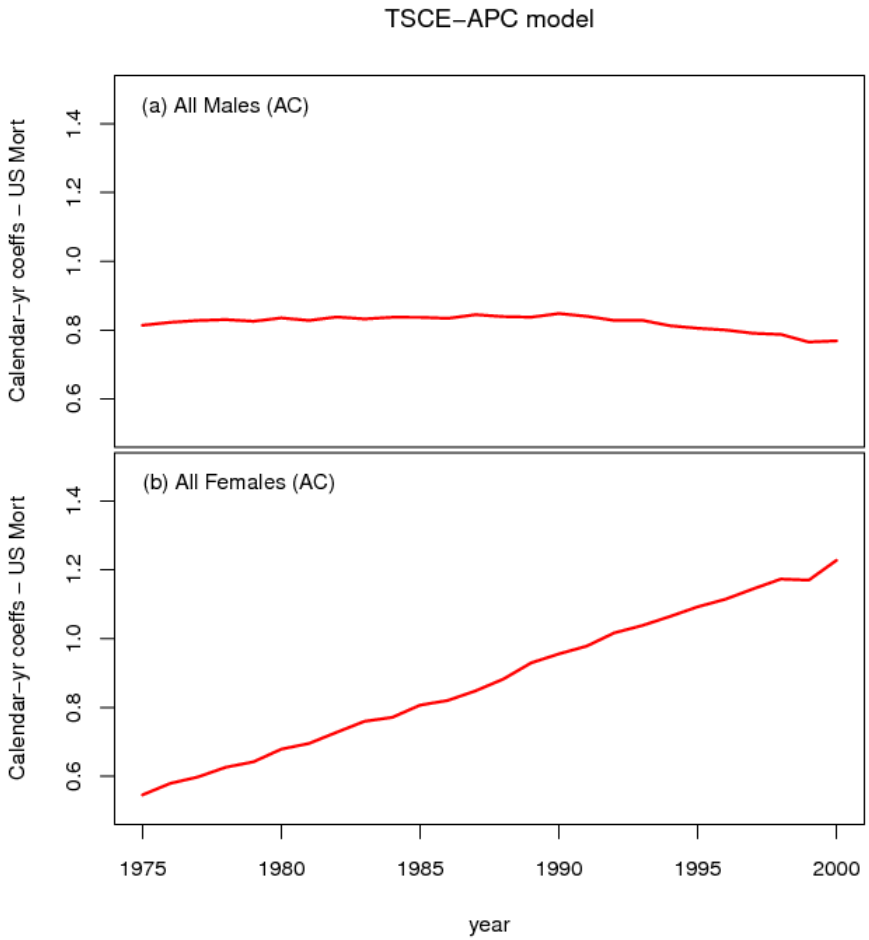


Figure 3: Calendar year effects (1975-2000). AC: all cohorts starting with birth cohort 1890, involves extrapolated smoking history data. Age-Age-Period-Cohort model is used.

FIGURE4

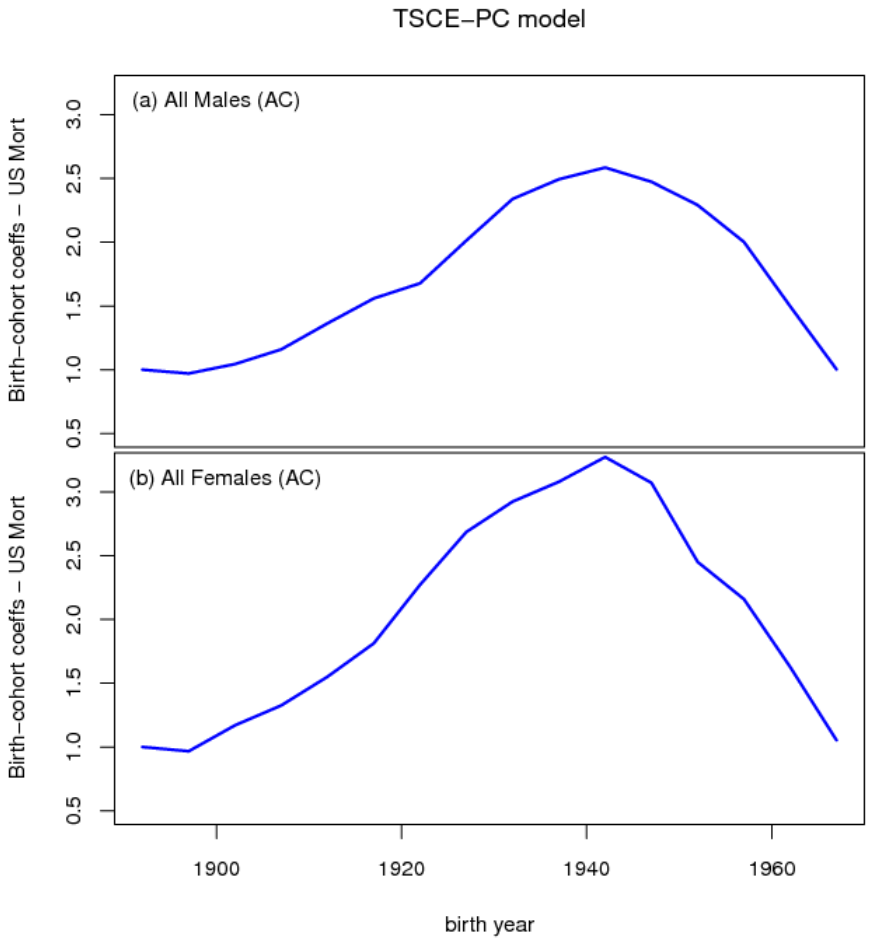


Figure 4: Birth cohort effects (birth cohorts 1890 – 1894, 1895 – 1899, ..., 1960 – 1964,>= 1965). AC: all cohorts starting with birth cohort 1890, involves extrapolated smoking history data. Age-Period-Cohort model is used.

FIGURE5

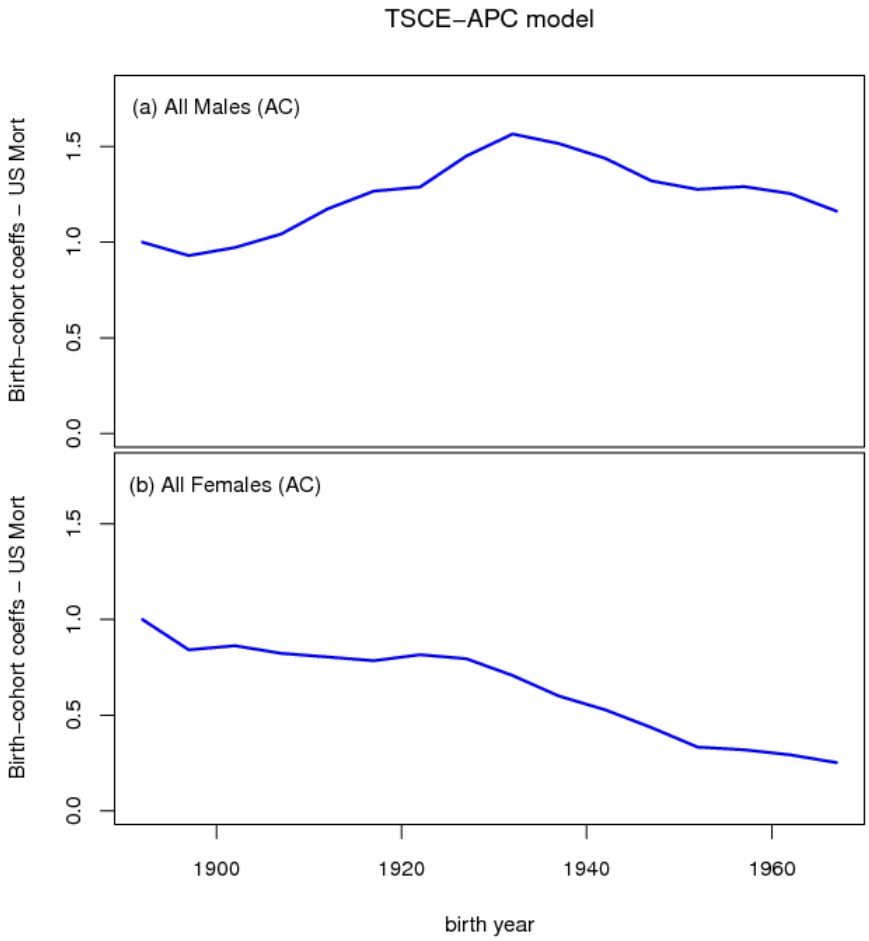


Figure 5: Birth cohort effects (birth cohorts 1890 – 1894, 1895 – 1899, ..., 1960 – 1964,>= 1965). AC: all cohorts starting with birth cohort 1890, involves extrapolated smoking history data. Age-Age-Period-Cohort model is used.

FIGURE6

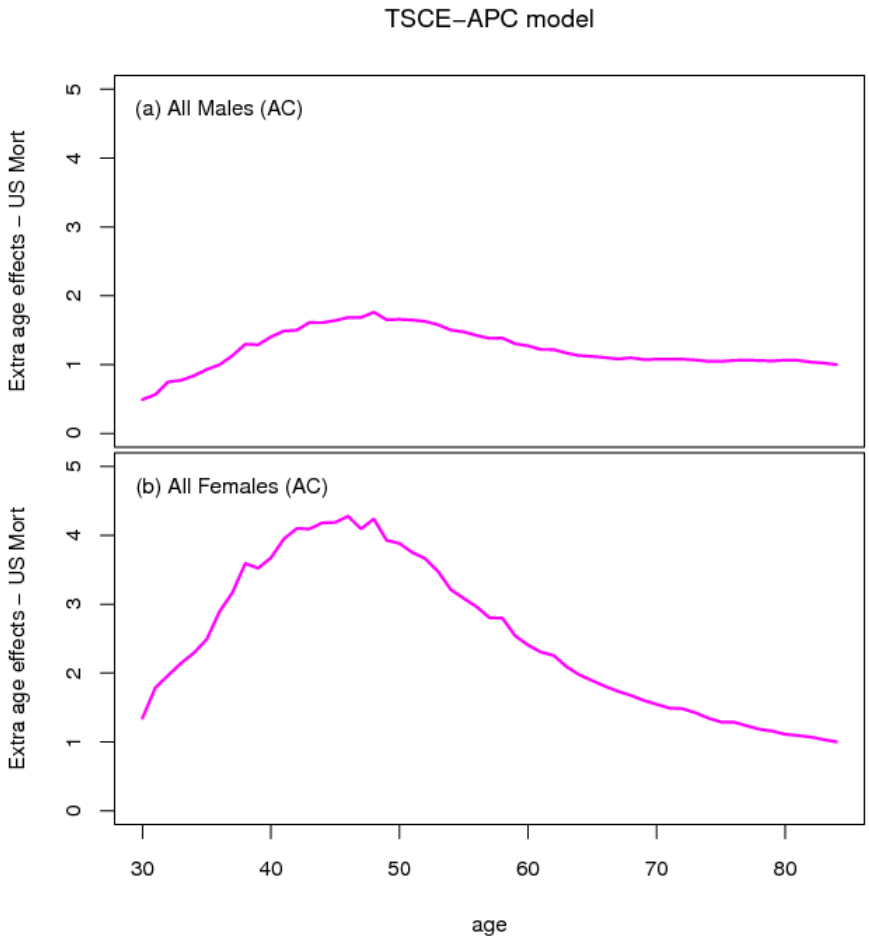


Figure 6: Additional age effects. AC: all cohorts starting with birth cohort 1890, involves extrapolated smoking history data. Age-Age-Period-Cohort model is used.

FIGURE7

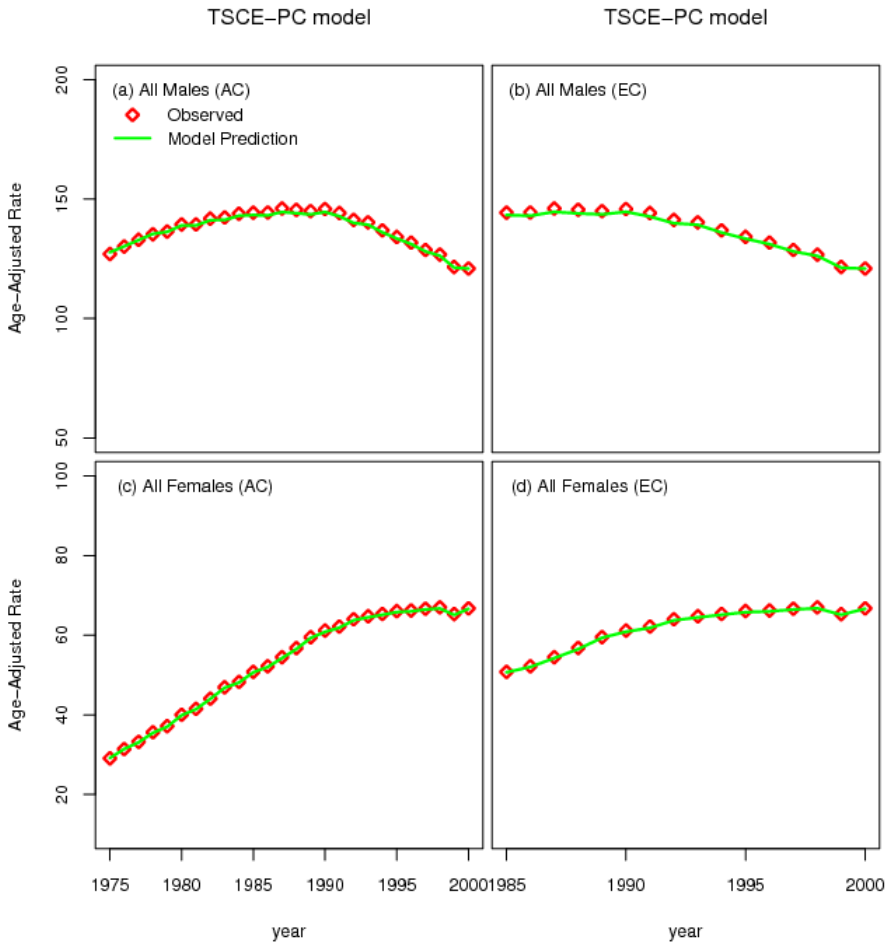


Figure 7: Age-standardized lung cancer mortality rates (using the 2000 US standard population, Census P25-1130) by calendar year. Diamond points represent the age-standardized lung cancer mortality rates in the observed US lung cancer mortality data, and the green line is the model prediction. AC: all cohorts starting with birth cohort 1890, involves extrapolated smoking history data. EC: empirical cohorts starting with birth cohort 1900, based on empirical data. Age-Period-Cohort model is used.

FIGURE8

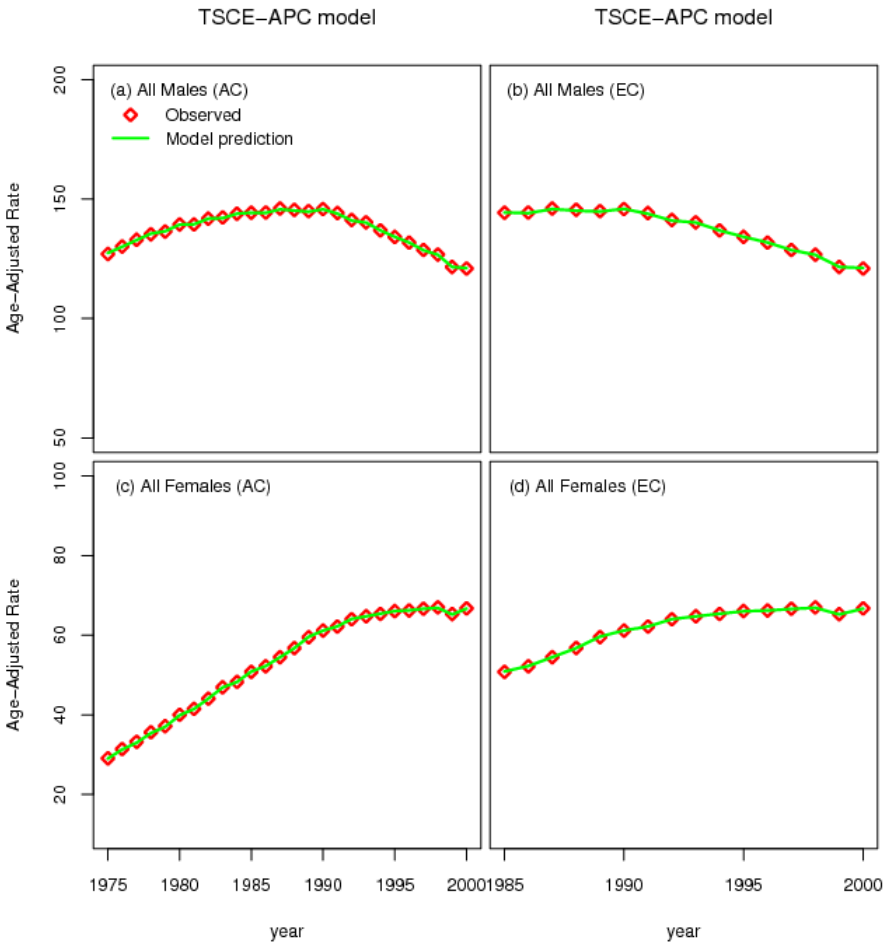


Figure 8: Age-standardized lung cancer mortality rates (using the 2000 US standard population, Census P25-1130) by calendar year. Diamond points represent the age-standardized lung cancer mortality rates in the observed US lung cancer mortality data, and the green line is the model prediction. AC: all cohorts starting with birth cohort 1890, involves extrapolated smoking history data. EC: empirical cohorts starting with birth cohort 1900, based on empirical data. Age-Age-Period-Cohort model is used.

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

TABLE1

Table 1: Full model and Reduced model dose-response

Full model.

Initial scale:

$X = 10^7$		Aassume 10^7 normal stem cells in both lungs*
$\nu_0 = \mu_0$		Equate background initiation, malignant conversion rates*

Background variables for cohort

Rate	Parameter	
α_0	p_1 (CPS-I) p_1' (Brit. docs.)	Background or nonsmoking cell division rate* (per cell per year)
$g_0 = \alpha_0 - \beta_0 - \mu_0$	p_2	Background or nonsmoking net cell promotion rate (per cell per year)
$\nu_0 = \mu_0$	p_3	Background initiation, malignant conversion* (per cell per year)

Full model dose-response for cohort

$\nu_i = \nu_0(1 + p_4\, dose_i^{p_8})$	Initiation rate* (per cell per year)
$g_i = g_0(1 + p_5\, dose_i^{p_6})$	Net initiated cell promotion rate (per cell per year)
$\alpha_i = \alpha_0(1 + (p_5 + p_9)\, dose_i^{p_6})$	Initiated cell division rate* (per cell per year)
$\mu_i = \mu_0(1 + p_{10}\, dose_i^{p_{11}})$	Malignant conversion rate* (per cell per year)

Fixed lag: $t_{lag} = p_7$

Gamma distribution with random variable X as lag time:
$f(x,a,b) = \frac{1}{b^a\Gamma(a)} x^{a-1} e^{-x/b}$, $\mu_{lag} = p_{12} = a\,b$, $\sigma_{lag} = p_{13} = \sqrt{a\,b^2}$

Reduced model.

X and background rates ν_0, α_0, g_0, and μ_0 are parameterized the same as in the full model above.

Reduced model dose-response for the British Doctors', CPS I and II cohorts

$\nu_i = \nu_0(1 + p_4)$, $p_4 = 0$ for nonsmokers	Initiation rate* (per cell per year)
$g_i = g_0(1 + p_5\, dose_i^{p_6})$	Net initiated cell promotion rate (per cell per year)
$\alpha_i = \alpha_0(1 + p_5\, dose_i^{p_6})$	Initiated cell division rate* (per cell per year)
$\mu_i = \mu_0$	Malignant conversion rate*, no dose-response (per cell per year)

Reduced model dose-response for the HPFS and NHS cohorts

$\nu_i = \nu_0$	Initiation rate*, no dose-response (per cell per year)
$g_i = g_0(1 + p_5\, dose_i^{p_6})$	Net initiated cell promotion rate (per cell per year)
$\alpha_i = \alpha_0(1 + p_5\, dose_i^{p_6})$	Initiated cell division rate* (per cell per year)
$\mu_i = \mu_0(1 + p_{10}\, dose_i^{p_{11}})$	Malignant conversion rate* (per cell per year)

Fixed lag:
$t_{lag} = 5$ years

Note: A full model was first tested where initiation, promotion, and malignant conversion are assumed to be potentially modified by smoking through flexible dose-response parameterization of the form: $a(1 + bx^c)$. Likelihood ratio tests were used to eliminate variables that were not significant for any of the cohorts, resulting in the reduced model.
*Only the product νX is identifiable. Thus, without loss of generality, we assume $X = 10^7$ normal stem cells at risk. This number may be changed by reciprocal change in ν. A separate identifiability condition relates variables α, μ, and ν. Thus, without loss of generality, we set $\mu_0 = \nu_0$. These variables may be rescaled without changing the model fit.

TABLE2

Table 2: Reduced model estimates for CPS-I, CPS-II and British Doctors' cohorts [MLE (MCMC 95% CI)]

Parameter	p_1	p_2	p_3	p_4	p_5	p_6
Cohort	Background cell division rate (per cell per year)	Background net cell proliferation rate (per cell per year)	Background initiation and malignant conversion rate (per cell per year)	Tobacco initiation rate	Tobacco promotion rate coefficient	Tobacco promotion rate power
CPS-I males	22.65(17.35,31.31)	0.075(0.065,0.085)*	1.40e-7(1.12e-7,1.76e-7)*	1.79(1.11,3.07)*	0.21(0.15,0.28)*	0.47(0.42,0.54)*
British Doctors	5.87(2.66,12.89)	0.075(0.065,0.085)*	1.40e-7(1.12e-7,1.76e-7)*	1.79(1.11,3.07)*	0.21(0.15,0.28)*	0.47(0.42,0.54)*
CPS-II males	7.70(6.45,12.99)	0.090(0.071,0.106)	7.16e-8(4.60e-8,1.21e-7)	~ 0.00(~ 0.00,1.76)	0.60(0.43,0.91)	0.22(0.12,0.30)
CPS-I females	71.56(49.08,100.0)	0.086(0.073,0.101)	8.93e-8(6.50e-8,1.19e-7)	1.23(0.32,2.80)	0.04(0.02,0.07)	0.98(0.80,1.15)
CPS-II females	15.82(13.39,42.12)	0.071(0.055,0.088)	1.07e-7(6.97e-8,1.62e-7)	0.02(~ 0.00,12.50)	0.50(0.27,0.86)	0.32(0.14,0.40)

Note: Estimates for net cell proliferation rates are tightly constrained, whereas cell division rates are not due to compensatory cell death.
* Shared variables in joint fit to White males in CPS-I and British Doctors' cohorts.

FIGURE1

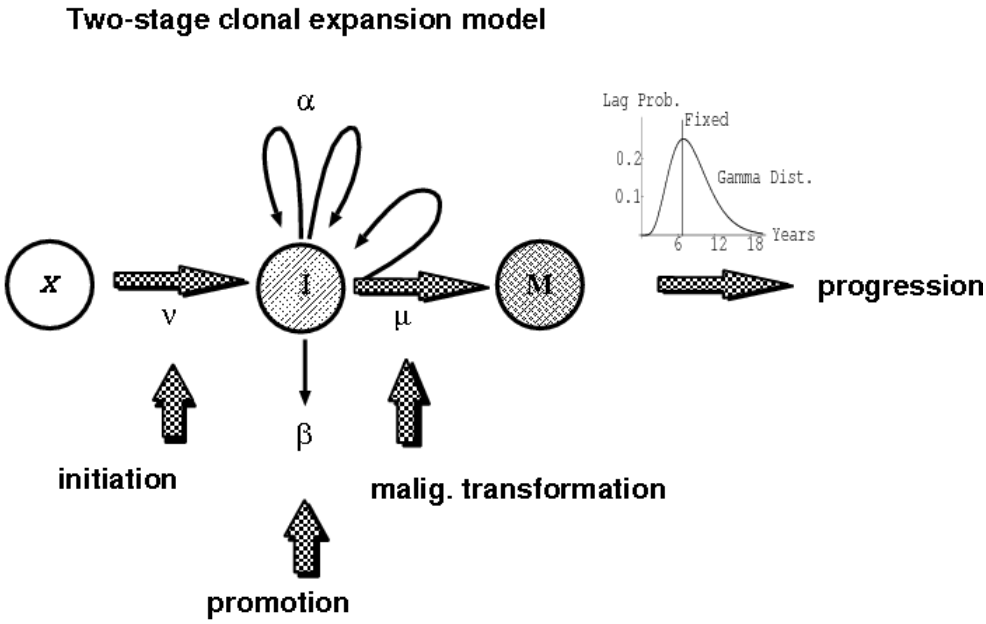


Figure 1: Two-stage clonal expansion (TSCE) model of lung cancer, including initiation, promotion, malignant conversion, and a lag time from first malignant cell to time of death from lung cancer.

KEY REFERENCES

- Alberg AJ, Ford JG, , Samet JM; American College of Chest Physicians. Epidemiology of lung cancer: ACCP evidence-based clinical practice guidelines (2nd edition). Chest 2007 Sep;132(3 Suppl):29S-55S.
- Jemal A, Siegel R, Ward E, Hao Y, Xu J, Murray T, Thun MJ. Cancer Statistics, 2008. CA Cancer J Clin. 2008 Mar-Apr;58(2):71-96.
- Stewart SL, Cardinez CJ, Richardson LC, Norman L, Kaufmann R, Pechacek TF, Thompson TD, Weir HK, Sabatino SA: Centers for Disease Control and Prevention (CDC). Surveillance for cancers associated with tobacco use: United States, 1999-2004. MMWR Surveill Summ. 2008 Sep 5;57(8):1-33.
- Coultas DB, Samet JM. Occupational lung cancer. Clin Chest Med. 1992 Jun;13(2):341-54.
- Moolgavkar SH, Knudson AG Jr. Mutation and cancer: a model for human carcinogenesis. J Natl Cancer Inst. 1981 Jun;66(6):1037-52.
- Stevens RG, Moolgavkar SH. Estimation of relative risk from vital data: smoking and cancers of the lung and bladder. J Natl Cancer Inst. 1979 Dec;63(6):1351-7.
- Moolgavkar SH, Venzon DJ. Two-event models for carcinogenesis: Incidence curves for childhood and adult tumors. Mathematical Biosciences 47:55-77, 1979.
- Irvine L, Elliott L, Wallace H, Crombie IK. A review of major influences on current public health policy in developed countries in the second half of the 20th century. J R Soc Health. 2006 Mar;126(2):73-8.
- Musk AW, de Klerk NH. History of tobacco and health. Respirology. 2003 Sep;8(3):286-90.
- Parascandola M, Weed DL, Dasgupta A. Two Surgeon General's reports on smoking and cancer: a historical investigation of the practice of causal inference. Emerg Themes Epidemiol. 2006 Jan 10;3:1.
- Parascandola M. Cigarettes and the US Public Health Service in the 1950s. Am J Public Health. 2001 Feb;91(2):196-205.
- Moolgavkar SH, Luebeck EG. Multistage carcinogenesis and the incidence of human cancer. Genes Chromosomes Cancer. 2003 Dec;38(4):302-6.
- Moolgavkar SH. Pollution analysis flawed by statistical model. Nature 2007 Jan 4;445(7123):21.
- Moolgavkar SH. Fine particles and mortality. Comment on: Inhal Toxicol. 2005 Dec 15;17(14):803-16. Inhal Toxicol. 2006 Jan;18(1):93-4.
- Thun MJ, Burns DM. Health impact of ”reduced yield” cigarettes: a critical assessment of the epidemiological evidence. Tob Control. 2001;10 Suppl 1:i4-11. Erratum in: Tob Control 2002 Mar;11(1):85.
- Kabat GC. Fifty years’ experience of reduced-tar cigarettes: what do we know about their health effects? Inhal Toxicol. 2003 Sep 15;15(11):1059-102. Hoffmann D, Hoffmann I. The changing cigarette, 1950-1995. J Toxicol Environ Health. 1997 Mar;50(4):307-64.
- Wynder EL, Muscat JE. The changing epidemiology of smoking and lung cancer histology. Environ Health Perspect. 1995 Nov;103 Suppl 8:143-8. Hoffmann D, Rivenson A, Hecht SS. The biological significance of tobacco-specific N-nitrosamines: smoking and adenocarcinoma of the lung. Crit Rev Toxicol. 1996;26(2):199-211.

- Bartsch H, Dally H, Popanda O, Risch A Schmezer P. Genetic risk profiles for cancer susceptibility and therapy response. Recent Results Cancer res. 2007;174:19-36.
- Clegg LX, Reichman ME, Miller BA, Hankey BF, Singh GK, Lin YD, Goodman MT, Lynch CF, Schwartz SM, Chen VW, Bernstein L, Gomez SL, Graff JJ, Lin CC, hohnson NJ, Edwards BK. Cancer Causes Control. 2008 Nov 12.
- Brske-Hohlfeld, I. Environmental and occupational risk factors for lung cancer. Methods Mol Bio. 2009;472:3-
- Gonzalez CA. The European Prospective Investigation into Cancer and Nutrition (EPIC). Public Health Nutr. 2006 Feb;9(1A):124-6.
- Benowitz NL, Hall SM, Stewart S, Wilson M, Dempsey D, Jacob P 3rd. Nicotine and carcinogen exposure with smoking of progressively reduced nicotine content cigarette. Cancer Epidemiology Biormarkers Prev. 2007 Nov;16(11):2479-85.
- Moolgavkar SH and Luebeck EG. Two event model for carcinogenesis: biological, mathematical, and statistical considerations. Risk Anal. 1990;10:323-341.
- Heidenreich WF, Luebeck EG, Moolgavkar SH. Some properties of the hazard function of the two-mutation clonal expansion model. Risk Anal. 1997;17:391-399.
- Meza R, Hazelton WD, Colditz GA, Moolgavkar SH. Analysis of lung cancer incidence in the Nurses' Health and the Health Professionals' Follow-Up Studies using a multistage carcinogenesis model. Cancer Causes Control. 2008 Apr;19(3):317-28. Epub 2007 Dec 6.
- Hazelton WD, Moolgavkar SH, Curtis SB, Zielinski JM, Ashmore JP, Krewski D. Biologically based analysis of lung cancer incidence in a large Canadian occupational cohort with low-dose ionizing radiation exposure, and comparison with Japanese atomic bomb survivors. J Toxicol Environ Health A. 2006 Jun;69(11):1013-
- Hazelton WD, Clements MS, Moolgavkar SH. Multistage carcinogenesis and lung cancer mortality in three cohorts. Cancer Epidemiol Biomarkers Prev. 2005 May;14(5):1171-81.
- Curtis SB, Hazelton WD, Luebeck EG, Moolgavkar SH. From mechanisms to risk estimation–bridging the chasm. Adv Space Res. 2004;34(6):1404-9.
- Clements MS, Armstrong BK, Moolgavkar SH. Lung cancer rate predictions using generalized additive models. Biostatistics. 2005 Oct;6(4):576-89. Epub 2005 Apr 28.
- Knoke JD, Shanks TG, Vaughn JW, Thun MJ, Burns DM. Lung cancer mortality is related to age in addition to duration and intensity of cigarette smoking: an analysis of CPS-I data. Cancer Epidemiol Biomarkers Prev. 2004 Jun;13(6):949-57.
- Haylock RG, Muirhead CR. Fitting the two-stage model of carcinogenesis to nested case-control data on the Colorado Plateau uranium miners: dependence on data assumptions. Radiat Environ Biophys. 2004 Feb;42(4):257-63. Epub 2003 Nov 15.
- Krewski D, Zielinski JM, Hazelton WD, Garner MJ, Moolgavkar SH. The use of biologically based cancer risk models in radiation epidemiology. Radiat Prot Dosimetry. 2003;104(4):367-76. Review.
- Curtis SB, Luebeck EG, Hazelton WD, Moolgavkar SH. A new perspective of carcinogenesis from protracted high-LET radiation arises from the two-stage clonal expansion model. Adv Space Res. 2002;30(4):937-44.
- Heidenreich WF, Luebeck EG, Hazelton WD, Paretzke HG, Moolgavkar SH. Multistage models and the incidence of cancer in the cohort of atomic bomb survivors. Radiat Res. 2002 Nov;158(5):607-14.

- Curtis SB, Luebeck EG, Hazelton WD, Moolgavkar SH. The role of promotion in carcinogenesis from protracted high-LET exposure. Phys Med. 2001;17 Suppl 1:157-60.
- Little MP, Haylock RG, Muirhead CR. Modelling lung tumour risk in radon-exposed uranium miners using generalizations of the two-mutation model of Moolgavkar, Venzon and Knudson. Int J Radiat Biol. 2002 Jan;78(1):49-68.
- Moolgavkar SH, Brown RC, Turim J. Biopersistence, fiber length, and cancer risk assessment for inhaled fibers. Inhal Toxicol. 2001 Sep;13(9):755-72.
- Hazelton WD, Luebeck EG, Heidenreich WF, Moolgavkar SH. Analysis of a historical cohort of Chinese tin miners with arsenic, radon, cigarette smoke, and pipe smoke exposures using the biologically based two-stage clonal expansion model. Radiat Res. 2001 Jul;156(1):78-94.
- Moolgavkar SH, Turim J, Brown RC. The power of the European Union protocol to test for carcinogenicity of inhaled fibers. Regul Toxicol Pharmacol. 2001 Jun;33(3):350-5.
- Moolgavkar SH, Turim J, Brown RC, Luebeck EG. Long man-made fibers and lung cancer risk. Regul Toxicol Pharmacol. 2001 Apr;33(2):138-46.
- Luebeck EG, Curtis SB, Hazelton WD, Moolgavkar SH. The importance of promotion in lung carcinogenesis for protracted exposures to radon and radon progeny. Radiat Res. 2000 Dec;154(6):730;discussion 730-1. No abstract available.
- Moolgavkar SH. Air pollution and daily mortality in three U.S. counties. Environ Health Perspect. 2000 Aug;108(8):777-84.
- Moolgavkar SH, Luebeck EG, Turim J, Hanna L. Quantitative assessment of the risk of lung cancer associated with occupational exposure to refractory ceramic fibers. Risk Anal. 1999 Aug;19(4):599-611.
- Neus H, Schmann M, Koss G. [Uncertainty analysis of quantitative risk assessment in temporally variable exposures: model observations based on biological and epidemiological risk models] Schriftenr Ver Wasser Boden Lufthyg. 999;103:140-200. German.
- Moolgavkar SH, Luebeck EG, Turim J, Brown RC. Lung cancer risk associated with exposure to man-made fibers. Drug Chem Toxicol. 2000 Feb;23(1):223-42.
- Cardis E, Zeise L, Schwarz M, Moolgavkar S. Review of specific examples of QEP. IARC Sci Publ. 1999;(131):239-304. Review. No abstract available.
- Luebeck EG, Heidenreich WF, Hazelton WD, Paretzke HG, Moolgavkar SH. Biologically based analysis of the data for the Colorado uranium miners cohort: age, dose and dose-rate effects. Radiat Res. 1999 Oct;152(4):339-51.
- Castrn O. Implications of a two-stage clonal expansion model to indoor radon risk assessment. Health Phys. 1999 Apr;76(4):393-7.
- Moolgavkar SH, Luebeck EG, Anderson EL. Estimation of unit risk for coke oven emissions. Risk Anal. 1998 Dec;18(6):813-25.
- Kai M, Luebeck EG, Moolgavkar SH. Analysis of the incidence of solid cancer among atomic bomb survivors using a two-stage model of carcinogenesis. Radiat Res. 1997 Oct;148(4):348-58. Erratum in: Radiat Res 1998 Mar;149(3):309.
- Moolgavkar SH, Luebeck EG, Anderson EL. Air pollution and hospital admissions for respiratory causes in Minneapolis-St. Paul and Birmingham. Epidemiology. 1997 Jul;8(4):364-70.

- Luebeck EG, Curtis SB, Cross FT, Moolgavkar SH. Two-stage model of radon-induced malignant lung tumors in rats: effects of cell killing. Radiat Res. 1996 Feb;145(2):163-73.
- Luebeck EG. http://cran.r-project.org/web/packages/Bhat/Bhat.pdf (2009).
- Stayner L, Smith R, Bailer AJ, Luebeck EG, Moolgavkar SH. Modeling epidemiologic studies of occupational cohorts for the quantitative assessment of carcinogenic hazards. Am J Ind Med. 1995 Feb;27(2):155-
- Review.
- Moolgavkar SH. Biological models of carcinogenesis and quantitative cancer risk assessment. Risk Anal. 1994 Dec;14(6):879-82.
- Moolgavkar SH. Air pollution and mortality. N Engl J Med. 1994 Apr 28;330(17):1237-8. No abstract available.
- Moolgavkar SH, Luebeck EG, Krewski D, Zielinski JM. Radon, cigarette smoke, and lung cancer: a reanalysis of the Colorado Plateau uranium miners' data. Epidemiology. 1993 May;4(3):204-17.
- Moolgavkar SH. Carcinogenesis models: an overview. Basic Life Sci. 1991;58:387-96; discussion 396-9. Review.
- Moolgavkar SH, Luebeck G. Two-event model for carcinogenesis: biological, mathematical, and statistical considerations. Risk Anal. 1990 Jun;10(2):323-41. Review.
- Moolgavkar SH, Cross FT, Luebeck G, Dagle GE. A two-mutation model for radon-induced lung tumors in rats. Radiat Res. 1990 Jan;121(1):28-37.
- Moolgavkar SH, Dewanji A, Luebeck G. Cigarette smoking and lung cancer: reanalysis of the British doctors' data. J Natl Cancer Inst. 1989 Mar 15;81(6):415-20.
- Stevens RG, Moolgavkar SH. A cohort analysis of lung cancer and smoking in British males. Am J Epidemiol. 1984 Apr;119(4):624-41.
- Moolgavkar SH. Model for human carcinogenesis: action of environmental agents. Environ Health Perspect. 1983 Apr;50:285-91.

FLEXKB DOCUMENT
Version: HI.001.03302011.70015
Document generated: 03/30/2011

Pacific Institute for Research and Evaluation

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.

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

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

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

Output Overview

Definitons and methodologies for the basic model outputs.

Results Overview

A guide to the results obtained from the model.

Key References

MODEL PURPOSE

SUMMARY

Brief overview of the purposes driving the development of this model.

PURPOSE

US lung cancer deaths rates have seen dramatic changes in the last twenty years. After increasing through the 1980s, the age-adjusted mortality rate for men has been declining steadily from a high of 90 per 100,000 in 1984 to about 70 per 100,000 in 2005[1]. After steep rises in previous decades, the rate of increase for women began to slow in the 1990s, and has remained at about 40 per 100,000 since 1998[1]. Smoking has been established as the leading cause of lung cancer deaths, with as much as 90% of the deaths due to smoking[2].

The analysis focuses on the role of smoking in explaining trends in lung cancer deaths using the Smoking Base Case data. The data contain information on smoking status, smoking intensity and smoking duration by cohort from 1975 through 2000. We employ a macro modeling approach that uses data aggregated over ages by year and gender. To incorporate smoking related factors, we apply smoking models of lung cancer death rates from past studies[8] to lung cancer deaths in terms of smoking intensity and duration.

The different smoking models yield estimates of lung cancer deaths that vary considerably among themselves and that vary from observed rates over time. Among other reasons, this variation may be due to the non-representative data used to estimate the models and the changing nature of lung cancer risks vis-à-vis smoking, other non-smoking factors (air pollution, asbestos, eating habits) and their interactions. Using the predictions from the smoking models, we consider the role of smoking and non-smoking related trends relative to the predictions of these models. Our model differs from other models in that it does not specifically distinguish age-period-cohort effects, but instead focuses on period effects.

The model also considers the role of tobacco control efforts in reducing lung cancer deaths. Beginning with the Surgeon General's Report in 1964, efforts have been aimed at reducing smoking; bans have been placed on certain types of advertising, clean air laws have been implemented, and higher taxes have been imposed on cigarettes. In addition to the analysis conducted under actual Tobacco Control (TC), we employ smoking data generated for the Smoking Base Case to consider the counterfactual cases of how lung cancer rates would have been affected: 1) in the absence of tobacco control policies since the early fifties, i.e., No Tobacco Control (NTC), and 2) if all smoking was terminated in 1965, i.e., Complete Tobacco Control (CTC).

In sum, the purpose of the analyses is three-fold:

- We compare the predictability of models that examine the relationship of smoking to lung cancer, and thereby provide independent evaluation of these models in explaining levels and trends in lung cancer. Because the models use either CPS-I or CPS-II data, we also indirectly consider how well these two data sets explain lung cancer death rates.
- We attempt to distinguish smoking and non-smoking related factors in explaining trends.
- We estimate how many lung cancer deaths have been avoided as a result of reductions in smoking due to tobacco control efforts implemented since the early fifties and how many deaths could have been avoided if all smoking had stopped as of 1965.

REFERENCES:

[1] JEMAL, A., THUN, M., RIES, L. A., HOWE, H. L, WEIR, H. K. "Annual Report to the Nation on the Status of Cancer, 1975-2005, Featuring Trends in Lung cancer, Tobacco Use, and Tobacco Control" in Journal of the National Cancer Institute 2008;

[2] USDHHS "The Health Consequences of Smoking (Atlanta, GA, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health)" 2004;

[3] DOLL, R., PETO, R., WHEATLEY, K., GRAY, R., SUTHERLAND, I. "Mortality in relation to smoking: 40 years' observations on male British doctors" in Bmj 1994; 309: 901-11

[4] FLANDERS, W. D., LALLY, C. A., ZHU, B. P., HENLEY, S. J., THUN, M. J. "Lung cancer mortality in relation to age, duration of smoking, and daily cigarette consumption: results from Cancer Prevention Study II" in Cancer Res 2003; 63: 6556-62

[5] KNOKE, J. D., SHANKS, T. G., VAUGHN, J. W., THUN, M. J., BURNS, D. M. "Lung Cancer Mortality Is Related to Age in Addition to Duration and Intensity of Cigarette Smoking: An Analysis of CPS-I Data" in Cancer Epidemiol Biomarkers Prev 2004; 13: 949-57

[6] KNOKE, J. D., BURNS, D. M., THUN, M. J. "The change in excess risk of lung cancer attributable to smoking following smoking cessation: an examination of different analytic approaches using CPS-I data" in Cancer Causes Control 2008; 19: 207-19

[7] MEZA, R., HAZELTON, W. D., COLDITZ, G. A., MOOLGAVKAR, S. H. "Analysis of lung cancer incidence in the Nurses' Health and the Health Professionals' Follow-Up Studies using a multistage carcinogenesis model" in Cancer Causes Control 2008; 19: 317-28

[8] HAZELTON, W. D., CLEMENTS, M. S., MOOLGAVKAR, S. H. "Multistage carcinogenesis and lung cancer mortality in three cohorts" in Epidemiol Biomarkers Prev 2005; 14: 1171-81

Model Overview

Purpose

See Model Purpose for details on the purposes behind this study.

Background

The relationship of smoking to lung cancer has been established by biological and epidemiological studies[1]. At a population level, lung cancer has been strongly linked to the smoking dose, both in terms of cigarettes smoked per day and the number of years smoked[7]. These studies are based primarily on large cohort studies, such as the Cancer Prevention Study CPS)-I, and the CPS-II. These data, however, over-represent the middle class, married, Whites, and more educated[9]. In addition, the relative risks of smoking appear to be changing over time[10].

We use the empirical models from 7 different studies of the smoking-lung cancer relationship[7] to predict lung cancer deaths over time. The estimates of lung cancer deaths obtained from inserting historical smoking rate base case data into the models are then compared to historical lung cancer deaths over time using regression analysis. The smoking models[7] are also used to estimate lung cancer rates: 1) in the absence of tobacco control policies since the early fifties (NTC), and 2) if all smoking was terminated in 1965 (CTC). These results are calibrated using the regression analysis from the historical lung cancer case. Finally, we estimate the number of lung cancer deaths avoided as a result of reductions in smoking due to tobacco control efforts and how many deaths could have been avoided if all smoking had stopped as of 1965.

Model Description

A macro-modeling approach is adopted to explain consider male and female lung cancer rates. Smoking Base Case data is employed for populations, lung cancer death and smoking data. The analysis is confined to ages 30-84 and the years 1975-2000.

The smoking data are applied to a series of different smoking models from the literature[7]. Each of the smoking models is a set of equations, as described in the Model Components sections (Component Overview) that relate smoking characteristics to lung cancer death rates. Separate equations are applied to the never, current and former smoker populations to estimate the lung cancer deaths for each gender and age group. For former smokers, individuals are further distinguished by years quit. For current or former smokers, Deaths are related to smoking intensity and duration.

The smoking models yield estimates of lung cancer deaths that differ from observed rates. We fit the predicted rates from the models to historical lung cancer death rates over time, using a method which distinguishes smoking-related and non-smoking-related factors. These analyses are conducted at the aggregated level over all ages by

year, as well as by age. The models are compared in terms of their predictive abilities, as well as their ability to distinguish smoking factors from non-smoking factors. The fitted equations also serve to calibrate the models to actual rates (with past tobacco control efforts, ATC).

To examine the results of tobacco control efforts, we use the smoking models to predict lung cancer deaths under the two counterfactual scenarios of no tobacco control (NTC) and complete tobacco control (CTC). The uncalibrated results are obtained by applying base case data for each of the counterfactual cases to each the smoking models. The results for the Counterfactual cases are calibrated by applying the results from the fitted TC models to predictions from the smoking models for the counterfactual cases.

Using each of the smoking models model, we compare the calibrated and un-calibrated predicted lung cancer death rates under the three scenarios: actual tobacco control, complete tobacco control and no tobacco control. Multiplying by population, we obtain lung cancer deaths by age and gender. The difference between the deaths under the ATC and NTC scenarios is the lives saved as a result of actual tobacco control. The difference between in deaths between the ATC and CTC scenarios is the number of lives that could be saved if smoking were eliminated in 1965. Summing over age groups yields total lives saved by gender.

REFERENCES:

[1] USDHHS "The Health Consequences of Smoking (Atlanta, GA, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health)" 2004;

[2] DOLL, R., PETO, R., WHEATLEY, K., GRAY, R., SUTHERLAND, I. "Mortality in relation to smoking: 40 years' observations on male British doctors" in Bmj 1994; 309: 901-11

[3] FLANDERS, W. D., LALLY, C. A., ZHU, B. P., HENLEY, S. J., THUN, M. J. "Lung cancer mortality in relation to age, duration of smoking, and daily cigarette consumption: results from Cancer Prevention Study II" in Cancer Res 2003; 63: 6556-62

[4] KNOKE, J. D., SHANKS, T. G., VAUGHN, J. W., THUN, M. J., BURNS, D. M. "Lung Cancer Mortality Is Related to Age in Addition to Duration and Intensity of Cigarette Smoking: An Analysis of CPS-I Data" in Cancer Epidemiol Biomarkers Prev 2004; 13: 949-57

[5] KNOKE, J. D., BURNS, D. M., THUN, M. J. "The change in excess risk of lung cancer attributable to smoking following smoking cessation: an examination of different analytic approaches using CPS-I data" in Cancer Causes Control 2008; 19: 207-19

[6] MEZA, R., HAZELTON, W. D., COLDITZ, G. A., MOOLGAVKAR, S. H. "Analysis of lung cancer incidence in the Nurses' Health and the Health Professionals' Follow-Up Studies using a multistage carcinogenesis model" in Cancer Causes Control 2008; 19: 317-28

[7] HAZELTON, W. D., CLEMENTS, M. S., MOOLGAVKAR, S. H. "Multistage carcinogenesis and lung cancer mortality in three cohorts" in Epidemiol Biomarkers Prev 2005; 14: 1171-81

[8] GARFINKEL, L. "Selection, follow-up, and analysis in the American Cancer Society prospective studies" in Natl. Cancer Inst. Monogr. 1985; 67: 49-52

[9] STELLMAN, S. D., GARFINKEL, L. "Smoking habits and tar levels in a new American Cancer Society prospective study of 1.2 million men and women" in J. Natl. Cancer Inst. 1986; 76: 1057-1063

[10] THUN, M. J., HEATH, C. W., JR. "Changes in mortality from smoking in two American Cancer Society prospective studies since 1959" in Prev Med 1997; 26: 422-6

ASSUMPTION OVERVIEW

BACKGROUND

The base case data is based on the large National Health Interview Survey (NHIS). The comparison of the Tobacco Control (ATC) No Tobacco Control (NTC), and Complete Tobacco Control (CTC) is dependent on the assumptions made in deriving the smoking data generated for the base case. Assumptions regarding the development of the Base Case data are described in the Base Case Description. We focus here on assumptions made in the application and at each of the three steps at each of analysis: 1) application of the smoking models, 2) Calibration/validation of the models, and 3) Application of the Models to the Counterfactuals.

LIST OF ASSUMPTIONS

Smoking Models

- The smoking models incorporate all relevant smoking characteristics, namely smoking intensity, smoking duration and years quit.
- The smoking models are time invariant. Studies indicate that relative risks may have changed over time, e.g., due to non-smoking related factors (radon or other indoor or outdoor pollution), composition of cigarettes, which is considered at a later step
- Non-linearities are captured at the age level. Because the models were estimated by age group (as defined by the Base Case data), we captured non-linearities as they apply to different ages. However, nonlinearities are implied by the different models within age groups due to the variation of duration and intensity among those a particular age group. Since duration and intensity are likely to be correlated, data on their joint distributions would be required to accurately capture these effects.
- Except for age and gender-related differences, the effects are homogeneous across socio-demographic groups. Some previous studies indicate important differences in lung cancer rates and the role of smoking by race[1], but the base case data and smoking models employed do not distinguish by race.

Assumptions related to specific models

- The TCSE equations were estimated only for Whites, and therefore may be biased as applied to other populations.
- Knoke et al. did not estimate equations for females and limited their analysis to the White population.
- Flanders used data for all racial-ethic groups, and thus may be subject to aggregation bias across racial-ethnic groups. Flanders has no former smoker equation, so the male former smoker equation from Knoke et al. was used to capture the decline in risk with years quit for male and female smokers.

Calibration/Validation of the Model

- The models were calibrated over the entire range of years and ages. Consequently, results could only be compared across models and could not be validated for individual models. The calibration equations assumed a specific form, whereby trends in smoking and non-smoking related factors are captured by quadratic equations.
- The effects of smoking are assumed to be captured by the smoking models. In particular, smoking effects by age and cohort are assumed to be captured by the smoking models.
- The calibration equations assumed a specific functional form, whereby log vs. linear and liner vs. quadratic forms are considered.

Application of the Models to the Counterfactuals

As indicated above, important assumptions are made in developing the base case data on smoking rates, as discussed in the Base Case Data description. Assumption made in our analysis include:

1. The smoking models can be applied to smoking rates outside the usual ranges. The results from the smoking models may be sensitive to functional form in moving to smoking rates that deviate from the historical, especially for the CTC models which are sensitive to number of years quit. With large changes in the percent of the population smoking, changes in the likelihood of second hand smoke exposure may affect never smoker risks.
2. The calibration of the models developed using historical smoking rates are assumed to be applicable to the counterfactual cases, where smoking rates are different. The calibrations may be sensitive to functional form in moving to smoking rates that deviate widely from the historical rates.

REFERENCES:

[1] USDHHS "The Health Consequences of Smoking (Atlanta, GA, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health)" 2004;

PARAMETER OVERVIEW

BACKGROUND

To predict lung cancer deaths, separate sets of equation are developed using empirical models from 7 different studies of the smoking-lung cancer relationship. Two sets of parameters are used to develop the final predictions of lung cancer deaths. The first set relates smoking characteristics to lung cancer death rates using different empirical models of the smoking–lung cancer relationship, and the second set is from equations that relate the predictions from the smoking models to historical lung cancer death rates.

Once models have been developed for historical lung cancer deaths, the same empirical models are applied to the counterfactual scenarios (no tobacco control and complete tobacco control), with the exception that the second set of equations from the historical lung cancer case is used to calibrate the counterfactuals.

PARAMETER LISTING OVERVIEW

Each of the smoking models is set of equations that relate smoking characteristics to lung cancer death rates. The smoking characteristics first distinguish whether the individual is a current, former or never smoker. For smoker smokers, individuals are further distinguished by years quit. If a current or former smoker, smoking intensity and smoking duration are considered. Data for the Smoking Base Case were provided on smoking prevalence, smoking intensity and smoking duration provided by gender, cohort (aggregated over 5 years beginning in 1900) and year. Smoking prevalence data were distinguished by 3 smoking categories: never smokers, current smokers, and former smokers, with former smokers further divided into 5 groups based on how long ago they quit: 1-2, 3-5, 6-10, 11-15, and 16+ years. Thus, smoking status is defined by 7 categories (never, current and 5 former). The smoking status prevalence rates were applied to the total population to obtain the number of never, current and former (by years quit) smokers by age and year.

For smokers and former smokers by years quit, intensity data were provided in terms

of the average number of cigarettes smoked per day (CPD). Duration of smoking (DUR) was calculated from average initiation data, as the current age minus the age of smoking initiation for smokers, with further correction for years quit by former smokers.

For each measure, we converted the data by (5 year) cohort and year into age groups by year starting with ages 30-34 and continuing through ages 80-84. To obtain data by single age, we smoothed (i.e. interpolated) over 5 ages starting at age 32 and ending at age 82. For ages 30-31 and 83-84, we extrapolated from the two nearest interpolated ages. Because reliable data were not available for cohorts before 1900, values in the base case were inferred for missing values of the smoking rate variables for ages 75-84 in the years 1975-1985.

Besides data for historical smoking prevalence, intensity and duration under Actual Tobacco Control (ATC) as implemented, separate data for each for the variables for the No Tobacco Control (NTC) and Complete Tobacco Control (CTC). Smoking Base Case Accordingly, separate data sets by gender, age, and year were created for each of these scenarios.

Parameter Listings for each model

1. Using models from past studies of the smoking-lung cancer relationship, Predicted Lung Cancer Rates (PLCR) are estimated as a function of:

- Smoking Status: Current, Former or Former smoker by Age (Base Case Data)
- If Current Former Smoker:
  - Smoking duration in years (Base Case Data)
  - Smoking intensity in terms of average number of cigarettes smoked per month (Base Case Data)
  - Age
- If Former Smoker:
  - Same variables as for Smokers with smoking duration and intensity related to the period when smoking occurred. (Base Case Data)
  - Years quits
- Using time series analysis, Historical Lung Cancer Rates (Base Case Data) are related to
- Predicted Lung Cancer Rates (PLCR) examining the following specifications:
  - Linear vs Log
  - Prediction, Prediction X Time Trend, Prediction X Time Trend^2
    - Time-trends correspond to years
- Time-trends independent of smoking predictions examining the following specifications
  - Linear vs log
  - Time trend vs. Time Trend^2 or both

COMPONENT OVERVIEW

OVERVIEW

The models relate smoking factors to deaths rates. As such they contain a population and a survival/mortality component.

Population Component

There is not a population model per se with births and deaths. Rather, historical population data is obtained from Smoking Base Case data from the Census data. The data used in the model is obtained by year beginning at age 30 and continuing through age 84, distinguished by gender. The smoking status prevalence rates were applied to the total population to obtain the number of never, current and former (by years quit) smokers by age, gender and year.

Survival Mortality Component

Two sets of equations are used to develop the final predictions of lung cancer deaths. The first set (smoking models) relate smoking characteristics to lung cancer death rates, and the second set (Prediction And Calibration Equations) takes the predictions from the smoking models and fits them to historical lung cancer death rates. The fitted models are applied then applied to the No Tobacco Control and Complete Tobacco Control Cases.

SMOKING MODELS

We consider three sets of models[6], all of which relate lung cancer mortality risk to age, duration, and intensity of smoking and employ either the CPS-I or CPS-II cohort data. Using each of these models, we estimate lung cancer death rates for never, current, and former smokers under each of the three smoking scenarios. An Excel program was used to conduct the analysis.

Knoke model

Knoke et al.[4] estimated separate equations for the absolute lung cancer death risk of never smokers, smokers and former smokers using only the white, male subpopulation of the CPS-I. Consequently, we only estimate death rates for males using the Knoke model.

The lung cancer death rates were estimated in terms of excess risk (ER), where the mean absolute risk (R) for smokers (S) and former smokers (FS) is added to never smokers (NS), or:

$$R_S = R_{NS} + ER_S,$$

$$R_{FS} = R_{NS} + ER_{FS}$$

Knoke et al. modeled the absolute risk of death due to lung cancer in nonsmokers to be a two-parameter Poisson regression model on attained age, in years:

$$R_{NS} = 9.21 \times 10^{-13} \times (age^{2.38}).$$

They modeled the excess risk of death due to lung cancer in continuing smokers as a Poisson regression model with modified offset, which they describe as an extended Doll and Peto model[7] as suggested by the multistage theory of carcinogenesis. The mean excess risk was calculated as follows:

$$ER_S = [1.51 \times 10^{-13} \times (AGE^{2.38}) \times (CPD^{0.867}) \times (DUR^{2.87})],$$

They modeled the lung cancer death rate as a function of years since quit (QT-YRS) and quit age (QT-AGE) for former smokers as:

$$R_{FS} = R_{NS} + f(QT_{YRS}, QT_{AGE}) \times ER_S,$$

The excess smoker risk, ER_S, continues to denote the excess risk as if the individual continued to smoke at the same intensity and duration when that individual quit. They found that there was no decline in risk for the first two years after quitting. For more than two years quit, they found no significant effect of CPD with the following equation:

$$f(QT_{YRS}, QT_{AGE}) = exp[-(0.274 - 0.00279 \times QT_{AGE}) \times (QT_{YRS-2})].$$

Flanders Model

Flanders et al.[2] used the CPS-II data for all races to estimate separate lung cancer death rate equations for male and female smokers by 10 year age groups. They did not estimate equations for never or former smokers.

For smokers' lung cancer death rates (LMR) for males (m) and females (f), they estimated equations by gender similar to the Doll and Peto[7] models and obtained:

ages 40–49: $LMRm = e^{17.9} \times DUR^{1.9} \times CPD^{0.95}$ $\quad$ $LMRf = e^{20.2} \times DUR^{2.8} \times CPD^{0.96}$

ages 50–59: $LMRm = e^{17.4} \times DUR^{2.6} \times CPD^{0.52}$ $\quad$ $LMRf = e^{17.2} \times DUR^{2.2} \times CPD^{0.75}$

ages 60–69: $LMRm = e^{15.7} \times DUR^{2.4} \times CPD^{0.37}$ $\quad$ $LMf = e^{14.1} \times DUR^{1.5} \times CPD^{0.78}$

ages 70–79: $LMRm = e^{13.0} \times DUR^{1.8} \times CPD^{0.39}$ $\quad$ $LMRf = e^{13.2} \times DUR^{1.3} \times CPD^{0.95}$

The estimates for ages 40-49 are applied to ages 30-39, and the estimates for ages 70-79 applied to ages 80-84.

Because they did not estimate equations for former smokers, we apply the former smoker equation from Knoke et al.[4] to smoker death rates from Flanders to obtain the

declining death rates of former smokers by years quit. Estimates of deaths rates for never smokers were obtained from Thun et al.[8] based on CPS-II estimates by age and gender for the White population.

Two stage clonal expansion model

Hazelton et al. (2005) separately applied the two-stage clonal expansion (TSCE) model to both the CPS-I and CPS-II, thus providing a common model applied for both data sets. They estimated separate equations for White males and White females. We apply their model separately by gender and for each of the different data sets, CPS-I and CPS-II.

Hazelton et al.[6] estimated a series of non-linear equations (not reproduced here, but available in the Appendix to their paper) in terms of the rates of initiation, cell division, apoptosis of initiated cells, and the rate of malignant conversion of initiated cells under the TSCE model. These rates are a function of smoking intensity and duration. Due to a lack of available data, they did not include former smokers in their estimation equations, but the model provides death risks for that group. The programming for the estimation equation models were made available to us by the authors as an excel add-in.

Summary

A total of seven different results are developed using each of the models (4 male and 3 female). The male Knoke et al. model and male and female TSCE CPS-I models use CPS-I data, while the Flanders et al. male and female models and the male and female TSCE CPS-II models use CPS-II data. The smoking data are applied by year and age and by smoking status.

Upon applying the predicted deaths rates by age, smoking status and year to their respective populations, we obtain the predicted number of lung cancer deaths by age, smoking status, and year for each of the models. The deaths are summed over the 7 smoking status categories for each model to get predicted total lung cancer deaths by age and year and then summed over age groups to get lung cancer deaths for all ages. We divide the number of lung cancer deaths in each age group by their respective age and year population to obtain lung cancer death rates per 100,000. For each model, the data are applied to the three smoking scenarios (Actual Tobacco Control (ATC), No Tobacco Control (NTC) and Complete Tobacco Control (CTC)) to get lung cancer deaths and death rates by age and year.

Calibration Methods

To control for population size, results from the TSCE models were converted to rates and calibrated against the lung cancer death rates rather than total lung cancer deaths. These models were estimated using data aggregated over the 30 to 84 age group for each of the years from 1975 to 2000. We consider general trends for the population as a whole after netting out smoking factors, as incorporated in the TSCE models. Specifically, for each model, we regressed the historical lung cancer rates (HLCR) against the predicted lung cancer rates (PLCR) alone, PLCR interacted with time

trends, and separate time trend factors:

HLCRt = b0+ b1*PLCRt+ b2 TTt*PLCRt+ b3 TTt2 *PLCRt+ a1TTt+ a2 TTt2+ e t,

where TT denotes a time trend (TT= 1 in 1975, TT = 2 in 1976, …, TT = 26 in 2000), subscript t = year (t= 1975, …, t = 2000), and e is the error term. The first part of the equation shows the influence of smoking as predicted by the TSCE model (b1) with any constant deviation (b0), while the next two terms correct the predictions of the smoking model for linear (b2) and non-linear (b3) biases over time. The second part of the equation allows for linear (a1) and non-linear (a2) trends not captured by the smoking model. We also estimated log specifications of the above model, which imply a proportional rather than constant deviation of model predictions from the historical rates, and a multiplicative rather than linear relationship between the variables.
Because reliable data were not available for cohorts born before 1900, values in the SBC data were inferred for smoking rate variables for the older missing ages in the years 1975 to 1985. To check for bias in the method used to calculate these values, we included a correction factor in the estimation equation equal to the number of age years with missing variables: 10 = 1975, 9= 1976,…, 1= 1984, 0 = 1985 and above. The correction factor was generally insignificant and induced autocorrelation. Consequently, it was dropped from the model.
Our goal was to determine the most parsimonious model with the highest level of predictability and with predictions that were not systematically biased over time. Predictability was gauged by the adjusted R-square. Systematically biased predictions were gauged by autocorrelation in the error terms. The Durbin-Watson (D-W) statistic was used to test for first order correlation of the error terms et and et-1. We began with a simple model that regressed HLCR on PLCR and a constant term. We then added variables to the equation, and kept those variables in the equation if the coefficient of the variable had a t-statistic

Lung Cancer Deaths under the No, Actual and Complete Tobacco Control Scenarios

For each of the 4 models (male and female TSCE CPS-I and TSCE CPS-II) we compared the un-calibrated and calibrated predicted lung cancer death rates under the NTC, ATC and CTC scenarios. To calibrate the counterfactual NTC and CTC predictions we applied the best-fitting calibration equations from the ATC model (for the corresponding gender and CPS model) to the predicted lung cancer rates for NTC and CTC. We also calibrated by multiplying the NTC and CTC predicted rates by a correction factor, measured as the ratio of the historical to the corresponding (gender, year and CPS-type data set) ATC rates. Corrections were separately applied to each estimate of lung cancer deaths in two ways: using a measure 1) aggregated over all ages by year and 2) distinguished by each age and year.

References:

[1] DOLL, R., PETO, R., WHEATLEY, K., GRAY, R., SUTHERLAND, I. “Mortality in relation to smoking: 40 years' observations on male British doctors” in Bmj 1994; 309: 901-11

[2] FLANDERS, W. D., LALLY, C. A., ZHU, B. P., HENLEY, S. J., THUN, M. J. “Lung cancer mortality in relation to age, duration of smoking, and daily cigarette consumption: results from Cancer Prevention Study II” in Cancer Res 2003; 63: 6556-62

[3] KNOKE, J. D., SHANKS, T. G., VAUGHN, J. W., THUN, M. J., BURNS, D. M. "Lung Cancer Mortality Is Related to Age in Addition to Duration and Intensity of Cigarette Smoking: An Analysis of CPS-I Data" in Cancer Epidemiol Biomarkers Prev 2004; 13: 949-57

[4] KNOKE, J. D., BURNS, D. M., THUN, M. J. "The change in excess risk of lung cancer attributable to smoking following smoking cessation: an examination of different analytic approaches using CPS-I data" in Cancer Causes Control 2008; 19: 207-19

[5] MEZA, R., HAZELTON, W. D., COLDITZ, G. A., MOOLGAVKAR, S. H. "Analysis of lung cancer incidence in the Nurses' Health and the Health Professionals' Follow-Up Studies using a multistage carcinogenesis model" in Cancer Causes Control 2008; 19: 317-28

[6] HAZELTON, W. D., CLEMENTS, M. S., MOOLGAVKAR, S. H. "Multistage carcinogenesis and lung cancer mortality in three cohorts" in Epidemiol Biomarkers Prev 2005; 14: 1171-81

[7] DOLL, R., PETO, R. "Cigarette smoking and bronchial carcinoma: dose and time relationships among regular smokers and lifelong non-smokers" in J Epidemiol Community Health 1978; 32: 303-13

[8] THUN, M. J., HENLEY, S. J., BURNS, D. et al. "Lung cancer death rates in lifelong nonsmokers" in J Natl Cancer Inst 2006; 98: 691-9

SMOKING HISTORY GENERATOR COMPONENT

SUMMARY

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 lung cancer consortium.

OVERVIEW

The core SHG software was parameterized using three tobacco control scenarios to 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-

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**

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

Inupt	**ATC**	**NTC**	**CTC**
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 (1890-1984)	User Defined		
Gender (Male/Female)	User Defined		
Race (All race)	User Defined		

[1] Cigarettes smoked per day,[2]Other 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

Parameter	Valid Values
Seed value for PRNG used for Initiation, Cessation, OCD[1], Smoking intensity quintile	Integer from -1 to 2147483647 (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, Smoking intensity quintile and CPD[4] data files	As derived from NHIS depending on the scenario

[1]Other 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. [3]Key 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. [4]Cigarettes 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 History	Smoking intensity quintile (5 quintiles ranging from light to heavy smoking), Yearly smoking dose (CPD[2])

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

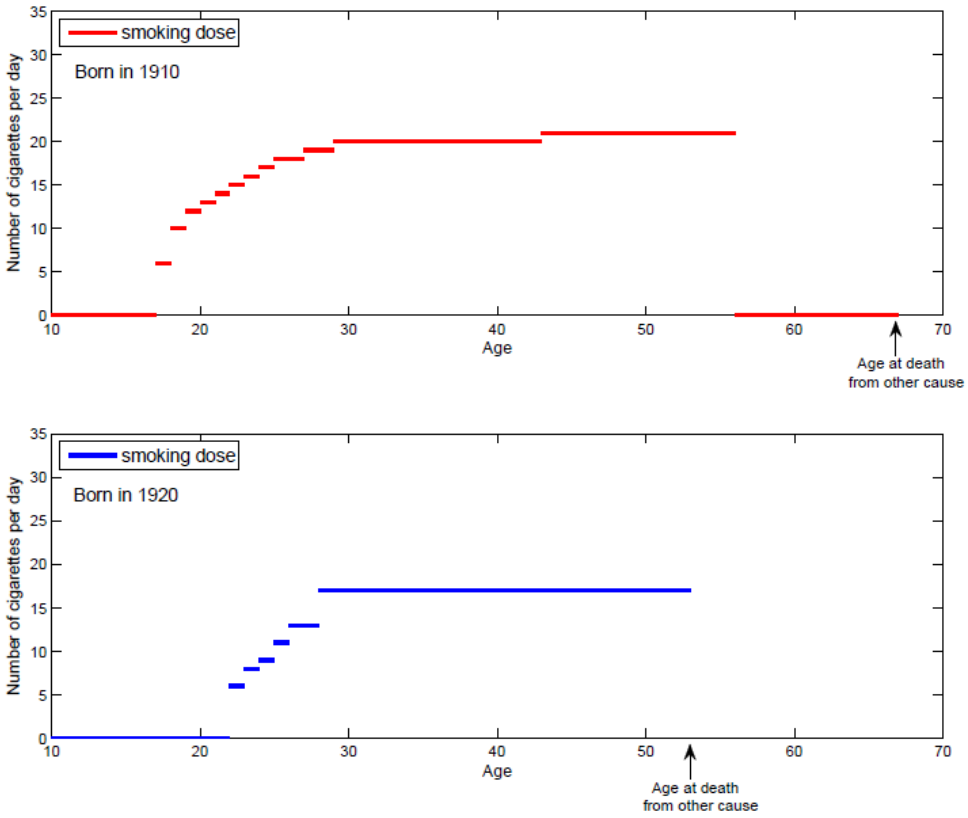


Figure SHG-1: Examples of the SHG-Generated Events

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." 1998; 8: 1: 3-30

OUTPUT OVERVIEW

OVERVIEW

Three sets of equations are used to develop the final predictions of lung cancer deaths, and each are an important part of the output. They are done sequentially.

The first set relates smoking characteristics to lung cancer death rates using the smoking models describeded in the Component Overview Section. Calculations of lung cancer rates are computed by age, year and gender. We calculate the predicted lung cancer rates separately for each of the smoking models (Knoke based on CPS-I, TSCE-CPS-I, TSCE-CPS-II, and Flanders-CPS-II). For a particular year, we sum the predicted lung cancer rates over ages 30-84 for a particular gender using each of the models. Thereby, for each smoking model, we obtain a predicted lung cancer death rate over ages 30-84 for each year and by gender.

The second set of outputs is from equations that relate the predictions from the smoking models to historical lung cancer death rates. For a particular gender and smoking model, we regress the predicted lung cancer rate (aggregated over ages) from the the smoking model on historical lung cancer rates, using the functional forms for the regression equations described in the components section. These results are used to calibrate the models and to distinguish smoking and non-smoking related factors using methods described in the Component Overview Section. They are conducted at two levels of aggregation: combined ages 30-84 by gender and year and by individual age by gender and year.

The third set of results is the development of the counterfactual cases for no tobacco control and complete tobacco control. We develop uncalibrated and calibrated results, with calibrations as developed using actual tobacco control predictions from the second set of results and period and age-period calibrations described in the Component Overview section. The no tobacco control and complete tobacco control results are combined with the actual tobacco control results to estimates lives saved as a result of tobacco control and the potential lives saved with complete tobacco control as described in the Component Overview Section

Results Overview

Overview

Macro Model Aggregate Approach

We have confined analysis presented here to the TSCE models, but using two different data sources: CPS-I and CPS-II. Separate models are developed by gender.

Predictions of Lung Cancer Deaths Rates from the TSCE Smoking Models

Figure1 shows male and female predictions of the TSCE models as well as historical lung cancer death rates.

Figure2a shows male and female breakdowns by age, and similarly Figure2b shows female rates. Figure3 shows male and female predicted lung cancer rates respectively for never, current, and former smokers.

Historical male lung cancer death rates for those ages 30 to 84 increase through 1985, plateau through 1990, and then fall from 1990 onward. Both male models predict rates that are considerably less than historical rates for all years. The trend predicted by the CPS-I model is similar to that of historical lung cancer rates, with similar percentage increases albeit at a lower initial level from 1975 to 1981. The lung cancer rates predicted by the male CPS-II model begin to decline almost immediately but are considerably closer to historical rates than the CPS-I model. The CPS-II model mirrors the trend albeit at a lower initial level from 1989 to 2000.

For the 30 to 49 age group, historical lung cancer death rates followed a continual downward trend with some flattening beginning in 1990. The rates predicted by both models showed similar patterns, but with less decline. The rates predicted by the CPS I model were above those predicted by the CPS-II model, but below historical rates except for 1995 to 2000. Unlike for those ages 50 and above, rates predicted by the CPS-I model were greater than the CPS-II model for those below age 50. For males in the 50 to 69 and 70 to 84 age groups, historical rates increased until about 1990 and then declined, with a steeper decline in the 50 to 69 age group. For males age 50 to 69, both smoking models predicted rates considerably below historical rates with relatively flat rates until about 1988, followed by a decline albeit at a slower rate than historical rates. For males ages 70 to 84, the age range during which most lung cancer deaths occur, the CPS-I model yielded considerably lower predictions than the CPS-II model and historical trends, but mirrored changes in historical trends more closely than the CPS-II model. The CPS-II model predictions begin near historical predictions, but diverge over time. Generally, the male models under-predict historical rates; the CPS-I models better mirrored changes in trends, but CPS-II models yielded predictions closer to those of historical rates and mirrored trends in the 1990 to 2000 period. The projected trends from both models converged toward historical rates for younger smokers, but diverged for older smokers.

For females, historical lung cancer death rates increased over the entire period studied, except for the year 1999, but begin to flatten around 1990. Although the rates predicted by the CPS-I model show similar trends to the historical rates, the gap widens from

50% of historical rates in 1975 to 25% of historical rates in 2000. The rates predicted by the CPS-2 model are close (about 90%) to historical rates in 1975, with the gap increasing to about 60%. For females age 30 to 49, the CPS-I and CPS-II models were considerably below historical rates and did not fall as rapidly over time. For females age 50 to 69, historical rates increased quite rapidly through 1993 and then fell. The predictions of the two models, especially the CPS-I model, were below and much flatter than historical rates. For females age 70 to 84, the upward trend exhibited by both models, especially the CPS-I model, was considerably less than the historical rates. The CPS-II model, and to a much greater extent the CPS-I model, predicted lung cancer death rates below historical rates, but both under-predict the change in rates relative to historical changes. Like for males, the projected trends from both models converged toward historical rates for younger smokers, but diverged for older smokers.

By smoking status, the smoking models (3a, 3b, 3c) predict that male lung cancer death rates vary little over time for never smokers and decrease for smokers. Different patterns are observed for ex-smokers from the CPS-I and CPS-II models, increasing until about 1990 and then decrease for former smokers using CPS II and continuously increasing using CPS I. The CPS-II model predicts higher death rates than the CPS-I model, except for never smokers. For females, the CPS-I and the CPS-II models predict a slow continuous decline for never smokers, a rise until 1988 followed by a decline for smokers, and a rapid rise until 1995 and then a tapering off for former smokers.

Calibration of Predictions of the ATC Model to Historical Lung Cancer Death Rates

Table1a, Table1b, Table2a, and Table2b show the results from our calibration equations for males and females respectively. The requirement of no serial correlation was rejected for models with only the non-interacted predicted values from the smoking models. For both the log and linear models, the predictability improved and serial correlation was reduced to acceptable levels when PLCR variables were interacted with time trends.

For the both the CPS-I and CPS-II male linear models, the models failed to reject serial correlated errors except when both the $PLCR_t \times TT_t$ and $PLCR_t \times TT_t^2$ variables were included in the estimation equation. In addition, the adjusted R-square was marginally higher when the non-interacted PLCR variable was dropped and the D-W statistic and adjusted R-square improved when a non-interacted time trend variable was added for both models. The CPS-I and CPS-II linear models with the $PLCR_t \times TT_t PLCR_t \times TT_t^2$ and TT variables (eqn. 6) have strong explanatory power (as indicated by R-square values that exceed 0.99), exhibit no significant autocorrelation (i.e. the D-W statistic is within the 1.7-2.3 range), and all variable coefficients are statistically significant. For the log models, the equations with a non-interacted $PLCR_t$ and $PLCR_t \times TT_t$ (eqn. 10

for CPS-I and CPS-II) performed best in terms of non-serially correlated errors and the adjusted R-square (above 0.99). When a non-interacted time trend was added, its coefficient was significant and serial correlation was reduced (eqn. 11).

Similar to the male models, the variables $PLCR_t \times TT_t$ and $PLCR_t \times TT_t^2$ were required in the linear models for females to reduce autocorrelation to insignificant levels. For the CPS-I model, the equations with $PLCR_t \times TT_t$ and $PLCR_t \times TT_t^2$ (eqn. 4) or with $PLCR_t \times TT_t$ and $PLCR_t \times TT_t^2$ and TT_t (eqn. 6) performed best in terms of the D-W statistic and the adjusted R-square. For the CPS-II linear model, the adjusted R-squares were higher when the non-interacted PLCR variable was dropped (eqns. 4 and 6), and the D-W statistic improved when a non-interacted TT variable was added (eqn. 6). In log form, the CPS-I models with either the variables $PLCR_t \times TT_t$ and $PLCR_t \times TT_t^2$ or with the variables $PLCR_t$ and $PLCR_t \times TT_t$ and TT_t^2 performed best. The log model with $PLCR_tPLCR_t \times TT_t$ and TT_t induced severe multicollinearity. The CPS-II female models in log form performed about equally well in terms of serial correlation and the adjusted R-square when either the variables $PLCR_tPLCR_t \times TT_t$ and TT_t (eqn. 11) or $PLCR_t \times TT_t$ and $PLCR_t \times TT_t^2$ (eqn. 12) were included. These models, along equations 4 and 6, had adjusted R-square values above 0.99 and no detectable serial correlation of the error terms.

Disaggregated Approach by Age

To be done

SMOKING BASE CASE

A consortium of investigators developed independent mathematical models to measure the impact of declines in smoking initiation and the rise in smoking cessation on lung cancer mortality. Using common inputs, they estimated the number of lung cancer deaths avoided over the period 1975-2000, and also the number of deaths that could have been averted had tobacco control been completely effective in eliminating smoking in 1965, shortly after the issuance of the Surgeon General's report in 1964.

Prediction And Calibration Equations

Overview

A second set of equations relate the predictions from the smoking models to historical lung cancer death rates Base Case data were provided on historical U.S. lung cancer deaths by age, gender and year for the years 1975 to 2000. Lung cancer deaths per 100,000 by age and year were calculated by dividing the total number of lung cancer deaths by population and multiplying by 100,000.

We seek to develop parsimonious models that capture age-related smoking and non-smoking trends in lung cancer deaths. Using lung cancer deaths rates estimated for each of the seven models under actual tobacco, each model's predictions under the TC scenario are separately calibrated against historical levels of lung cancer death rates.

The age, period and cohort-related changes are implicitly assumed to be captured by the smoking models. Because the smoking models were estimated using either CPS-1 or CPS-II data, which provide different relative risks and neither of which are representative of the U.S. population, their predictions of levels and trends in lung cancer rates may be biased for the population at large. In addition, the models applied from the literature differ. To correct for these potential biases, we apply correction factors to the predicted rates from the smoking models. We assume that age and cohort-related changes are captured by the smoking models described above. In addition, we simultaneously consider trends not related to the smoking model predictions to control for omitted non-smoking factors.

To control for the size of population, the models are calibrated against the lung cancer death rates, rather than total lung cancer deaths. The models are estimated using data for each of the years 1975-2000 for 1) all age groups aggregated (i.e., using lung cancer death rates for all 30-84 year olds), 2) by five year age group, and 3) stacked by age and year.

Detail

Models Aggregated over all Ages

In the models aggregated over all age groups, age and cohort-related changes are implicitly assumed to be captured by the smoking models described in the last section in an attempt to isolate the role of age- and smoking-related factors and unaccounted for trends in the effect of smoking. We also consider other trends, i.e., residual trends not captured by the smoking models.

We regress the historical lung cancer rate (HLCR) on the lung cancer rate predicted by a particular model (PLCR) alone and interacted with trends, as well as with separate trend factors. Specifically, our complete estimation equation for each of the seven models is the following:

$$HLCR_t = {}_{a0} + {}_{b0} \times PLCR_t + b_1 TT_t \times PLCR_t + {}_{b2} TT_t^2 \times PLCR_t + a_1 TT_t + a_2 TT_t^2 + e_t,$$

where TT denotes a time trend (TT= 1 in 1975, TT = 2 in 1976, …TT = 26 in 2000), subscript t = year (t= 1975, …, t = 2000), and e is the error term. The first part of the equation examines the role of smoking as predicted by a particular model (b_0) correcting the predictions of the smoking model for linear (b_1) and non-linear (b_2) trends and the second part allows for differences in linear (a_1) and non-linear (a_2) trends not captured by the smoking model.

Equations are estimated for each of the seven models. Our goal is to develop the simplest "unbiased" model with the highest adjusted R-square. Serial correlation of the error terms, e_t and e_{t-1}, is an indication of bias in the model from omitted factors, and will lead to biased predictions. To test for serial correlation, we use the Durbin Watson statistic. To gain model simplicity, we drop terms if the coefficient of a variable has a t We also considered log specifications of the above models to determine if that functional form provides a better fit. These models did not improve the fit and are not reported. In addition, to consider bias from values inferred for the missing smoking rate variables for ages 75-84 in the years 1975-1985, we created a correction factor, equal to the number of age years with missing variables: 10 = 1975, 9= 1976,…, 1= 1984, 0 = 1985 and above), and included that variable in the estimation equation to detect potential measurement error. This variable was generally insignificant and induced serial correlation, and was thus dropped from the model.

As a test of whether the smoking models provide predictability beyond simple trends, we compare each of the models to a quadratic trend model. The time trend model is of the form

$$HLCR_t = (a_{0,a} + a_{1,a} TT_t + a_{2,a} TT_t^2).$$

A statistical test proposed by Harvey et al.[1] is used to test the full model against the trend model. We also consider serial correlation as well as improved fit.

Once estimates are developed for each of the seven models, we compare the different models for each gender separately, also distinguishing models that used CPS-I against those using CPS-II data to generate their original models. The models are compared in terms of the parsimony of the fitted calibration equation (i.e., the ability of predictions of the smoking model alone to capture trends), the adjusted R-square, and serial correlation.

Equations by Age and Year

In this analysis, a single equation is estimated as in our aggregated model, except that there are separate observations by age group stacked by year. The estimation equation is:

$$HLCR_{t,a} = (b_{0,a} + b_1 TT_{t,a} + b_{2,a} TT_{t,a}^{\,2}) \times PLCR_{t,a} + (a_{0,a} + a_{1,a} TT_{t,a} + a_{2,a} TT_{t,a}^{\,2}) + e_{t,a}$$

where TT denotes a time trend (TT= 1 in 1975, TT = 2 in 1976, …TT = 25 in 2000), subscript a is the 5 year age group (a = 1 for ages 30-34, a = 2 for ages 35-39,…, a = 11 for

ages 80-84). subscript t = year (t= 1 in 1975, t = 2 in 1976, …t = 25 in 2000), and e is the error term. The above model is a fixed effects model for age a_{0,a}. We estimate random components models where the fixed effect is a special case.
In this case, we explicitly allow for differences in age effects, along with time trends. While cohort effects are assumed to be captured by the smoking models, we also directly test for cohort effects in the 5 year age group models by examining the correlation between e_{t,a} and e_{t-5,a-5}, where observations for ages 30-34 are dropped. While the coefficients are by age, the results are examined for similar patterns in the estimated coefficients for contiguous age groups.

Lung Cancer Deaths under the No, Actual and Complete Tobacco Control Scenarios

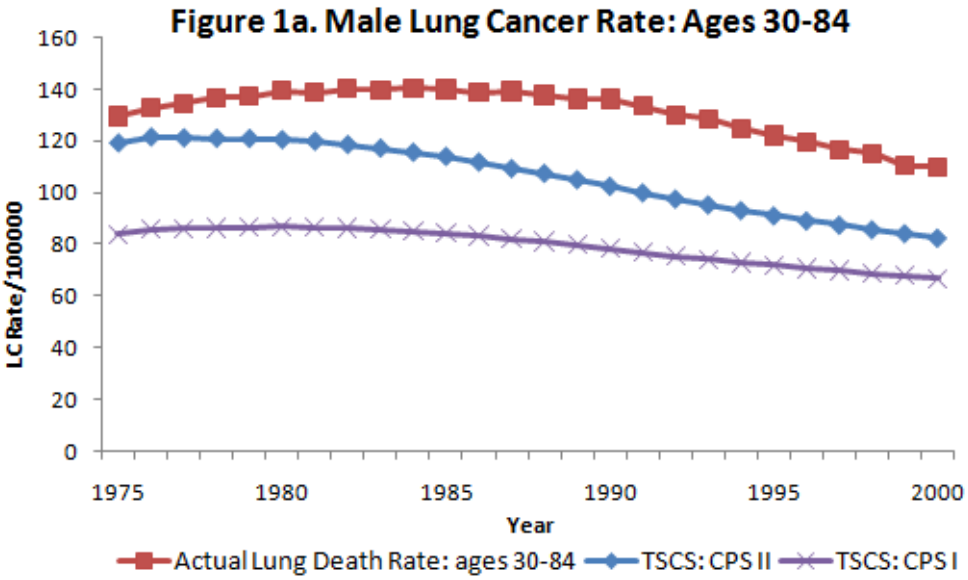
Under the TC, NTC and CTC scenarios, we summed the values under each model over age groups to obtain the total lung cancer deaths for ages 30 through 84 under the respective scenario. For each or the seven models, (TSCE CPS-I, TSCE CPS-II and Flanders both male and female and Knoke Male), we compare the predicted lung cancer deaths and death rates under the three scenarios. The difference between lung cancer deaths between the actual and no tobacco control scenario is the lives saved as a result of actual tobacco control. The difference between actual and complete tobacco control is the lives that could be saved if smoking were eliminated in 1965. Summing over genders yields total lives saved. Using population data to estimates rates, we also comparing lung cancer death rates under each scenario.

We conduct the analysis for both the un-calibrated and for the best fitting, parsimonious calibrated models. To calibrate the counterfactual NTC and CTC predictions we applied the best-fitting calibration equations from the ATC model (for the corresponding gender and CPS model) to the predicted lung cancer rates for NTC and CTC, under the assumption that the lung cancer risks would otherwise be influenced by the same smoking and non-smoking trends. We also calibrated by multiplying the NTC and CTC predicted rates by a correction factor, measured as the ratio of the historical to the corresponding (gender, year and CPS-type data set) ATC rates. Corrections were separately applied to each estimate of lung cancer deaths in two ways: using a measure 1) aggregated over all ages by year and 2) distinguished by each age and year.

REFERENCES:

[1] HARVEY, D., LEYBOURNE, S., NEWBOLD, P. “Tests for Forecast Encompassing” in Journal of Business and Economic Statistics, American Statistical Association 1998; 16: 254-69

FIGURE1



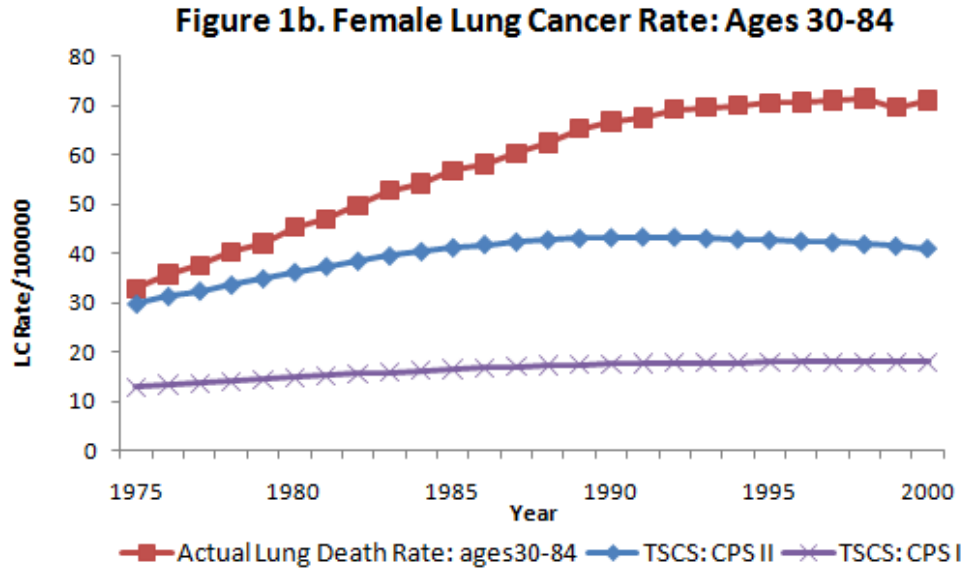
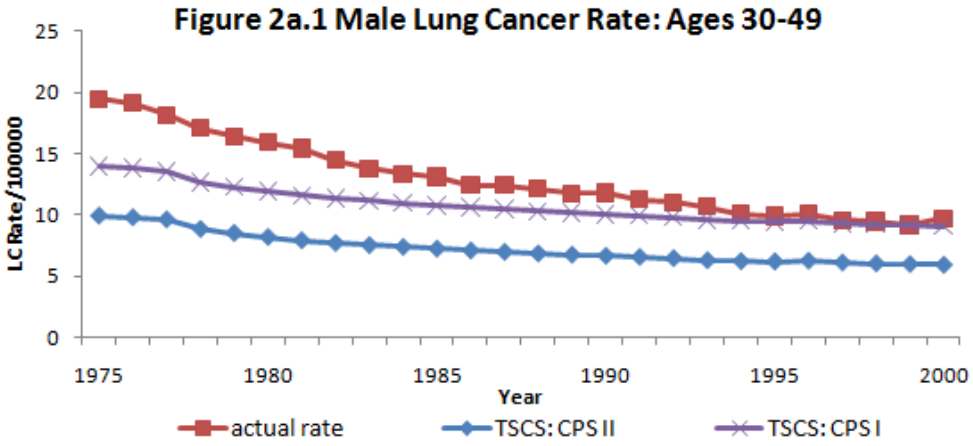
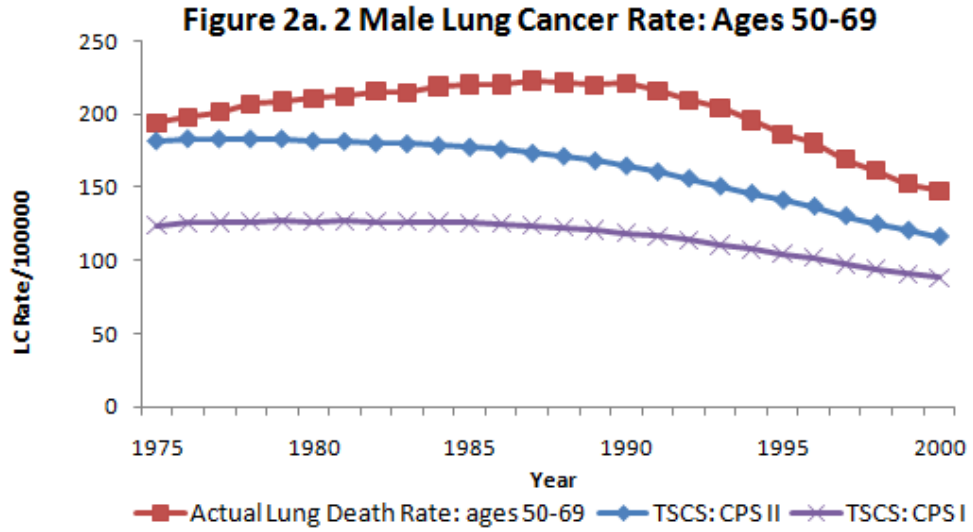


FIGURE2A





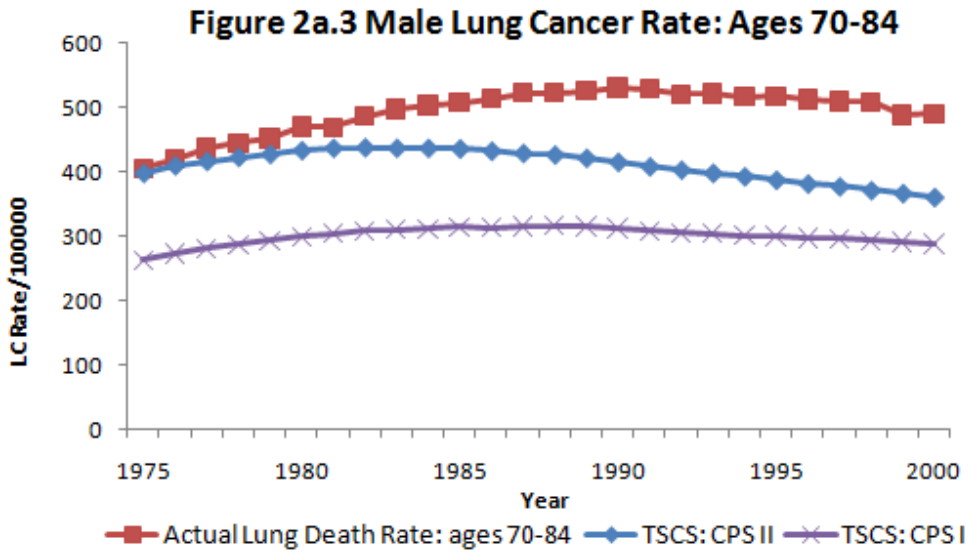
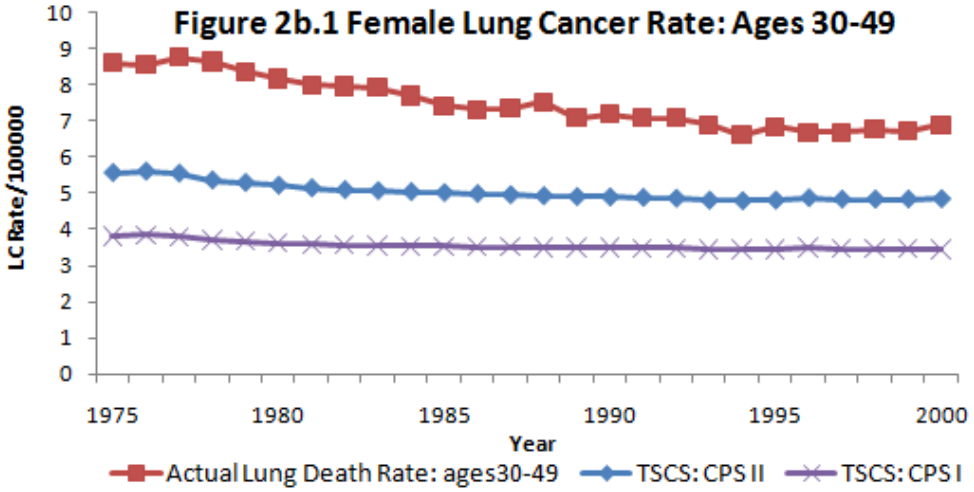
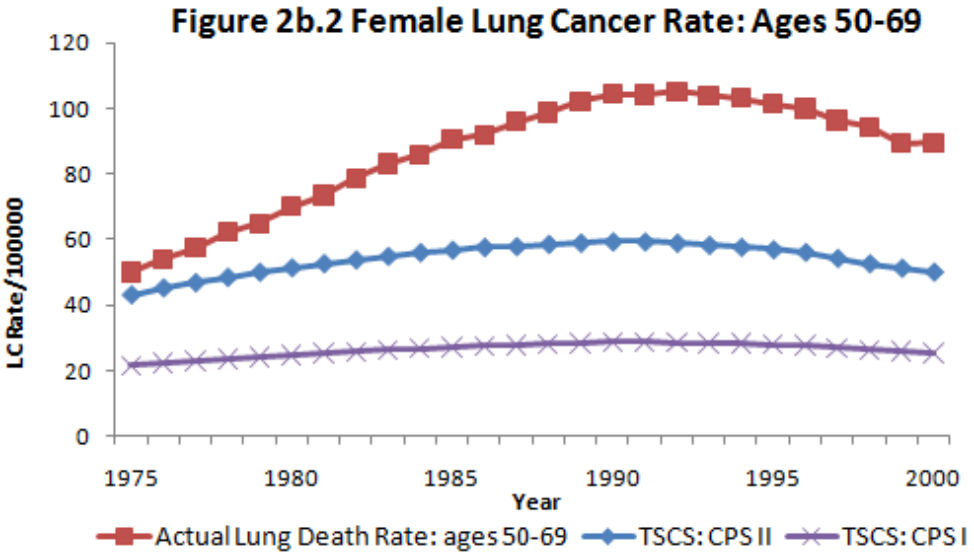


FIGURE2B





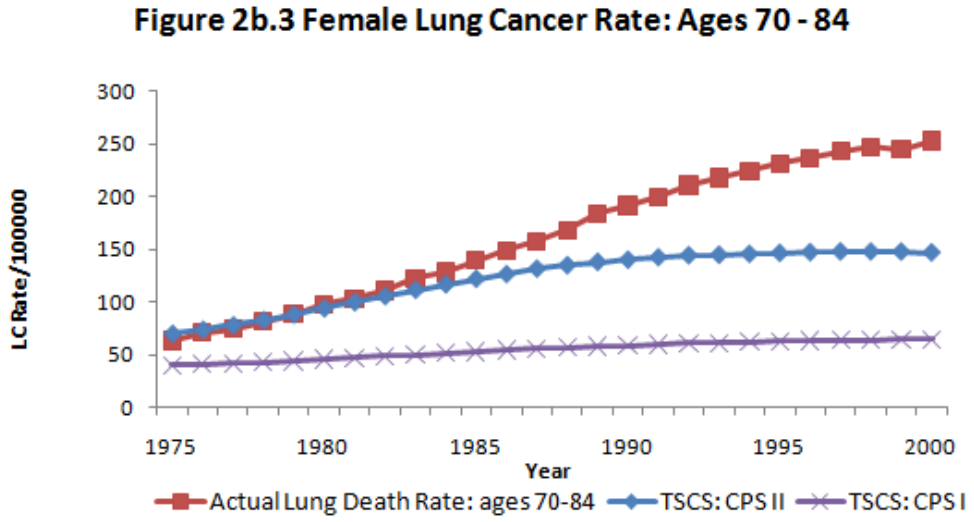
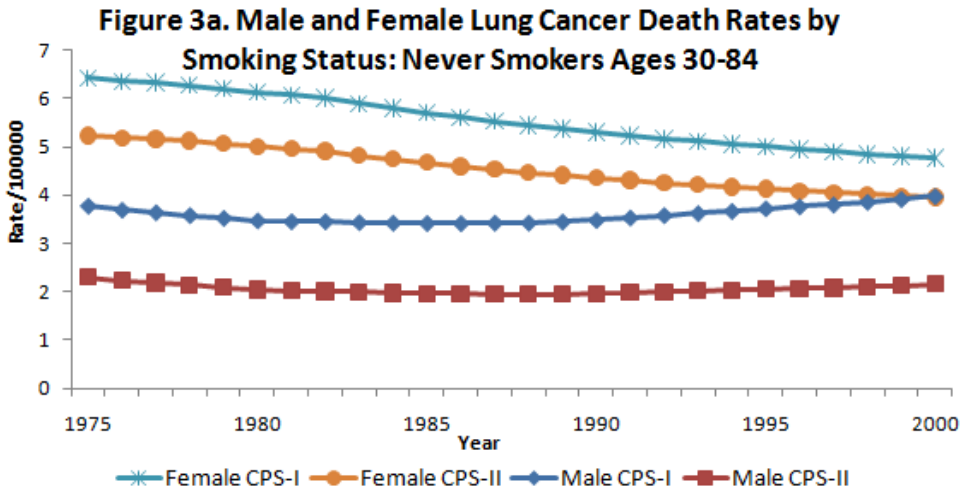
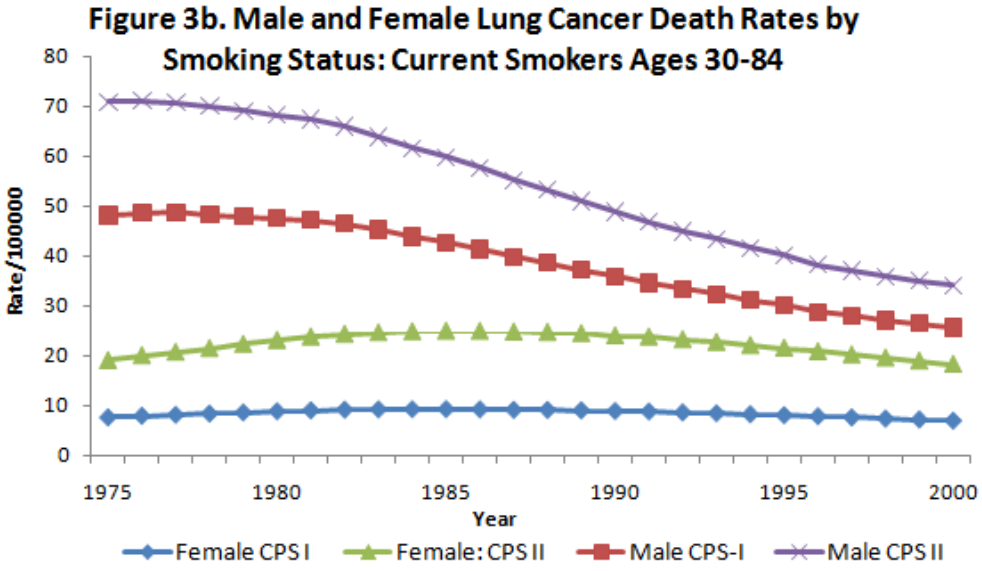


FIGURE3





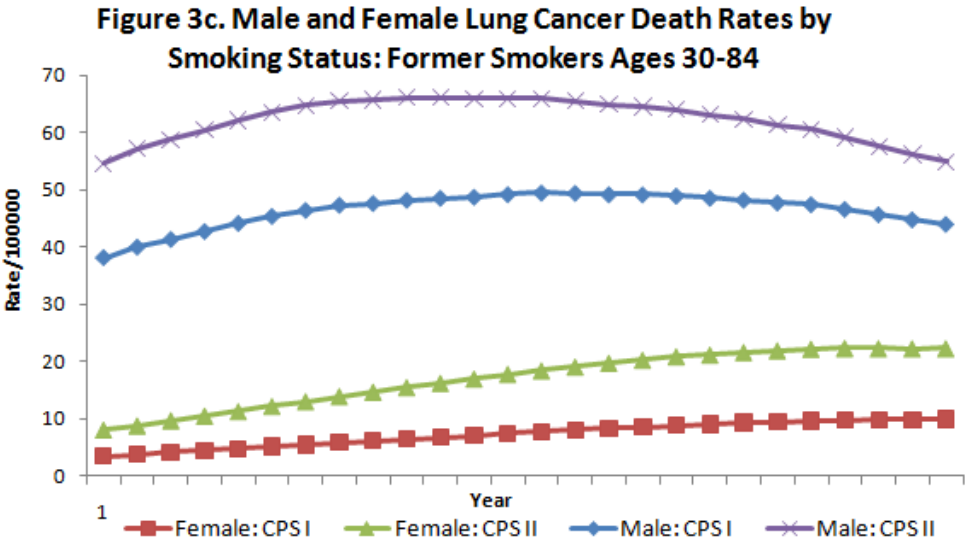


TABLE1A

Model	Constant	Linear time trend	Quadratic time trend	Predicted rate per 100K pop	predictT	predictTT	R Square	Adjusted R square	Std. Error of the Estimate	Durbin-Watson
				CPS-I: Linear Form						
1	29.5			1.28			0.830	0.822	4.095	0.13
t-stat	3.1			10.81						**
sig	0.0									
2	-68.5			2.31	0.016		0.952	0.948	2.218	0.43
t-stat	-5.0			15.49	7.67					**
sig										
3	121.1			0.07	0.032	-0.0016	0.992	0.991	0.902	1.70
t-stat	6.6			0.34	18.6	-10.8				
sig				0.74						
4	127.3				0.032	-0.0017	0.992	0.992	0.885	1.70
t-stat	221.1				27.5	-40.4				
sig										
5	131.3	-7.42			0.098		0.989	0.988	1.060	1.51
t-stat	215.1	-33.6			29.5					*
sig										
6	128.3	-1.93			0.049	-0.0012	0.993	0.992	0.876	1.89
t-stat	127.3	-1.19			3.40	-3.41				
sig		0.25			0.003	0.0030				
7	107.8	-2.83		0.25	0.056	-0.0009	0.993	0.992	0.875	1.98
t-stat	5.4	-1.54		0.24	0.016	0.0010				
sig		0.14		0.31	0.002	0.10				
				Natural Log Form						
8	1.41			0.79			0.844	0.837	0.03114	0.14
t-stat	4.63			11.38						**
sig										
9	-1.96			1.53	0.0022		0.949	0.945	0.018	0.41
t-stat	-3.80			13.47	6.96					**
sig										
10	4.85				0.0043	-0.00023	0.992	0.991	0.0070	1.78
t-stat	1086.2				25.1	-36.69				
sig										
11	3.36	-0.20		0.34	0.047		0.992	0.991	0.0072	1.92
t-stat	6.42	-11.04		2.90	11.6					
sig				0.01						
12	3.66		-0.00076	0.27	0.0038		0.993	0.992	0.0067	1.98
t-stat	7.25		-12.02	2.37	21.45					
sig				0.03						

TABLE1B

Model	Constant	Linear time trend	Quadratic time trend	Predicted rate per 100K pop	predictT	predictTT	R Square	Adjusted R square	Std. Error of the Estimate	Durbin-Watson
CPS-II: Linear Form										
1	65.84			0.62			0.741	0.730	5.051	0.11
t-stat	8.29			8.28						**
sig										
2	-46.98			1.44	0.019		0.947	0.943	2.326	0.41
t-stat	-3.78			15.43	9.49					**
sig										
3	146.95			-0.17	0.026	-0.0010	0.991	0.991	0.929	1.60
t-stat	8.06			-1.12	25.80	-11.05				**
sig				0.27						
4	126.5				0.025	-0.0010	0.991	0.991	0.934	1.50
t-stat	203.1				27.16	-39.95				*
sig										
5	130.1	-5.53			0.057		0.992	0.991	0.920	1.84
t-stat	234.7	-40.6			34.04					
sig										
6	128.3	-2.90			0.042	-0.00062	0.993	0.992	0.863	1.96
t-stat	128.2	-2.23			5.56	-2.04				
sig		0.04				0.05				
7	127.7	-2.93		0.0054	0.042	-0.00061	0.993	0.992	0.883	1.96
t-stat	6.30	-1.83		0.03	4.66	-1.30				
sig		0.08		0.98		0.21				
Natural Log Form										
8	2.48			0.51			0.774	0.765	0.03743	0.12
t-stat	9.43			9.07						**
sig										
9	-1.80			1.39	0.0036		0.945	0.939	0.019	0.39
t-stat	-3.42			12.88	8.42					**
sig	0.0023									
10	4.85				0.0041	-0.00021	0.993	0.992	0.0069	1.82
t-stat	1093.4				25.7	-37.5				
sig										
11	3.80	-0.17		0.22	0.04		0.993	0.991	0.0070	2.030
t-stat	7.63	-12.2		2.14	13.45					
sig				0.04						
12	3.77		-0.0008	0.225	0.0038		0.993	0.992	0.0067	1.98
t-stat	7.87		-12.6	2.26	24.9					
sig				0.033855						

TABLE2A

Model	(constant)	Linear time trend	Quadratic time trend	Predicted rate per 100K pop	predictT	predictTT	R Square	Adjusted R square	Std. Error of the Estimate	Durbin-Watson
					CPS-I: Linear Form					
1	-127.6			9.69			0.936	0.934	3.31	0.12
t-stat	-12.9			18.77						**
sig										
2	-54.7			5.32	0.04		0.994	0.993	10.6	0.60
t-stat	-9.3			15.56	14.61					**
sig										
3	105.7			-4.79	0.27	-0.01	0.998	0.998	0.573	2.05
t-stat	4.9			-3.51	8.79	-7.49				
sig				0.00						
4	29.6				0.16	-0.003	0.997	0.997	0.700	1.23
t-stat	72.0				44.84	-24.07				**
sig										
5	40.9	-7.04			0.42		0.937	0.932	3.353	0.14
t-stat	16.2	-1.80			2.22					**
sig		0.08			0.04					
6	31.3	-2.62			0.29	-0.003	0.998	0.998	0.543	2.19
t-stat	59.8	-4.04			9.31	-29.26				
sig		0.001								
7	57.1	-1.95		-1.65	0.29	0.00	0.998	0.998	0.548	2.26
t-stat	1.64	-1.74		-0.74	9.14	-3.07				
sig	0.12	0.10		0.47		0.01				
					Natural Log Form					
8	-2.37			2.29			0.996	0.996	0.015950	0.68
t-stat	-28.4			76.8						**
sig										
9	-1.70			2.03	0.0010		0.998	0.997	0.01245	1.11
t-stat	-12.12			38.14	5.29					**
sig										
10	3.44				0.02	-0.00053	0.999	0.999	0.0000787	2.43
t-stat	626.16				76.35	-46.21				
sig										
11	-3.00	-0.1113			2.340043057	0.04		0.999	0.999	0.00011253
t-stat	-16.92	-3.11		37.72	3.43					
sig										
12	1.66		-0.0010	0.644	0.019		0.999	0.999	0.00876	2.39
t-stat	1.70		-4.94	1.83	6.20					
sig	0.103			0.081						

TABLE2B

Model	(constant)	Linear time trend	Quadratic time trend	Predicted rate per 100K pop	predictT	predictTT	R Square	Adjusted R square	Std. Error of the Estimate	Durbin-Watson	
CPS-II: Linear Form											
1	-57.50			2.91			0.878	0.873	4.576	0.09	
t-stat	-6.52			13.16						**	
sig											
2	-7.50			1.31	0.02		0.996	0.995	0.866	0.82	
t-stat	-2.91			17.40	25.43					**	
sig	0.01										
3	48.56			-0.60	0.09	-0.0020	0.998	0.998	0.540	2.20	
t-stat	5.21			-1.90	8.14	-6.10					
sig				0.07							
4	30.82				0.07	-0.0014	0.998	0.998	0.570	1.82	
t-stat	96.76				53.8	-27.0					
sig											
5		40.54	-5.28			0.16		0.964	0.961	2.525	0.23
t-stat	30.43	-3.92			5.12					**	
sig		0.00									
6	31.42	-0.54			0.08	-0.0013	0.998	0.998	0.557	2.04	
t-stat	60.44	-1.44			10.91	-21.24					
sig		0.16									
7	60.34	0.53		-1.03	0.10	-0.0024	0.998	0.998	0.548	2.24	
t-stat	2.74	0.59		-1.31	6.99	-2.82				**	
sig	0.01	0.56		0.20		0.010					
Natural Log Form											
8	-3.71			2.11			0.920	0.916	0.0711	0.084	
t-stat	-7.94			16.57						**	
sig											
9	-0.73			1.24	0.0041		0.998	0.997	0.0123	1.12	
t-stat	-5.46			32.5	27.8					**	
sig											
10	3.44				0.019	-0.00041	0.998	0.998	0.00919	2.21	
t-stat	612.0				73.4	-43.8					
sig											
11	-0.52	-0.07		1.18	0.02		0.998	0.998	0.0104	1.71	
t-stat	-3.93	-3.22		32.1	3.91						
sig		0.00									
12	1.91		0.00	0.45	0.01		0.999	0.999	0.00896	2.30	
t-stat	3.31		-4.64	2.65	6.65						
sig	0.003			0.01							

FLEXKB DOCUMENT
Version: HI.001.03302011.70017
Document generated: 03/30/2011

MGH Institute for Technology Assessment

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.

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

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

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
- Population Component
- Incidence Component
- Natural History Component
- Screening Component
- Treatment 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.

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.

Advanced Reading

These topics denote more detailed documentation about specific and important aspects of the model structure

MODEL PURPOSE

SUMMARY

This document provides a brief overview of two versions of the Lung Cancer Policy Model (LCPM) as of the time this model profile was archived. The Summary Of Versions table lists differences between the single cohort and dynamic cohort versions of the model and provides examples of their uses to date. The Model Overview gives more detail and links to model components.

PURPOSE

The original single-cohort LCPM was designed to evaluate the effectiveness, costs, and cost-effectiveness of helical computed tomography (CT) screening for lung cancer in the U.S. The single-cohort model can also be used to evaluate both future screening technologies and advances in treatment effectiveness.

The LCPM was designed to reproduce observed lung cancer incidence and survival rates in a specified cohort, in the absence of screening. A screening component allows comparison of mortality rates in the same cohorts under multiple scenarios, e.g., no screening versus screening. Individual-level outputs include the probability of positive screening tests. A notable limitation of the current model is that individuals are simulated as receiving care consistent with clinical practice guidelines.

A dynamic-cohort Population LCPM was developed to evaluate U.S. population trends in incidence and mortality.

MODEL OVERVIEW

SUMMARY

This document provides an overview of the lung cancer simulation model developed by researchers at the MGH Institute for Technology Assessment/Harvard Medical School.

PURPOSE

The Lung Cancer Policy Model (LCPM) was originally designed to evaluate screening programs in a specified cohort. When we originally joined CISNET (as an Affiliate group), the LCPM did not simulate populations. The Population LCPM was developed with CISNET funding and was used to participate in the Smoking Base Case. See Summary Of Versions for an overview of the differences between the versions.

Designed to evaluate the effectiveness, costs, and cost-effectiveness of helical computed tomography (CT) screening for lung cancer in the U.S., the LCPM will inform screening decisions prior to completion of ongoing trials, address limitations of published cost-effectiveness analyses of lung cancer screening and offer an opportunity to evaluate both future screening technologies and advances in treatment effectiveness.

BACKGROUND

An effective means of reducing mortality from lung cancer, the leading cause of cancer death in the U.S., is urgently needed. Unfortunately, even a sharp reduction in current smoking rates -- an obvious first step -- would not eliminate lung cancer in the near term: a former smoker's risk for lung cancer remains elevated for decades after smoking cessation. To date, no screening program has been demonstrated effective at reducing lung cancer mortality.

Ongoing trials of helical CT screening will contribute critical information on effectiveness, but debates over past cancer screening trials (e.g.,[1]) should remind us that publication of completed trial results is unlikely to eliminate uncertainty about the effectiveness of lung cancer screening.

Advances in screening technologies, staging examinations, and therapies are being made simultaneously, yet conducting controlled trials on all of these aspects at once is simply not feasible. The comprehensive modeling approach used in the LCPM, however, permits an evaluation of all three inter-related areas. Specifically, modeling can be used to: 1) estimate effects of several combined screening, workup, and treatment strategies; 2) interpret and reconcile the results of different screening trials; 3) evaluate the potential effects of improved adherence to staging and treatment guidelines; and 4) determine the effect that improvements in staging and treatment might have on screening effectiveness. Finally, by including costs as well as effectiveness outcomes, our model will provide information concerning the relative cost-effectiveness of interventions spanning the spectrum from screening to treatment, and thereby provide information which is useful to physicians, policy makers, legislators and the public.

Model Description

The LCPM is a state-transition model, analyzed as Monte Carlo to allow for individual heterogeneity in risk factors and event rates. Individuals can move through 5 possible states: general population, follow-up, diagnosis & staging, treatment & survival, and dead. Please see the Component Overview and links provided for further details.

The model employs a lifetime time horizon and a cycle length of one month to capture the short survival times of late-stage lung cancers and to allow for a wide variety of event recurrence frequencies. The model was populated with individuals in an age-, race-, gender-, and calendar year-specific cohort representative of the U.S. in terms of smoking history (Population Component).

Inputs include national survey data for assigning smoking histories, type-specific distributions of doubling times for lung cancers (Natural History Component), rates of thoracic imaging exams performed for reasons unrelated to lung cancer, and response rates of treatments.

Outputs include estimation of incident cancers (Incidence Component), stratified by age, type, and stage, as well as mortality by detected stage and treatment (Survival Mortality Component). Calibration to observed incidence and stage-specific survival curves from the NCI SEER tumor registry allowed estimation of parameters governing unobservable events, such as development of the first cancerous cell and of metastasis. Some endpoints from CT screening trials and other literature sources describing clinical experience were used as secondary calibration targets. Validation of the model was performed by reproducing observed results of a past lung cancer screening trial and cohort studies. See Calibration Validation Results for a summary of model calibration and validation.

As with any model, simplifying assumptions were made (Assumption Overview). Increasing complexity of the model must be balanced against the number of parameters that can be estimated using available data; calibration and validation can show that model outputs are consistent with observed data, but do not guarantee that the model accurately represents the underlying biology. The model currently omits radon and second-hand smoke exposure, two known risk factors for lung cancer.

Contributors

Pamela M. Mc Mahon, PhD
Chung Yin Kong, PhD
G. Scott Gazelle, MD, MPH, PhD

References:

[1] Gøtzsche, P. C., Nielsen, M. “Screening for breast cancer with mammography” in Cochrane Database of Systematic Reviews 2006; 4

ASSUMPTION OVERVIEW

SUMMARY

This document discusses key assumptions underlying the Lung Cancer Policy Model (LCPM) and their possible implications. See the linked Component documents for context and further details.

BACKGROUND

The LCPM is a comprehensive model of lung cancer development, detection, treatment, and survival. Unlike a stage-shift model, the LCPM's underlying natural history model does not require estimates of Screening Biases (eg lead-time) as model inputs. To estimate parameters governing unobservable events (e.g., timing of metastasis), we calibrated to multiple endpoints in observed data (see Calibration Details). Using the calibrated model, we can simulate a screening program and generate estimates of the screening biases as model outputs.

Archived versions of this document will remain as technical appendices for publications, but newer versions of this document will reflect updates and refinements. Like the Coronary Heart Disease Policy Model developed by Dr. Milton Weinstein (a Co-Investigator on the LCPM) and colleagues[1,2], the LCPM was designed to be a model with a long lifespan.

The LCPM does not rely on data from a single trial to inform the parameter estimates (but rather incorporates trial data as they emerge), so can be used to evaluate screening in populations not included in ongoing trials, and can address the 'moving target' problem of improved test sensitivity (e.g., CT resolution), as well as other late-breaking topics, such as treatment interventions.

ASSUMPTION LISTING

Population Component Assumptions

- To allow for undetected lung cancers in the cohort, each individual is first 'regressed' to age 20 and assumed to be free of lung cancer. Upon entering the general population state, he can develop one or more lung cancers as he ages and acquires his (known[3,4]) smoking exposure. Two procedures insure that the cohort still reflects the U.S. population upon reaching the cohort age: 1) individuals face no competing risks of death until reaching the cohort age; and 2) any individual who dies of lung cancer prior to reaching the cohort age is re-started at age 20 with the same smoking history. Aggressive cancers that would have been fatal at ages younger than the cohort age are thus appropriately removed.
- In the single cohort LCPM, all current smokers after 1990 face a 3% annual chance of quitting, based on estimates of 2.5% to 3.2%.[5,6] Cessation rates in the Population LCPM rely on the Smoking History Generator.

- Smoking histories in the single cohort LCPM did not incorporate the tendency of beginning smokers to gradually increase the number of cigarettes smoked per day. The Smoking History Generator used in the Population LCPM more closely approximates such smoking behavior influences the apportioning of lung cancer risk across the population (see Natural History, below) and alters the proportion of individuals in a cohort eligible for screening.

Natural History Component Assumptions

- The risk of developing a lung cancer is modeled using a tolerance model: increasing age, smoking exposure, and genetic susceptibility contribute to risks of developing one of 5 histologic types of lung cancer.
- A person can develop a maximum of 3 cancers in a lifetime, of any of the 5 types.
- The growth rate assigned to each cancer is drawn from a distribution specific to the histologic type, is assumed to decrease with increasing size, and was allowed to vary by smoking status during model calibration.
- Disease progression is modeled through monthly probabilities of involvement of lymph nodes and development of distant metastases. Progression risks are functions of characteristics of existing cancers (location, volume, doubling time, and type), nodal status, and random individual variation.
- Incidence of benign lesions varies with age and geographical region but not with smoking history. Few benign lesions exhibit cancer-like growth.
- The proportion of mixed adenocarcinoma/BAC that is pure BAC (bronchioloalveolar carcinoma) was estimated via calibration, not taken from the literature.

Incidence Component Assumptions

- Benign lesions and asymptomatic lung cancers can be detected incidentally during a thoracic imaging exam performed for an unrelated reason (non-screening). Risks of incidental imaging are functions of age, gender, and geography. Sensitivity varies with size and location and was estimated during calibration.
- Symptom detection can occur via symptoms from the largest primary cancer, by distant metastases, or both.
- Incidence rates reported in SEER reflect a negligible rate of lung cancer screening in the population.

Follow Up Component and Workup And Staging Component Assumptions

- Lesions suspicious for lung cancer (from symptoms or incidental detection) are biopsied if over a minimum diameter or followed with serial high-resolution CT exams (even in the absence of screening).
- Lesions that exhibit no detectable growth after 2 years of follow-up are assumed to be benign and to require no further surveillance. Reflecting clinical practice, a proportion of benign lesions are diagnosed as benign on the basis of a high-resolution CT (a proxy for modeling calcification patterns).

- Biopsy-confirmed malignancies are clinically staged based on guidelines recommended by the National Comprehensive Cancer Network (NCCN).

Treatment Component Assumptions

- Treatments are assigned following NCCN guidelines.
- Effectiveness of systemic treatments are based on probabilities of complete or partial response. See below for relationship of treatment effectiveness to survival.
- Effectiveness of resection depends on the existence of undetected second lung cancers and/or occult metastases.

Survival Mortality Component Assumptions

- Survival is a function of both underlying disease state and treatment received (which itself depends on the accuracy of staging). Patients with M1 (stage IV or ES) cancers are assigned exponential survival, based on observed median survival rates[7]. (Observed stage-specific survival rates[7] for patients with M0 cancers are used as calibration targets, not inputs.)
- Once a patient is diagnosed as stage IV, survival is as observed in SEER (by age, decade, race, gender, and cell type).

REFERENCES:

[1] Weinstein, MC, Coxson, PG, Williams, LW, Pass, TM, Stason, WB, Goldman, L "Forecasting coronary heart disease incidence, mortality, and cost: The Coronary Heart Disease Policy Model" in American Journal of Public Health 1987; 77: : 1417-1426

[2] Hunink, MGM, Goldman, L, Tosteson, ANA, Mittleman, MA, Goldman, PA, Williams, LW, Tsevat, J, Weinstein, MC "The recent decline in mortality from coronary heart disease, 1980-1990" in JAMA 1997; 277: 7: 535-542

[3] Massey, JT, Moore, TF, Parsons, VL, Tadros, W "Design and Estimation for the National Health Interview Survey, 1985-94" in Vital Health Statistics, Series 2, No. 110 1989;

[4] U.S. Department of Health and Human Services, National Center for Health Statistics, "The National Health and Nutrition Examination Survey III Data file 1988-1994." in Public Use Data file Series 11 1997;

[5] Centers for Disease Control "Smoking Cessation During Previous Year Among Adults -- United States, 1990 and 1991" in Morbidity and Mortality Weekly Report 1993; 42: 26: 504-507

[6] The Commit Research Group "Community Intervention Trial for Smoking Cessation (COMMIT): II. Changes in Adult Cigarette Smoking Prevalence" in Am J Public Health 1995; 85: : 193-200

[7] Surveillance Research Program, National Cancer Institute, "Surveillance, Epidemiology, and End Results (SEER) Program Seer*Stat Database: Incidence - SEER 13 Regs Public-Use, Nov 2004 Sub (1973-2002 varying)" in Seer*Stat software version 6.1.4 2005;

PARAMETER OVERVIEW

SUMMARY

This document provides an overview of the major classes of parameters in the LCPM, and provides links to parameter documents.

BACKGROUND

Most natural history parameters were estimated by calibration. The NCI's SEER registry was the primary data source for calibration targets. In the absence of screening, the model should accurately predict observed tumor registry (SEER) incidence by year, age, sex, and race. Characteristics of incident cancers predicted by the model should also correspond to observed distributions of cell types, stages, and sizes, and stage-specific survival rates for M0 cancers should be accurately predicted. Additional calibration targets were defined from the literature (see Calibration Details and Output Overview).

PARAMETER LISTING OVERVIEW

1. Unobservable parameters define unobservable events: development of lung cancer, disease progression, and symptom detection. Note that the actual values of many of these parameters are not meaningful outside the context of the LCPM (although their relative magnitudes may reveal insights into biology). Of more interest are outputs of the model, such as estimates of Screening Biases.

See Parameters Natural History and Natural History Component for details.

2. Uncertain parameters were those for which literature estimates provided ranges of values. Categories of uncertain parameters included test characteristics, operative mortality rates for interventions, response rates for systemic therapies, and probabilities of clinical events such as wedge biopsy of a growing pulmonary lesion.

See Parameters Test Performance and Parameters Treatment for details.

3. Structural parameters were fixed during calibration, but included for future analyses. These included a parameter to allow simulation of African-American cohorts. Additional structural parameters are described in their relevant model components.

The Assumption Overview describes major assumptions underlying the LCPM.

4. Other parameters include estimates of costs and weights for adjustments in quality of life due to lung cancer diagnosis and treatment (to allow estimation of cost-effectiveness ratios; see Results Overview).

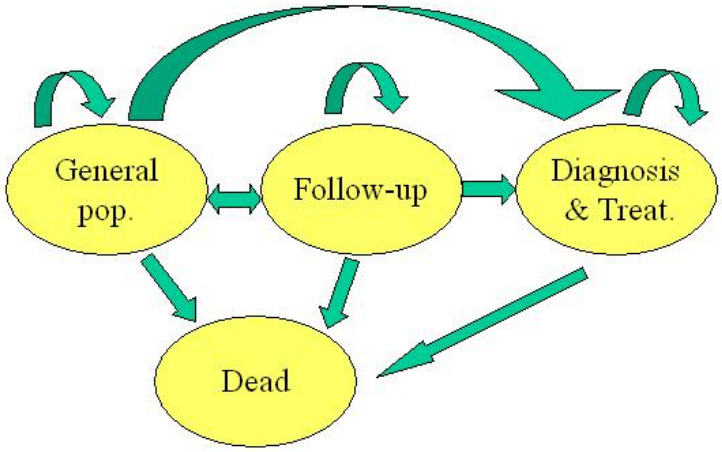
COMPONENT OVERVIEW

SUMMARY

This document describes typical sequences of component processes for a hypothetical individual simulated by the LCPM.

OVERVIEW

Persons start the model in the general population state.
See schematic.



Each month, persons in the general population state face competing risks of death from causes other than lung cancer.[1] While in the general population state, benign pulmonary nodules and lung cancers can develop, and lung cancers can grow, progress to nodal involvement and/or distant metastases, or cause symptoms. Additionally, persons face risks of undergoing non-screening thoracic imaging exams for reasons unrelated to lung cancer (e.g., for trauma). In a screening scenario, persons can undergo screening if they are eligible for the specific program and adherent to the screening protocol.

Persons with small incidentally detected lesions undergo sequential imaging exams in the follow up state.

Larger incidentally detected lesions, lesions exhibiting growth on serial imaging exams, and symptomatic cancers are sent for work up and staging.

Once the diagnosis of lung cancer is made, the cancer is staged, and the person moves to the treatment and survival state.

In the next section, we provide available links to component processes for each of the states indicated above.

Component Listing

General population

The Natural History Component is included, as well as the Screening Component and the Incidental Imaging Component.

Follow-Up

In the Follow Up Component, incidentally-detected nodules smaller than the cutoff threshold are managed expectantly with periodic high-resolution CT exams. While a patient is being followed up, he also cycles through the Natural History Component.

Work-up and Staging

In a single cycle (one month), workup and staging tests are used to establish both the presence of lung cancer as well as the extent of disease progression. See the Workup And Staging Component. Patients also cycle through the Natural History Component.

Treatment and Survival

In addition to the Treatment Component and the Survival Mortality Component, the Natural History Component is also included here. This allows for development of second lung cancers as well as disease progression of existing primary cancers or occult metastases.

See the Assumption Overview for key assumptions and links to parameter documents from the corresponding component documents.

References:

[1] McMahon, P. M., Zaslavsky, A. M., et al. “Estimation of mortality rates for disease simulation models using Bayesian evidence synthesis” in Medical Decision Making 2006; 26: : 497-511

SMOKING HISTORY GENERATOR COMPONENT

SUMMARY

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 lung cancer consortium.

OVERVIEW

The core SHG software was parameterized using three tobacco control scenarios to 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-

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**

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**	**CTC**
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 (1890-1984)	User Defined		
Gender (Male/Female)	User Defined		
Race (All race)	User Defined		

[1] Cigarettes smoked per day,[2]Other 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

Parameter	Valid Values
Seed value for PRNG used for Initiation, Cessation, OCD[1], Smoking intensity quintile	Integer from -1 to 2147483647 (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, Smoking intensity quintile and CPD[4] data files	As derived from NHIS depending on the scenario

[1]Other 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. [3]Key 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. [4]Cigarettes 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 History	Smoking intensity quintile (5 quintiles ranging from light to heavy smoking), Yearly smoking dose (CPD[2])

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

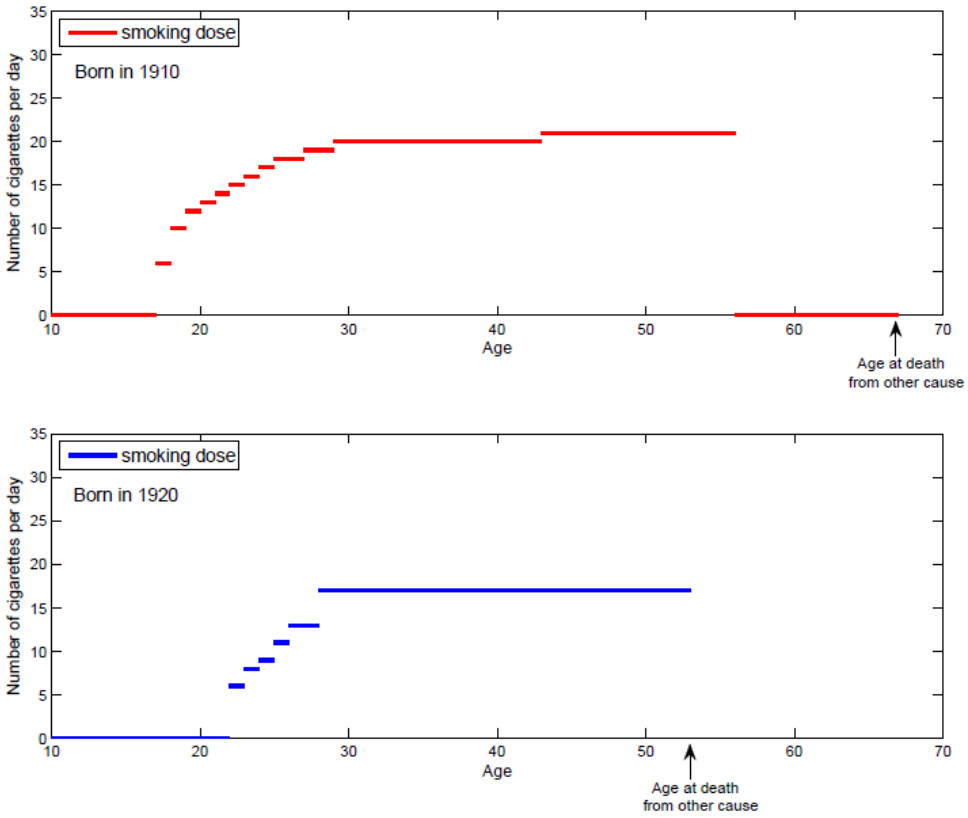


Figure SHG-1: Examples of the SHG-Generated Events

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

POPULATION COMPONENT

SUMMARY

This document describes both the original single cohort LCPM and the Population LCPM. See Summary Of Versions for an overview of how the versions compare to each other.

OVERVIEW

The population component defines the initial characteristics of the population entering the LCPM for a given simulation. This is the first component in the modeling process. Once each individual in the population is initiated, the individual moves to the general population state (see Component Overview).

Initial characteristics assigned to each hypothetical person include gender, race, ethnicity, age, and smoking history. Smoking history includes current status, age at smoking initiation (if applicable), age at smoking cessation (if applicable), and cigarettes per day. In the single-cohort LCPM, the cigarettes smoked per day is assumed constant for cycles in which the individual is a current smoker. In the Population LCPM, cigarettes per day could vary over time.

Additional characteristics include an indicator for genetic susceptibility to lung cancer (see Natural History Component).

Simulations begin in a specified calendar year, so that:
1) the proportions of ethnicities and the prevalence of smoking is representative of the cohort being simulated, and
2) the model-predicted incidence rates can be compared to the corresponding SEER data (see Incidence Component).

QUANTITATIVE DESCRIPTION

The LCPM is a state transition (Markov) model, analyzed as Monte Carlo (i.e. it simulates life histories of individuals). The possible states are described in the Component Overview. The model uses a lifetime time horizon and a cycle length of one month to capture the short survival times of late-stage lung cancers and to allow for a wide variety of event recurrence frequencies.

POPULATION DYNAMICS

The original version of the LCPM is a single-cohort model. Individuals enter the model in specified calendar years, however, and carry appropriate smoking histories (informed by national survey data).

The Population LCPM simulates multiple birth cohorts to generate annual incidence and mortality rates.

Recurrence

The LCPM does model recurrence, via either clinical detection of (previously) occult metastases or by development of a new primary lung cancer. Each individual in the LCPM can develop up to 3 lung cancers (of any of 5 cell types), and up to 3 benign lesions. See the Natural History Component and the Survival Mortality Component.

Disease Distribution

We model the risks of developing each of the 5 cell types of lung cancer (adenocarcinoma/BAC, large cell, small cell, and squamous, as well as other) as independent, conditional on risk factors. (In other words, we do not assign a distribution of lung cancer histologies to the population.)

Detail

Approximately 10% of lung cancers occur in life-long non-smokers and SEER data are not stratified by smoking history. Therefore, the LCPM is populated with entire age, race and gender cohorts, representative of the U.S. in terms of smoking history. Smoking history includes current status, age at smoking initiation (if applicable), age at smoking cessation (if applicable), and cigarettes per day.

An indicator for genetic susceptibility to lung cancer (see Natural History Component) is assigned randomly.

Single cohort LCPM

The single-cohort LCPM simulates cohorts of white males and females aged 50, 60 or 70 in 1990. Cohorts entered the model in calendar year 1990 for calibration to SEER data from 1990 to 2000. Joint distributions of ethnicity and geographic region of the U.S. were derived from the 1990 Census. Ethnicity (Hispanic/non), region of the country, and smoking history were assigned to each individual.

Using the 1990 National Health Interview Survey we fit a multinomial logistic model to estimate the proportion of each smoking status using the predictors age, sex, race, ethnicity, and region. Data from the third National Health and Nutrition Examination Survey fielded in 1988-1994, were used to estimate normal distributions of ages of starting and stopping smoking and the average number of cigarettes smoked per day, conditional on smoking status, age group, and ethnicity. Cigarettes per day was assumed constant for cycles in which the individual is a current smoker.

Trial populations can also be simulated - see Screening Component.

Population LCPM

The Population LCPM uses the Smoking History Generator common to all CISNET lung groups. The Smoking History Generator allows for beginning smokers to 'ramp up' the number of cigarettes per day and yields a wider range of accumulated pack-years than the smoking histories used for the single cohort model described above.

Relevant Assumptions

See the Assumption Overview.

Relevant Parameters

Inputs

As described above, smoking histories for the single cohort LCPM were derived from the NHIS and NHANES, large sample surveys that yielded precise estimates of cigarettes per day and ages of starting and stopping smoking. For the same 6 cohorts in 1990, the Smoking History Generator yields a wider range of pack-years.

Impacts of Smoking History Inputs

Smoking is the strongest risk factor for lung cancer, so even small variations in smoking histories will influence lung cancer outcomes. To assess the downstream effects of the observed differences between the smoking histories from the original single cohort LCPM and those from the Smoking History Generator, we used the Smoking History Generator to provide ages of starting/stopping smoking and cigarettes per day and re-calibrated the model, allowing effects of smoke-years, cigarettes per day, and age to vary from their original estimates. We identified a parameter set that apportioned the lung cancer risk across smokers differently. See Parameters Natural History for further details.

Relevant Components

The Population Component is necessary to specify the characteristics of the cohort entering the LCPM. Different cohorts will have different lung cancer risks and therefore outcomes.

Dependent Outputs

All outputs will be affected by the characteristics of the population being simulated. Heavier smokers, for instance, will have higher rates of lung cancer death and possibly poorer outcomes from treatment. Characteristics (e.g., doubling times, sizes) of detected lung cancers will also vary across input populations. See the Output Overview document.

Relevant Results

See the Results Overview for a summary of relevant results from the single cohort LCPM and Population LCPM.

INCIDENCE COMPONENT

SUMMARY

To be counted as an incident case, a lung cancer must first develop, then be detected (by any of several possible modalities) and finally be diagnosed.

This document describes how the LCPM counts incident lung cancers and provides links to descriptions of components that involve development of cancer, detection, and diagnosis.

OVERVIEW

In the LCPM, we model the development of lung cancers, followed by tumor growth and metastasis (see Natural History Component). An individual with undetected lung cancer remains in the general population state (see Component Overview).

After the last individual in a cohort is simulated, we essentially count up the numbers of cancers in various categories. We count as incident cancers only cancers that were diagnosed during the patient's lifetime. Incident cancers are further categorized by stage, size, type, etc.

Non-screening scenarios:

Lung cancer can be diagnosed symptomatically (either the primary cancer obstructing an airway or from distant metastases) or asymptomatically (found incidentally during a thoracic imaging exam performed for unrelated causes - see Relevant Components, below).

Age-specific incidence rates are calculated and then compared to observed data.

Screening scenarios:

Cancers may also be detected by screening (see Screening Component). The model tracks the mode of detection of each cancer.

DISEASE RISK

For each of the 5 lung cancer cell types, we estimate a logistic function to predict monthly risks of developing a cancer. For each cancer type, we estimated independent coefficients for age, age squared, cigarettes per day, years of smoking, and years since quitting. There is also a randomly-assigned indicator for increased genetic risk (equivalent to HR=2). See Natural History Component.

To account for observed birth cohort trends in lung cancer risks and allow for differences in baseline risk by gender, we added a term that modifies the monthly risk of lung cancer development (all cell types), stratified by gender. See Calibration Details.

Impact of Screening

As described above, we do distinguish between incidence in the absence of screening vs. the presence.
We calibrate to SEER (no screening) and validate with screening trial data (with screening).

Detail

Development of cancer

Further details on the way the LCPM simulates development of cancer is provided in the Natural History Component.

Detection of cancer

Three modes of detection are possible in the LCPM:

1. Symptoms of previously undiagnosed lung cancers (either the primary cancer or distant metastases) can prompt detection. See the Symptom Detection Component.

2. During each cycle spent in the general population, persons may undergo imaging exams (thoracic CT, or CXR) performed for reasons unrelated to screening for lung cancer. See the Incidental Imaging Component.

3. Screen detection can occur in eligible individuals, in scenarios which include screening. See Screening Component.

Diagnosis of cancer

In the LCPM, a diagnosis of lung cancer is required before a person transitions into the Treatment Component. Diagnosis is operationalized by a biopsy that returns a specific diagnosis of lung cancer. Biopsies and staging both occur in the one-month Workup And Staging Component.

Relevant Assumptions

For the single-cohort LCPM, we used national survey data (NHIS, NHANES) to assign smoking histories to the individuals in the cohort (see Population Component), and calibrate to SEER data for incidence.

- If the SEER registries are not representative of the US, calibration to SEER data may yield biased parameter estimates.
- The smoking histories used for the single-cohort LCPM do not reflect the tendency of individuals to increase their smoking intake over time (i.e., light smokers become heavy smokers), which overestimates the pack-years accrued. This could have resulted in biased estimates of the cumulative dose-response relationship between smoking and lung cancer risk.

Also see the Assumption Overview.

Relevant Parameters

This component relies most directly on the natural history parameters (see Natural History Component and the Symptom Detection Component).

However, parameters in other components can influence the incidence rates, such as patterns of imaging examinations (and their test characteristics) in the general population (see the Incidental Imaging Component) and whether screening is occurring (see the Screening Component).

Relevant Components

The incidence component operates after the last individual in a simulated cohort 'dies.' It does not contain any other components, per se, but merely functions as a bookkeeping component.

Components that influence the predicted incidence rates include the Natural History Component, as well as the Follow-up, Workup And Staging Component and the Incidental Imaging Component.

Dependent Outputs

Most outputs of interest will depend on the Incidence component, including incidence rates and therefore mortality rates.

Relevant Results

See Calibration Validation Results for a description of outputs from the LCPM after calibration and validation and links to specific outputs.

Natural History Component

Summary

This document describes various aspects of the model processes responsible for generating the natural history of lung cancer. Benign pulmonary nodules are described in the Benign Component.

Overview

The natural history component occurs in every cycle of the model, so that new lung cancers may develop (and existing lung cancers grow and progress) throughout life.

The Population Component initiates the population entering the LCPM and therefore precedes the natural history component. The natural history component has sub-components for lung cancer development, disease (tumor) growth, disease progression, and symptom detection. All of the sub-components are described below and/or in linked documents.

Approximately 6% of patients with lung cancer develop more than one primary tumor, and only half of synchronous multiple primaries are the same type. Therefore, we model up to three cancers per person, of any of the 4 main types of lung cancer, plus a 5th type to represent Carcinoma, Not Otherwise Specified (ICD-O-2 code 80103). We modeled pure bronchioloalveolar carcinoma (BAC) as a subset of adenocarcinoma +/- some BAC, reflecting their differences yet typically mixed histology and misclassification.[3]

Disease Stages

A 'true' disease stage is assigned based on the individual's simulated disease characteristics (tumor size, nodal involvement, distant spread). This true stage is updated every cycle. See Details, below. An observed disease stage is also assigned, based on the individual's 'true' disease characteristics and the results from any diagnostic or staging tests performed. Observed and true stages may not match if a cancer is undiagnosed or mis-staged by a test result.

Disease growth

We assume continuous Gompertz tumor growth, assigning a growth parameter for each new cancer that is drawn from distributions specific for the 5 cell types of lung cancer. We also include a term to allow cancers in smokers to exhibit accelerated growth. See Details, below and Parameters Natural History.

Stage Transition Trends

No temporal trends are imposed on stage transitions.

Disease evolution

One birth cohort parameter is changed over calendar time:

To account for observed birth cohort trends in lung cancer risks and allow for differences in baseline risk by gender,[6] we added a term that modifies the monthly risk of lung cancer development (all cell types), stratified by decade of birth and gender. See Calibration Details.

Remaining natural history parameters are not changed over (calendar) time. (Smoking histories do change over time, however, so will influence lung cancer trends.)

Regression

The model assumes an irreversible (in the absence of resection) progression of lung cancer disease stages. The speed of progression varies greatly, however, so that some cancers would never be detected during life in the absence of screening. The growth of BACs was truncated at a maximum diameter of 1 cm (detectable by X-ray).

Detail

Lung Cancer Development

The LCPM employs a simple 'tolerance' model of cancer development (so-called because cancer may only develop after an individual's tolerance to risk factors has been exceeded).

The monthly probability of developing the first malignant cell of cancer type k = 1-5 is a logistic function with a type-specific intercept and type-specific coefficients for age, age^2, years of cigarette exposure (smoke-years, SY), average number of cigarettes smoked per day (cigarettes per day, CPD), and the years since quitting (YSQ) smoking, if applicable. We also allow for random individual variation (highrisk, a proxy for genetic susceptibility), constant for all 5 types.

A logistic model produced nearly as high an R^2 goodness-of-fit statistic as a two-stage model[7] (R^2 of 0.61 and 0.67, respectively) in a comparison of 5 models for lung cancer's dose-response to tobacco,[8] and studies of case control data showed good fit using a logistic function to predict lung cancer (all types combined).[9] The MVK 2-stage model[7] models each initiated cell as growing instantaneously into a malignant tumor after a fixed period of time,[10] an assumption that precludes size-dependent sensitivity of imaging exams.

Lung Cancer Characteristics and Growth

Indicators are assigned to each new cancer for cell type, size (initial diameter of 0.01 mm), lobe in the lung, and central or peripheral location (varied by type).

In each cycle, the diameter and volume of existing cancers (and any growing benign lesions) are incremented according to a Gompertz function for tumor growth. Consistent with biological mechanisms of tumor growth (e.g., angiogenesis and necrosis of the tumor core), tumor doubling times decrease as the volume asymptotes to its maximum possible.

Mean doubling times for large, small, and squamous cell cancers were estimated from the literature (see Table Growth Parameters) and used to derive distributions of growth rate parameters. Distributions of growth rate parameters for adenocarcinoma/ BAC and "other" cell types were estimated via calibration.

A modification term (estimated during calibration) allows slower growth rates in non-smokers.

Lung Cancer Progression

Disease progression of an existing lung cancer can occur via nodal involvement and distant metastasis. Risks of disease progression depend on characteristics of any cancers present, and random individual variation that allows for more or less aggressive cancers, given a cancer's size and growth rate.

For each individual, 8 threshold volumes are drawn randomly from distributions for each nodal stage (N1, N2, N3) and for distant spread (M1), stratified by cell type (NSCLC/SCLC). Threshold volumes are adjusted to allow variation by growth rate. In each cycle, development of metastases and involvement of lymph nodes (N1, 2, 3) occurs if and only if the current volume of the largest cancer is greater than the corresponding adjusted threshold volume.

Symptom Detection

Each month, individuals with distant metastases and/or a primary lung cancer may develop symptoms that result in lung cancer detection. A person with symptom-detected cancer begins the following cycle in the Workup And Staging Component. See Symptom Detection Component for details.

Relevant Assumptions

Lung Cancer Development

The probabilities of developing each cancer type are assumed independent, conditional on the covariates (see Relevant Parameters, below). Each month >age 20, only one cancer can develop. Because the monthly probabilities are on the order of 10E-7, bias resulting from development of more than one cancer type is negligible.

Lung Cancer Growth

After Spratt, a maximum possible tumor size of 277 mm is assumed (this is consistent with the largest reported size of 201-300mm diameter in the SEER*Stat database for 60-64 year old white males, 1990-1994). As a simplification, we assume equal growth in all directions (i.e., spherical), allowing only one diameter to be tracked. The growth of BACs was truncated at a maximum diameter of 1cm.

Lung Cancer Progression

By definition, BACs do not progress. Because AJCC stage T3 cancers (i.e., cancers with extension into adjacent organs) represent only about 5% of NSCLC, we modeled tumor stage as T1 (≤3cm) or T2+ (>3cm). Involvement of lymph nodes (stages N0, N1, N2, and N3) dictates treatment options, so nodal status is modeled explicitly (but not specific nodes within each stage). Once distant spread (M1) has occurred, survival is poor, so explicit modeling of types of metastases was assumed to be unnecessary. As hypothesized for breast cancer, growth rate is related to the probability of metastasis. To reflect observed variations in propensity to metastasize for each histological type, adenocarcinomas are often more indolent, while small cell lung cancers develop metastases earlier. We assume that lymph nodes typically (but not always) become involved before distant spread occurs.

Lung Cancer Symptom Detection

We assume that peripheral cancers must be at least 10mm in diameter to cause symptoms. Central cancers have a smaller minimum diameter, because they are more likely to obstruct airways. We assume that metastases from SCLC cause symptoms faster than metastases from NSCLC. Benign nodules and lymph node involvement do not cause symptoms that result in lung cancer detection.

Relevant Parameters

The parameters in the Natural History Component are informed by calibration (see Calibration Details and Parameters Natural History).

Relevant Components

The Natural History Component occurs in every cycle, so can be thought of as a sub-component of the major states in the LCPM (see Component Overview for schematic).

Sub-components in the Natural History Component are described above or in linked documents:
lung cancer development, lung cancer growth, lung cancer progression, Symptom Detection Component, and Benign Component.

Dependent Outputs

The natural history component primarily determines the lung cancer incidence rate, as well as the type and stage distributions of incident cancers. The natural history component also primarily determines the survival rates of incident cancers, in conjunction with the Treatment Component.

The particular staging, work-up, and follow-up algorithms used in a scenario will also influence the stage distribution and rate of incident cancers, as well as the stage-specific survival rates (see the Workup And Staging Component). And the rates of thoracic imaging exams performed for reasons unrelated to screening (Incidental Imaging Component) will also influence incidence rates, although to a lesser extent than the natural history components.

Relevant Results

See the Results Overview for a description of the outputs from the base case LCPM.

REFERENCES:

[1] Raz, D J, He, B, Rosell, R, Jablons, D M "Current concepts in bronchioloalveolar carcinoma biology" in Clin Cancer Res 2005; 12: : 3698-3704

[2] Travis, W D, Garg, K, Franklin, W A, et al. "Evolving concepts in the pathology and computed tomography imaging of lung adenocarcinoma and bronchioloalveolar carcinoma" in J Clin Oncol 2005; 23: : 3279-3287

[3] Jackman, DM, Chirieac, LR, Janne, PA "Bronchioloalveolar carcinoma: A review of the epidemiology, pathology, and treatment." in Sem Respir Crit Care Med 2006; 26: : 342-352

[4] Jemal, A, Chu, K, Tarone, R "Cancer Surveillance Series: Recent trends in lung cancer mortality in the United States." in J Natl Cancer Inst 2001; 93: : 277-283

[5] Zheng, T, Holford, TR, Boyle, P, et al. "Time trend and the age-period-cohort effect on the incidence of histologic types of lung cancer in Connecticut, 1960-1989" in Cancer 1994; 74: : 1556-1567

[6] Wakelee, HA, Chang, ET, Gomez, SL, et al. "Lung cancer incidence in never smokers" in J Clin Oncol 2007; 25: : 472-478

[7] Moolgavkar, S. H., Knudson, A. G., Jr. "Mutation and cancer: a model for human carcinogenesis" in J Natl Cancer Inst 1981; 66: 6: 1037-52

[8] Puntoni, R., Toninelli, F., Zhankui, L., Bonassi, S. "Mathematical modelling in risk/ exposure assessment of tobacco related lung cancer" in Carcinogenesis 1995; 16: 7: 1465-71

[9] Rachet, B., Siemiatycki, J., Abrahamowicz, M., Leffondre, K. "A flexible modeling approach to estimating the component effects of smoking behavior on lung cancer" in J Clin Epidemiol 2004; 57: 10: 1076-85

[10] Tan, W. Y., Zhang, L.-J. "Stochastic modeling of carcinogenesis: state space models and estimation of parameters" in Discrete and Continuous Dynamical Systems - Series B 2004; 4: 1: 297-322

SCREENING COMPONENT

SUMMARY

This document describes the processes in the model responsible for screen-detection of asymptomatic lung cancers.

OVERVIEW

Note that during model calibration (to SEER registry data), the screening component was turned off.

To define a screening program, we can specify eligibility in terms of age and pack-year histories, as well as screen frequencies and probabilities of adherence to recommended screenings.

To reproduce results from a particular screening study (e.g., for validation), the model is populated with simulated trial participants and the screening component is turned on.

An individual with a positive screening exam proceeds in the next cycle to either the Follow Up Component or the Workup And Staging Component. The particular scenario being modeled determines which of these components a person will transition into.

DISEASE DETECTION MECHANISM

Lung cancer detection can occur in one of 4 ways: 1) by symptom detection of distant metastases, as a function of the time since metastases developed, varied by N/SCLC; 2) symptom detection of the primary cancer as a function of size and location; 3) by incidental detection of an asymptomatic lung cancer on a chest imaging exam performed for unrelated reasons; and 4) by a screening exam (imaging or biomarker), in a scenario with screening operating.

For any imaging exam, the probability of detection of asymptomatic cancers is a function of size, location (peripheral/central) and test characteristics.

SCREENING DISSEMINATION

For screening exams, individuals are screened if they are both 1) eligible for the screening program being modeled and 2) adherent, which is currently assigned randomly according to the population-wide probability of adherence.

The probability of incidental (non-screening) imaging exams is a function of age, region of the U.S., and race. See Incidental Imaging Component. Temporal trends in these background rates have not yet been explicitly incorporated.

Type / Detection Interaction

The probability of detection on an imaging exam is a function of nodule size (and therefore growth rate, indirectly) and location: central lesions are less likely to be detected. Both growth rate and the proportion central vs. peripheral vary by cell type. See the Natural History Component.

Stage / Detection Interaction

As stated above, the probability of detection of a pulmonary lesion on an imaging exam is a function of nodule size and location (central/peripheral) and the test characteristics of the imaging exam. Nodal involvement and distant metastases are not detected on a screening imaging exam so do not influence screen detection (but do of course influence symptom-detection - see the Natural History Component).

Length Bias

Slower-growing lesions persist in the asymptomatic state and are therefore more often 'available' to be screen-detected, on average, than faster-growing lesions. The probability of detection on an imaging exam is a function of lesion size (and therefore growth rate, indirectly). The growth rate varies by cell type. See the Natural History Component.

On average, lung cancers detected on annual screening exams would be expected to have longer doubling times (i.e., slower growth rates) than interval-detected lung cancers. Note that a small, slow-growing lung cancer may be referred for follow-up serial CT exams; if no growth is detectable over a two-year period, the cancer would be incorrectly diagnosed as benign.

See Screening Biases for background information on lead-time, length-time and overdiagnosis biases.

Detail

Parallel random number generation allows simulation of the same individuals in screening vs. non-screening scenarios. This allows us to compare the outcomes of individuals in the two scenarios, as well as the mean life expectancy across a cohort, for a better understanding of the range of individual outcomes attributable to screening.

When simulating a specific screening study, individual-level data from the study (if available) is used to populate the LCPM with a cohort similar to the study participants. See Protected Health Information.

Relevant Assumptions

We assume that nodal involvement and distant metastases are not detected on a screening imaging exam. See also the Assumption Overview.

Relevant Parameters

To enable simulation of screening, one parameter is set in the input file (intervention = 1 for screening, vs. 0 for no screening), and additional parameters define eligibility (based on pack-years of smoking exposure, years since quitting, and age), adherence rates, and screening frequency (modality, frequency, maximum number of screens, and follow-up algorithm).

Sensitivity and specificity of the screening exam also affect the efficacy of the screening program. See Parameters Test Performance.

Indicators record screen results and cancers detected.

Relevant Components

Under a screening scenario, individuals in the general population state are screened if they are 1) eligible, and 2) adherent. Persons in the Follow Up Component, the Workup And Staging Component, and the Treatment Component are not screened.

Screening will increase the rate of detection of lung cancer in a population and therefore impact the Incidence Component. Similarly, by detecting a lung cancer earlier, screening can alter the treatment a patient would receive in the Treatment Component.

Dependent Outputs

Because screening detects asymptomatic cancers, prevalence and incidence rates depend on the screening program in place (if any), as do stage distributions and cell types.

To date, we have used outputs from simulations of two single-arm screening studies to calibrate certain endpoints, to validate the LCPM, and to predict outcomes from hypothetical control arms.

Relevant Results

The Mayo Clinic conducted a single-arm study of helical CT screening for lung cancer in current and former smokers. Using data provided by the Mayo Clinic (see Protected Health Information), we replicated the trial population by bootstrapping demographics and smoking histories from individual records. One endpoint (baseline prevalence) was used to calibrate the proportion of adenocarcinoma that was BAC (see Natural History Component). Remaining endpoints were reserved for use as validation endpoints.

See the validation section of Calibration Validation Results for a link to a description of validation of the LCPM using the LSS study endpoints.

See the Results Overview for analyses of screening programs.

TREATMENT COMPONENT

SUMMARY

This document describes how treatment after diagnosis is modeled.

OVERVIEW

To enter the treatment component, patients must have been diagnosed with lung cancer in the Workup And Staging Component. Treatment is modeled as occurring in the month(s) after reaching the 'treatment and survival' state. Patients remain in this state until death (from any cause). See also the Survival Mortality Component.

Treatment consists of either removal of the primary lung cancer (i.e., resection) or systematic therapy. Tumors which respond to systemic therapy are reduced in size (diameter), following conventional guidelines for solid tumors.[1]

TREATMENT DISSEMINATION

Treatment is assigned based on the diagnosed stage and type (NSCLC/SCLC). We currently assume all patients receive care according to consensus guidelines (e.g., National Comprehensive Cancer Network, NCCN).

We are in the process of adding a 'usual care' option that more closely approximates observed practice patterns. The usual care option will allow us to explicitly vary treatments with calendar year, which will be particularly important in the Population LCPM.

TREATMENT EFFICACY

Treatment effectiveness is incorporated as follows: a person with no occult metastases whose primary cancer is resected is assigned competing risks consistent with a person of the same smoking history – not stage I survival from SEER. On the other hand, if occult metastases are present in a person who undergoes resection for an apparent stage I cancer, the metastases continue to develop as before. (The presence of undetected micro metastases is likely the cause of the poor observed survival after "curative" resection in many patients.)

If a second, undetected primary tumor remains (in a non-resected lobe), metastasis can occur. Note that removal or sampling nodes at resection can result in re-assigning stage at diagnosis, but provides no survival benefit.

For systemic therapies, we use probabilities of partial and complete responses as published in the literature (see Parameters Treatment). A response results in a reduction of the size(s) of existing lung cancer(s), and thereby may delay disease progression.

We do model adverse effects of screening and treatment. Operative mortality can occur during resection, mediastinoscopy, or VATS, in diseased or non-diseased persons (iatrogenic deaths are tracked). We have not yet incorporated complications (e.g., pneumothorax) or quality of life.

Detail

The default Treatment Component simulates all patients as receiving guideline care.

Only individuals assigned the status of operative candidate were eligible for surgical resection, regardless of stage. To account for patients who were not operative candidates, we estimated proportions of SEER cases who were NSCLC stage I and II and either refused surgery or had contraindications.

As observed in clinical practice, a small proportion of operative candidates with stage LS (limited stage) SCLC underwent resection, with the remaining patients receiving chemoradiation.

Operative candidates with NSCLC stages I through IIIa were assigned resection, with the remainder and all stage IIIb cases receiving chemoradiation.

Stages IV (NSCLC) and ES (SCLC) were assigned chemotherapy.

See Parameters Treatment.

Relevant Assumptions

Death from lung cancer is unlikely to occur without detection of metastases (due to symptoms or otherwise), so we assume that once metastases are detected (by symptoms or any modality), survival is as observed in SEER for stage IV-detected cancer. We estimated cause-specific (net) survival for cases diagnosed as stage IV in the years 1990 – 2000, stratified by 10-year age group, race, and gender. Net survival (i.e., in absence of other causes) was used because persons face competing risks elsewhere in the model.[2]

Median survival of stage IV lung cancer is uniformly less than one year, so survival for M1 (stages IV or ES) is modeled as exponential.

See also the Assumption Overview.

Relevant Parameters

Parameter values that define treatment efficacy are probabilities of complete or partial response, using the definition of complete as no visible cancer at 4 week follow-up and partial as ≥30% decrease in diameter.[1] Probabilities of complete and partial responses vary by type, with estimates taken from the literature. A cancer that partially responds to therapy is decreased in diameter by 30%, and a cancer that completely responds to therapy is reduced to 1.5mm diameter, or below the 2mm detection threshold assumed for helical CT.

Based on the new diameter, an adjusted 'time since cancer developed' is calculated, retaining the original growth parameter, alpha. The new 'time since cancer developed' is used to increment growth in all future cycles.

To account for observed differences in growth rates of cancers pre- and post-therapy, we include a parameter that allows faster-growing cancers (cancers with a parameters over a specified cutoff) to be more or less likely to respond to therapy. These parameters were initially set to values that conferred no effect (probabilities of response vary by cell type and treatment) but were varied during calibration.

See Parameters Treatment.

Relevant Components

Treatment assignment occurs as the final step in the Workup And Staging Component (i.e., after the stage at diagnosis is assigned). The treatment itself is the first step in the Treatment and Survival state.

The sensitivity and specificity of the staging algorithm influences the treatment assignment. The specific treatment assigned and the treatment's effectiveness both influence the survival rate.

Dependent Outputs

Survival by stage is dependent on the treatment assigned and the treatment's effectiveness. For example, assigning systemic therapy instead of resection to a patient with resectable disease will result in a shorter survival time.

Further, the sensitivity and specificity of the staging algorithm (Parameters Test Performance) influence the treatment assignment. For example, if the staging tests performed on a hypothetical patient miss the involvement in a contralateral node, the individual will be understaged and receive an inappropriate treatment.

Mortality rates are calculated as a secondary output, based on age at death among lung cancer patients.

Incidence rates, on the other hand, depend on neither the treatment assigned nor the treatment effectiveness.

Relevant Results

See Output Overview and Results Overview.

REFERENCES:

[1] Therasse, P, Arbuck, SG, Eisenhauer, EA, Wanders, J, Kaplan, RS, Rubinstein, L, Verweij, J, Van Glabbeke, M, van Oosterom, AT, Christian, MC, Gwyther, SG "New guidelines to evaluate the response to treatment in solid tumors" in Journal of the National Cancer Institute 2000; 92: : 205-216

[2] McMahon, P. M., Zaslavsky, A. M., et al. "Estimation of mortality rates for disease simulation models using Bayesian evidence synthesis" in Medical Decision Making 2006; 26: : 497-511

Survival Mortality Component

Summary

This document describes how lung cancer survival is modeled and describes estimation of both lung cancer-specific and other-cause mortality.

Overview

Once a hypothetical patient is diagnosed with lung cancer, he or she moves into the 'treatment and survival state', and remains there until death. The Treatment Component precedes the Survival Mortality Component.

The patient can die from any cause while in this state. Relative risks of each cause of death are a function of underlying disease state, treatment received, any surveillance performed, age, gender, race, and smoking history.

Survival Estimation Covariates

Lung cancer-specific survival for patients diagnosed at stages I-III (i.e., M0) is based on the true, underlying disease stage and the treatment assigned. Stage-specific survival for patients with M0 cancers is a calibration target for the LCPM, not an input. See Calibration Details and below.

Once a person is diagnosed (by symptoms or any modality) as stage IV (i.e., M1), lung cancer specific survival is assumed exponential. Cell-type (N/SCLC) specific median survival by age, gender, and race group was estimated from SEER for appropriate calendar years as inputs.

Survival after clinical detection

Same as above.

Screen Detection Benefit

Screening may detect cancers that have not yet metastasized.

Mortality Reduction

Resection of an early-stage cancer is curative if 1) no occult metastases remain and 2) no additional lung cancers arise. The person is subject to competing mortality risks appropriate to his/her age and smoking history (see below).

Resection of an early-stage cancer in a patient with occult metastases does not confer a survival benefit.

Mortality rates are calculated after the last individual in a simulation run 'dies', and are merely counts of lung cancer deaths by age at death, divided by the population at risk. Mortality rates are therefore completely dependent on the incidence and survival rates, and are merely an additional way to present the same information.

Other Cause Mortality

The Population LCPM assigns other cause mortality according to estimates from the Smoking History Generator, a common input for the lung cancer models in CISNET.

The original single-cohort LCPM assigns other-cause mortality risks using results from an independently-conducted analysis. We developed a Bayesian evidence synthesis model to estimate cause-specific mortality rates stratified by age, sex, race, and smoking status.[1] We combined three data sources: 1) individual survey data on smoking status, demographics, and date and cause of death; 2) population data on numbers of deaths by cause; and 3) cohort study estimates of smoking-related mortality risks, correcting for known inconsistencies between two of the data sources. Risks of heart disease and other causes (i.e., non-lung cancer, non-heart disease) are used as inputs for the LCPM.

Leadtime

Stage-specific survival is not an input, but rather a calibration target (see Calibration Details and Output Overview).

Estimates of lead, length, and overdiagnosis biases (see Screening Biases) are outputs of the model, obtained by simulating the same cohort with and without screening.

Detail

If a patient undergoes curative resection of an early stage lung cancer and harbors no occult metastases, the patient is assigned competing mortality risks appropriate for his or her age, gender, race, and smoking status.

If a patient undergoes resection of an early stage lung cancer but does harbor occult metastases, the metastases continue to develop as before, and can cause symptoms. Once symptomatic, the person is assigned a stage IV survival as above.

Patients who undergo systemic therapies may respond to the therapy, which results in a reduction in the size of the primary lung cancer(s), and therefore a reduction in the monthly probability of disease progression (see Natural History Component).

Relevant Assumptions

See the Assumption Overview, the Treatment Component and the Natural History Component.

Relevant Parameters

See Parameters Treatment for values of parameters that govern effectiveness of treatment and influence survival rates.

Relevant Components

See the Treatment Component and the Natural History Component.

Dependent Outputs

Stage-specific survival is dependent on this component (Calibration Survival1). Incidence rates, however, are not dependent on this component.

Relevant Results

See Results Overview and Output Overview for more information on calibration and validation outputs.

References:

[1] McMahon, P. M., Zaslavsky, A. M., et al. "Estimation of mortality rates for disease simulation models using Bayesian evidence synthesis" in Medical Decision Making 2006; 26: : 497-511

OUTPUT OVERVIEW

SUMMARY

This document describes the types of outputs generated by the LCPM.

OVERVIEW

See Results Overview for a summary of how the various outputs are used for calibration, validation, predictions, and analyses.

Some general categories of outputs include:

incidence rates;
characteristics of incident cancers;
survival and mortality rates;
screening test results;
effectiveness of screening tests;
estimation of screening biases.

OUTPUT LISTING

Within each general category, some examples of specific outputs include:

incidence rates;
Age-specific incidence rates, by gender, race, and calendar year (used as calibration endpoints for model development; see Calibration Incidence1)
Age-adjusted incidence rates (predictions from the Population LCPM)

characteristics of incident cancers;
Size, type and stage distributions of incident cancers (used in Calibration Size Type Stage1 as calibration endpoints for model development)

survival and mortality rates;

Survival curves by type (SCLC vs. NSCLC) and stage at diagnosis (used in Calibration Survival1 as calibration endpoints for model development)

Mortality rates (used in Validation Cohort Studies1 as validation endpoints, in comparison to published cohort studies)

Age-adjusted mortality rates (used in Calibration USMortality1 as calibration endpoints for the Population LCPM)

screening endpoints;

Estimation of effectiveness of screening (see Results Overview for publications)

Reproduction of observed endpoints in the LSS screening trial (used in Validation Lung Screening Study1 as validation endpoints)

Screening trial endpoints:

-stage shift

-number of surgeries (appropriate and inappropriate)

-number of invasive work-up procedures (appropriate and inappropriate)

estimation of screening biases

By simulating both screened and unscreened scenarios, the model estimates lead-time, length-time, and overdiagnosis (see Screening Biases).

RESULTS OVERVIEW

SUMMARY

This document will discuss results from the LCPM and provide links to published evaluations of lung cancer control interventions.

OVERVIEW

We see simulation as a tool for integrating data from national sources such as SEER with individual-level data from screening trials. By synthesizing available evidence, we can impute unobserved results.
This is conceptually similar to the 'borrowing strength' idea from Bayesian statistics. Extrapolating from available data allows us to pose a wide range of interesting questions regarding cancer control interventions in a wider variety of populations than represented in trials.

Results from the LCPM have been used to calibrate and validate the model, and evaluate screening programs, smoking cessation programs, and treatments. Additional types of results could include estimation of natural history parameters.

RESULTS LIST

CALIBRATION AND VALIDATION

See Calibration Validation Results for results of calibration and validation.

Kong CY, Mc Mahon PM, Gazelle GS. Calibration of Disease Simulation Models Using an Engineering Approach. Value in Health. 2008 In Press. See Calibration Methods Research for a description of this evaluation of advanced engineering methods used for calibration of the LCPM.

SCREENING EVALUATIONS

McMahon PM, Kong CY, Johnson BE, Weinstein MC, Weeks JC, Kuntz KM, Shepard JA, Swensen SJ, Gazelle GS. Estimating Long-term Effectiveness of Lung Cancer Screening in the Mayo CT Screening Study. Radiology. 2008 Jul;248(1):278-87. Epub 2008 May 5. PMID: 18458247 [Pub Med - as supplied by publisher]

See also Validation Lung Screening Study1 for results from simulating the CT-screened arm of the LSS study.

McMahon PM, Kong CY, Weinstein MC, Tramontano AC, Cipriano LE, Johnson BE, Weeks JC, Gazelle GS. Adopting helical CT screening for lung cancer: potential health consequences during a 15-year period. Cancer. 2008 Dec 15;113(12):3440-9. PMID: 18988293 [Pub Med - indexed for MEDLINE]
See Index Supplement Cancer for information that might be helpful for readers of this analysis.

An evaluation of the cost-effectiveness of helical CT screening for lung cancer is underway and will be described in a future results document.

Treatment Evaluations

Evaluations of specific treatments (ablation, other) will be described in future results documents.

Population Trends

Trends in lung cancer incidence and mortality under various scenarios (e.g., the Smoking Base Case) will be described in future results documents.

SUMMARY OF VERSIONS

Summary of differences and similarities between the original single cohort LCPM (v.1) and the Population LCPM (v.2).

Original single cohort LCPM (v.1)

*Used for published screening evaluations and to document calibration methods.[1,1,2]
*Smoking histories for U.S. cohorts were derived from survey data, as described in the Population Component.
*Competing mortality risks are stratified by smoking status, age group, race, and sex.[3]
*Natural history parameter values are described in Parameters Natural History.
*Calibration targets for birth cohort terms were age-specific incidence for 5 cohorts in 1990 (males aged 50 and 70, and females aged 50, 60, and 70).
*Calibration targets for period terms - not applicable.

Population LCPM (v.2)

*Used for all analyses in Risk Analysis Monograph and Moolgavkar, et al. (forthcoming)
*Smoking histories for U.S. cohorts and competing mortality risks by smoking status, age, and sex were from the shared smoking history generator (cite Chapter 3 and the Smoking History Generator Component.
*Natural history parameter values consistent with a stronger relationship between years of smoking and lung cancer risk; weaker relationship between cigarettes per day and lung cancer risk; greater benefit from quitting. See also Calibration Details.
*Calibration targets for birth cohort and period terms were age-adjusted mortality rates for US population 1975-2000

Similarities

Calibration targets for natural history parameters (excluding birth cohort and period terms) were age-specific incidence rates for cohort of white males aged 60 in 1990, and cell, stage, and size distributions as described previously (cite) and in Calibration Details.

REFERENCES:

1No Reference found for: McMahon, 2008

2No Reference found for: Kong, 2009

3 McMahon, P. M., Zaslavsky, A. M., et al. "Estimation of mortality rates for disease simulation models using Bayesian evidence synthesis" in Medical Decision Making 2006; 26: : 497-511

Calibration Validation Results

Summary

This document summarizes the calibration of the LCPM and validation to additional endpoints.

Result Type

Other

Overview

Calibration and validation results indicate no immediately obvious departure from observed data, lending credence to simulations of hypothetical scenarios (i.e., those that extrapolate past observed data).

Many limitations of the LCPM are common to all studies employing simulation models. Tradeoffs must be made between increasing complexity and practical limits on the number of unknown parameters that can be identified using available data. A 'deep' model like the LCPM has more complexity (which allows us to evaluate different workup algorithms) than a 'shallow' statistical model that estimates transition probabilities (e.g., stage I to stage II), but at the cost of greater risk of identifiability problems. To reduce the risk of identifiability issues biasing results, we continue to select additional calibration targets, refine calibration approaches, remove parameters where possible, and identify additional sources of data for inputs and validation.

Methods

See the Population Component and Calibration Details for descriptions of calibration of the original single cohort LCPM and the more recent Population LCPM.

For both the single cohort LCPM and the Population LCPM, birth cohort terms were estimated.

Result

Fit to Calibration Targets

Primary Targets, Derived from SEER

The single cohort LCPM produced a good fit to incidence by age for cohorts of 50-, 60-,

and 70 year-old whites (Calibration Incidence1). We also achieved good fits to size, type, and stage distributions (Calibration Size Type Stage1). The best-fitting set slightly overestimated ≥3-year survival for NSCLC stages I and II (Calibration Survival1). This overestimation is due to our assumption that all patients underwent guideline staging and treatment; many patients represented in SEER did not receive guideline treatments. Survival of patients with limited stage small-cell lung cancer was accurately predicted.

The Population LCPM produced a good fit to age-adjusted lung cancer mortality rates over the period 1975 to 2000 (see Calibration USMortality1).

Secondary Targets, Derived from Cohort Studies and Literature

The single cohort LCPM predicted annual mortality rates per 100,000 non-smokers and lung cancer-specific mortality ratios for current (vs. never) smokers (by 5-year age group) that agree with observed data (Calibration Cohort Study1).

The LCPM predicted lung cancer outcomes in non-smokers and in autopsy studies that agreed with published findings (Calibration Non Smokers Autopsy1).

Validation

Validation is documented here (Validation Cohort Studies1 and Validation Lung Screening Study1).

Discussion

After calibration and validation of the LCPM, the model could be used to evaluate screening programs.

Because screening is not part of usual clinical practice, most lung cancers in the SEER registry were diagnosed on the basis of symptoms. The SEER calibration targets used to inform estimates of incidence and survival were therefore supplemented with screening trial data to refine estimates regarding noninvasive BACs, which appear with greater frequency in screening studies.

Conclusion

The LCPM generates outputs consistent with multiple data sources. Predictions from the model regarding the effectiveness of screening or other interventions are however extrapolations beyond available data, and are subject to all assumptions built in to the model.

Relevant Assumptions

The base case assumption that all individuals receive guideline care is necessary, given the lack of data on staging practices in the US. However, this assumption likely yields fewer understaged patients and therefore higher survival for early-stage cancers than in SEER data used for calibration.

See also Assumption Overview.

Relevant Parameters

Calibration was used to estimate unobservable parameters (e.g., those that govern metastasis). See Parameters Natural History for details.

Relevant Outputs

See Output Overview.

SCREENING BIASES

Screening trial results are affected by several well-known biases that make interpretation of results challenging.[3,4] A test that detects earlier-stage disease, will, by definition, prolong observed survival times (lead-time bias). Assuming some individual heterogeneity in disease progression rates, periodic screening will preferentially detect slowly progressing cases, simply because such cases persist longer in the asymptomatic state (length-time bias). The extreme of length-time bias, overdiagnosis refers to both screen detection of cases that would not have caused symptom detection or death (i.e., without screening, the person would die of competing causes, unaware of the presence of the disease) and to detection of pseudo-disease (e.g., cases with a self-resolving clinical course).[6]

REFERENCES:

[1] Reich, J. M. "Improved survival and higher mortality: the conundrum of lung cancer screening.[see comment]." in Chest 2002; 122: 1: 329-37

[2] Black, W. "Overdiagnosis: An underrecognized cause of confusion and harm in cancer screening [Editorials]" in Journal of the National Cancer Institute 2000; 92: 16: 1280-1282

[3] Patz, E., Goodman, P., et al. "Screening for lung cancer [Review]" in New England Journal of Medicine 2000; 343: 22: 1627-1633

[4] Eddy, D. "Screening for lung cancer." in Annals of Internal Medicine 1989; 111: : 232-237

[5] Woods, William G., Gao, Ru-Nie, Shuster, Jonathan J., Robison, Leslie L., Bernstein, Mark, Weitzman, Sheila, Bunin, Greta, Levy, Isra, Brossard, Josee, Dougherty, Geoffrey, Tuchman, Mendel, Lemieux, Bernard "Screening of infants and mortality due to neuroblastoma" in N Engl J Med 2002; 346: 14: 1041-1046

[6] Schilling, Freimut H., Spix, Claudia, Berthold, Frank, Erttmann, Rudolf, Fehse, Natalja, Hero, Barbara, Klein, Gisela, Sander, Johannes, Schwarz, Kerstin, Treuner, Joern, Zorn, Ulrich, Michaelis, Joerg "Neuroblastoma screening at one year of age" in N Engl J Med 2002; 346: 14: 1047-1053

CALIBRATION DETAILS

OVERVIEW

Calibration was used to estimate values of unobservable natural history parameters and uncertain parameters (those for which literature estimates provided ranges of values). Each unique combination of model inputs (tables, distributions, and scalar values) is referred to as a parameter set. A combination of grid search and simulated annealing was used to identify the parameter set that minimized the total sum of squared errors between model output and 8 primary calibration targets. Of the 25 parameter sets with the best fit to the primary calibration targets, we chose the set with the best fit to 5 secondary calibration targets. Extensive debugging was performed throughout model development and prior to final calibration.

CALIBRATION TARGETS

Primary: Age-specific incidence, cell type, stage, and size distributions of incident cancers, survival curves (4 stages).
Secondary: Autopsy data, mortality in never-smokers, percent symptomatic at detection, lung cancer mortality.

Please see Calibration Methods Research for information on a comparison of calibration approaches.

ORIGINAL SINGLE COHORT LCPM

We chose a large cohort (white males aged 60 in 1990) as the initial calibration cohort, setting the birth cohort term to 1.0 (reference group). Once calibration to this large cohort was completed, the same parameter set was used to generate incidence by age predictions for cohorts of 50 and 70-year old white males and 50, 60, and 70-year old white females. To account for observed birth cohort trends in lung cancer risks and allow for differences in baseline risk by gender,[1,2,3] we added a term that modifies the monthly risk of lung cancer development (all cell types), stratified by gender. The birth cohort term was adjusted in these cohorts such that the LCPM generated observed age-specific incidence rates. See Population Component for a description of smoking histories used in the single cohort LCPM and[4] for a description of other cause mortality rates, and how these differed from inputs for the Population LCPM.

Population LCPM

The Population LCPM assigns smoking histories and other cause (non- lung cancer) mortality risks from the Smoking History Generator that is common to all CISNET lung models that simulate populations. The Population LCPM was re-calibrated to the same calibration targets used for the original single cohort LCPM (see below) to generate a revised parameter set ("version 2") that assigns a stronger dose-response relationship between years of smoking (duration) and lung cancer risk. The effect of cigarettes per day (dose) is correspondingly lower in the "version 2" parameter set. See Parameters Natural History for further details on smoking parameters and differences in birth cohort terms between parameter sets (versions 1 and 2).

Targets

See Output Overview for links to comparisons of targets and outputs from the calibrated LCPM.

Defining Ranges for Unobservable Natural History Parameters

See Natural History Component and Parameters Natural History. During calibration, some parameter values could be ruled out as implausible, after consultation with clinical experts and past research. For example, the intercept terms were ordered to reflect observed risks of each cell type among non-smokers. Lung cancer risks increase with age and SY and decrease with YSQ. SY has the strongest effect on development of small cell cancers, and the effect of YSQ was weakest for adenocarcinoma. The amount of BAC as a proportion of adenocarcinoma was varied from 0 to 0.4[5,6] and estimated to be 0.2.

Initial values of symptom detection parameters were selected so that the cumulative probability of symptom detection from (true) distant metastases was nearly 1.0 by 3 years, i.e., very few patients had asymptomatic/undetected metastases at 3 years after diagnosis, but it was not impossible. By comparison, the estimated growth duration of metastases was 3.8 years (faster growth than the primary tumor) in a published breast cancer model. Initial values of the intercept term and coefficient on tumor volume for symptom detection of primary cancers were chosen so that once a cancer passed the threshold size, the probability of symptom detection increased slowly to yield lung cancers of similar sizes as those observed in SEER. Adenocarcinomas were assumed to be less aggressive than SCLCs.

Defining Ranges for Uncertain Parameters

We classified parameters as uncertain if literature estimates provided ranges of values. During calibration, test characteristics were allowed to vary because verification bias likely affects many published sensitivity and specificity values. See Parameters Test Performance and Parameters Treatment for details.

Results

See Calibration Validation Results for model calibration and validation.

References:

[1] Jemal, A, Chu, K, Tarone, R "Cancer Surveillance Series: Recent trends in lung cancer mortality in the United States." in J Natl Cancer Inst 2001; 93: : 277-283

[2] Wakelee, HA, Chang, ET, Gomez, SL, et al. "Lung cancer incidence in never smokers" in J Clin Oncol 2007; 25: : 472-478

[3] Zheng, T, Holford, TR, Boyle, P, et al. "Time trend and the age-period-cohort effect on the incidence of histologic types of lung cancer in Connecticut, 1960-1989" in Cancer 1994; 74: : 1556-1567

[4] McMahon, P. M., Zaslavsky, A. M., et al. "Estimation of mortality rates for disease simulation models using Bayesian evidence synthesis" in Medical Decision Making 2006; 26: : 497-511

[5] Raz, D J, He, B, Rosell, R, Jablons, D M "Current concepts in bronchioloalveolar carcinoma biology" in Clin Cancer Res 2005; 12: : 3698-3704

[6] Crestanello, J. A., Allen, M. S., Jett, J. R., Cassivi, S. D., Nichols, F. C., 3rd, Swensen, S. J., Deschamps, C., Pairolero, P. C. "Thoracic surgical operations in patients enrolled in a computed tomographic screening trial" in Journal of Thoracic & Cardiovascular Surgery 2004; 128: 2: 254-259

FOLLOW UP COMPONENT

As described in the Incidental Imaging Component, patients with incidentally detected nodules suspicious for lung cancer are triaged according to the size of the nodule. Patients with nodules over the threshold diameter are sent to the Workup And Staging Component.

In sensitivity analyses of a manuscript in press, we examined scenarios where nodules smaller than a cutoff size (e.g., 4mm diameter) are ignored, with patients returning to the general population.

Smaller nodules are followed-up with serial high-resolution CT exams (even in the absence of screening), with a specified periodicity (see below) over 24 months. Detection of new small nodules re-starts the 24-month follow-up sequence. Nodules that exhibit no detectable growth (see below) after 2 years of follow-up are diagnosed as benign;[1] detectable growth on any subsequent imaging exam is considered sufficient to cause suspicion for lung cancer.

Structural Parameters in the Follow Up Component
(See Parameter Overview for definition):
1) A threshold (or 'cutoff') diameter of 8mm was used and is generally in agreement with a low (5%) biopsy rate for 4-9mm nodules in a recent trial[2].
2) Depending on the scenario, follow-up could occur with a fixed periodicity of 1, 3, 6, 12, and 24 months[3] or be managed according to the size of the largest nodule found (similar to published algorithms from CT screening trials)[4].
3) For the base case, the minimum detectable growth on sequential exams was 2mm[1].
4) An estimated 50% of growing nodules are excisionally biopsied using VATS (video-assisted thoracic surgery).

REFERENCES:

[1] Benjamin, MS, Drucker, EA, McLoud, TC, Shepard, JO "Small pulmonary nodules: detection at chest CT and outcome" in Radiology 2002; 226: : 489-493
[2] Pinsky, P. F., Marcus, P. M., Kramer, B. S., Freedman, M., Nath, H., Kvale, P., Reding, D. "Diagnostic procedures after a positive spiral computed tomography lung carcinoma screen" in Cancer 2005; 103: 1: 157-163
[3] Rubins, JB, Rubins, HB "Temporal trends in the prevalence of malignancy in resected solitary pulmonary lesions" in Chest 1996; 109: : 100-103
[4] Swensen, S., Jett, J., et al. "Screening for lung cancer with low-dose spiral computed tomography" in American Journal of Respiratory and Critical Care Medicine 2002; 165: : 508-513

Workup And Staging Component

Summary

This document describes the one-month cycle during which a patient with a pulmonary nodule suspicious for lung cancer undergoes clinical workup to establish a diagnosis of cancer or benign histology. Patients with cancer then undergo staging tests to establish the extent of disease progression.

Overview

Clinical algorithms for workup and staging are modeled explicitly, so that differences in patient management strategies can be compared.

Patients enter this component if they have a pulmonary nodule suspicious for lung cancer, detected by any modality and large enough to be biopsied. Patients with incidentally detected nodules smaller than the threshold go to the Follow Up Component.

Biopsy-confirmed malignancies are clinically staged (in the same cycle) based on practice guidelines from the National Comprehensive Cancer Network (NCCN, version 2000, for calibration to 1990-2000) and assigned both TNM and AJCC stages.

Patients whose pulmonary nodules are definitively diagnosed as benign start the next cycle in the general population state. Patients with a diagnosis of lung cancer begin the next cycle in the Treatment Component.

Detail

Patients presenting with symptom-detected cancers undergo biopsy to establish the histological type and a high-resolution CT examination to stage lymph nodes and determine tumor size, if not already known.

All patients undergo one high-resolution CT examination to determine calcification pattern and/or stage lymph nodes, per NCCN guidelines.

An estimated 50% of patients with N0/1 and evidence of primary tumor growth on CT are sent for excisional biopsies using VATS (video-assisted thoracic surgery). Non-operative candidates and remaining N0/1 patients undergo biopsy of the primary tumor (bronchoscopy for central nodules and TTNA for peripheral nodules).

Patients with clinically evident enlarged mediastinal lymph nodes (N2/3 on CT) undergo mediastinoscopy, which can establish a diagnosis of lung cancer and provide staging information. Patients with negative mediastinoscopy results are treated as N0/1.

Relevant Assumptions

The assumption that all patients undergo staging tests in accordance with consensus practice guidelines is a limitation of the current LCPM. Many publications have demonstrated that a large proportion of lung cancer patients do not receive guideline treatments, but information on staging practices in the U.S. is limited.

Our assumption of guideline staging, which is often more invasive/aggressive than usual care, will result in higher survival rates for earlier staged cancers, because patients with more advanced disease (and lower survival) will be correctly identified and categorized into later stages.

Relevant Parameters

Sensitivity and specificity of the diagnostic and staging tests will influence patient trajectories through this state (Parameters Test Performance).

Other input parameters define the scenario being simulated. For example, we will have the flexibility to simulate either guideline or usual care staging patterns.

Relevant Components

The Workup and Staging Component includes sub-components specific for NSCLC and SCLC, which have different characteristics and therefore different staging systems.

A "Usual Care" version of the staging component is in development and will be available for comparison to guideline staging.

Dependent Outputs

Distributions of stage at diagnosis and survival outputs rely on this component. Natural History and Incidence do not depend on this component.

Relevant Results

Size, Type and Stage Distributions of Incident Cancers (Calibration Size Type Stage1)

Survival Curves for NSCLC and SCLC (Calibration Survival1)

PARAMETERS NATURAL HISTORY

See the Natural History Component and Calibration Details for explanations of the way the LCPM models lung cancer natural history and how the unobservable parameters were estimated. This document contains additional detail not already provided.

LUNG CANCER DEVELOPMENT

The logistic model for development of a lung cancer was described in the Natural History Component. There are type-specific parameters for the effects of age and smoking history. We assume that beta_highrisk = ln(2), equivalent to a HR of 2 for individuals positive for the susceptible genotype, independent of age, gender, smoking status and pack-years. As a candidate 'susceptibility' genotype, we model the combined genotype of GST P1 (GG) and p53 (Arg/Pro or Pro/Pro) to occur with an estimated population frequency of 4.7% (no linkage).[1] The amount of BAC as a proportion of adenocarcinoma was varied from 0 to 0.4 during calibration and was estimated to equal 0.2.

As described in Calibration Details, values of smoking-related natural history parameters differ between the original single cohort LCPM and the Population LCPM.

LUNG CANCER LOCATION

Each newly-developed lung cancer is assigned a location, with indicators for the specific lobe in the lung and central/peripheral location. Most lung cancers occur in upper lobes, and the proportion central varies by cell type (more SCLC are central compared to adenocarcinomas).

LUNG CANCER GROWTH

See Table Growth Parameters. Lung cancers was assumed to grow 2-fold faster in smokers, although the difference may be due at least in part to 'type mix', i.e., non-smokers are more likely to develop slow-growing adenocarcinomas.[2,3,4] The growth of BACs was truncated at 1.0 cm diameter (detectable by CXR). For non-BACs, we assumed a maximum possible tumor size of 27.7 cm,[5] consistent with the largest reported size of 20.1-30.0cm diameter in the SEER*Stat database.

Lung Cancer Progression

Mean (SD) unadjusted threshold volumes for SCLC ranged from 0.61 (0.65) ml for N1 involvement to 4.07 (4.13) for N3 involvement and 4.71 (4.14) for distant metastases. Corresponding unadjusted threshold volumes for NSCLC ranged from 3.34 (4.09) ml for N1 involvement to 3.8 (4.64) for N3 involvement and 2.62 (3.18) for distant metastases. Adjustments were then estimated via calibration to allow the propensity to progress to vary by cell type and be correlated with the growth parameter assigned to the person's cancer. Note that the final estimated volume at metastasis development is an output of the model and will vary across populations that differ in terms of age, smoking history, and scenario (e.g., whether screening is in place).

Symptom Detection

See the Symptom Detection Component. The cumulative probability of symptom detection from (true) distant metastases was over 95% by 3 years (all cell types combined). By comparison, the estimated growth duration of metastases was 3.8 years in a breast cancer model.[6] Treatment with targeted therapies (eg, erlotinib) will influence the rate of symptom detection from metastases.

References:

[1] Miller, D. P., Liu, G., et al. “Combinations of the variant genotypes of GSTP1, GSTM1, and p53 are associated with an increased lung cancer risk” in Cancer Research 2002; 62: : 2819-2823

[2] Usuda, K., Saito, Y., et al. “Tumor doubling time and prognostic assessment of patients with primary lung cancer” in Cancer 1994; 74: : 2239-2244

[3] Song, P, Sekhon, HA, Jia, Y, Keller, JA, Blusztajn, JK, Mark, GP, Spindel, ER “Acetylcholine is synthesized by and acts as an autocrine growth factor for small cell lung carcinoma” in Cancer Research 2003; 63: : 214-221

[4] Hasegawa, M, Sone, S, Takashima, S, Li, F, Yang, Z-G, Maruyama, Y, Watanabe, T “Growth rate of small lung cancers detected on mass CT screening” in British Journal of Radiology 2000; 73: : 1252-1259

[5] Spratt, J. S., Spratt, J. A. “Growth Rates” in Cancer of the Breast 2002;

[6] Koscielny, S, Tubiana, M, Valleron, A-J “A simulation model of the natural history of breast cancer” in British Journal of Cancer 1985; 52: : 515-524

Parameters Test Performance

Diagnostic test characteristics (sensitivity and specificity) determine the probabilities of detection and/or diagnosis of lung cancers or benign lesions.

Pulmonary Nodules

Sensitivities of imaging examinations for peripheral pulmonary nodules are input from tables indexed by diameter of the lesion. Sensitivities for a central lesion of the same diameter were assumed to be 25% lower than those for peripheral lesions. We derived sensitivities from the literature to test during calibration.

Computed Tomography (CT)

Helical CT was estimated to have a sensitivity of 0.63 for 1-4mm peripheral nodules, 0.77 for 4-8mm peripheral nodules, and 1.0 for peripheral nodules >8mm. High-resolution CT was assumed to have equivalent sensitivity for detecting presence of a nodule as helical CT (by size), but to have greater resolution for calcification patterns. As in clinical practice, an estimated 11% of benign nodules are diagnosed by high-resolution CT as benign on the basis of calcification pattern (not explicitly modeled). CT could occur in several components: Incidental Imaging Component, Workup And Staging Component, Screening Component, and during surveillance for recurrent disease in the Survival Mortality Component.

Chest X-ray (CXR)

The sensitivity of CXR was estimated to be approximately 25% to 50% of that of helical CT, and to be less than 1.0 at 16mm. The minimum detectable size was assumed to be 7.5mm. CXR could occur in the Incidental Imaging Component.

Specificity for both helical CT and CXR is assumed to be 0.98 (per person, or 0.997 per nodule). Specificity of high-resolution CT for pulmonary nodules was assumed to be 1.0.

Staging Examinations

Computed Tomography

High-resolution CT was estimated to have a sensitivity for nodal involvement of 0.63 (average of N1, N2/3) and a specificity of 0.6.[1,2]

Bronchoscopy

Sensitivity of bronchoscopy is defined as the probability of establishing a definitive diagnosis on the basis of cells recovered from the nodule. The sensitivity increases with increasing size of the nodule. For cancer, the sensitivity is 5% for nodules less than 20mm diameter, 20% for nodules 20-29mm diameter, and 48% for nodules 30-40mm diameter. Establishing a specific diagnosis (of the many possible) for a benign nodule is more difficult; the sensitivities are lower for benign nodules of the same size.[3] Bronchoscopy was assumed to have a sensitivity of 0.5 for malignant nodal involvement.[4]

MEDIASTINOSCOPY

Sensitivity of mediastinoscopy for cancer in patients with enlarged lymph nodes is estimated at 0.92 (range, 0.88, 0.94),[2,5] and operative mortality is estimated at 0.3%.[5] Reflecting common practice of not initiating therapy without pathological proof of lung cancer, we assume perfect specificity for mediastinoscopy.[5]

TTNA

The sensitivities of trans-thoracic needle aspiration (TTNA) for malignancy and benign diagnoses were indexed by the size of the nodule, informed by literature estimates.[6,7,8]

VATS

VATS is assumed to have perfect accuracy at identifying malignant vs. benign disease and to include sampling or removal of nodes for confirmation of stage (perfect sensitivity and specificity). VATS had an operative mortality of 0.5%.[9]

METASTASES

A generic test for staging metastases (i.e., those not diagnosed on the basis of symptomatic presentation) is modeled with perfect specificity. Sensitivity for metastases was estimated during calibration; we tested functions of time since metastases developed and constant values of 0.4 to 0.5 derived from published sensitivities for bone scintigraphy and brain/bone MRI.[10,11]

OMITTED TESTS

We omitted sputum cytology due to its low sensitivity relative to bronchoscopy and positron emission tomography (PET) staging because it was uncommon during our calibration period (1990 to 2000).

REFERENCES:

[1] Beadsmoore, C. J., Screaton, N. J. "Classification, staging and prognosis of lung cancer" in European Journal of Radiology 2003; 45: 1: 8-17

[2] McLoud, TC, Bourgouin, PM, Greenberg, RW, Kosuik, JP, Templeton, PA, Shepard, JO, Moore, EH, Wain, JC, Mathesen, DJ, Grillo, HC "Bronchogenic carcinoma: Analysis of staging in the mediastinum with CT by correlative lymph node mapping and sampling" in Radiology 1992; 182: : 319-323

[3] Weinberger, SE "Differential diagnosis and evaluation of the solitary pulmonary nodule" in UpToDate 2004; version 12:

[4] Mandel, J, Weinberger, SE "Overview of non-small cell lung cancer staging" in UpToDate 2005; version 13:

[5] Esnaola, NF, Lazarides, SN, Mentzer, SJ, Kuntz, KM "Outcomes and cost-effectiveness of alternative staging strategies for non-small-cell lung cancer" in JCO 2002; 20: 1: 263-273

[6] Tan, BB, Flaherty, KR, Kazerooni, EA, Iannettoni, MD "The solitary pulmonary nodule" in Chest 2003; 123: 1: 89S-96S

[7] Li, H, Boiselle, PM, Shepard, JO, Trotman-Dickenson, B, McLoud, TC "Diagnostic accuracy and safety of CT-guided percutaneous needle aspiration biopsy of the lung: comparison of small and large pulmonary nodules" in American Journal of Radiology 1996; 167: : 105-109

[8] Wallace, MJ, Krishnamurthy, S, Broemeling, LD, Gupta, S, Ahrar, K, Morello, FA, Hicks, ME "CT-guided percutaneous fine-needle aspiration biopsy of small (less than or equal to 1-cm) pulmonary lesions" in Radiology 2002; 225: : 823-828

[9] Detterbeck, FC, Rivera, MP "Clinical presentation and diagnosis" in Diagnosis and Treatment of Lung Cancer 2001;

[10] Pope, RJE, Hansell, DM "Extra-thoracic staging of lung cancer" in European Journal of Radiology 2003; 45: 1: 31-38

[11] Earnest, Franklin, Ryu, Jay H., Miller, Gary M., Luetmer, Patrick H., Forstrom, Lee A., Burnett, Omer L., Rowland, Charles M., Swensen, Stephen J., Midthun, David E. "Suspected non–small cell lung cancer: Incidence of occult brain and skeletal metastases and effectiveness of imaging for detection—Pilot Study" in Radiology 1999; 211: 1: 137-145

PARAMETERS TREATMENT

Eligibility for Surgery

Individuals were randomly assigned as ineligible for surgical resection based on proportions of NSCLC stage I and II (all ages) cases that did not undergo surgery (where the reason was documented). Surgery was explicitly contraindicated for 5.6%, and offered but refused in 2.1% of cases. (Estimated from public release files using SEER*Stat 4.2.3 software.) We allow a small proportion (base case 13%, SEER-Medicare) of LS to be resected, reflecting the minority of cases which present with localized SCLC.

Resection

Effectiveness of resection is incorporated as follows: a person with no occult metastases whose single primary cancer is resected is assigned competing risks consistent with a person of the same smoking history – not stage I survival from SEER. However, if a second, undetected primary tumor remains (in a non-resected lobe), lung cancer can recur (see Natural History Component). The presence of undetected micro metastases is likely the cause of the poor observed survival after "curative" resection in many patients.[1] Removal of (or sampling from) nodes at resection can result in re-assigning stage at diagnosis, but provides no survival benefit.[2] The base case operative mortality rate for lobectomy is estimated at 4%[3] (value in sensitivity analysis, 3%). No increase in mortality due to late (post-30 day) effects of surgery (e.g., infection) was modeled.

Systemic Therapies

Parameter values that define efficacies of chemotherapy and radiotherapy are the probabilities of complete or partial response, using the definition of complete as no visible cancer at 4 week follow-up and partial as greater than or equal to a 30% decrease in diameter.[4] Probabilities of complete and partial responses vary by histologic type, with estimates taken from the literature. A cancer that partially responds to therapy is decreased in diameter by 30%, and a cancer that completely responds to therapy is reduced to 1.5mm diameter,[4] or below the 2mm detection threshold assumed for helical CT. Based on the new diameter, an adjusted 'time since cancer developed' is calculated and used to increment growth in all future cycles, retaining the original growth parameter alpha (Table Growth Parameters). To account for differences in growth rates of cancers pre- and post-therapy, we included parameters (estimated during calibration) that allowed faster-growing cancers to be more or less likely to respond to therapy.

Surveillance

Surveillance for recurrent lung cancer is modeled as helical CT at 6, 12, 24, 36, 48, and 60 months.[5]

REFERENCES:

[1] Spira, A, Ettinger, DS "Multidisciplinary management of lung cancer" in New England Journal of Medicine 2004; 350: : 379-392

[2] Smythe, W. R., American College of Chest Physicians "Treatment of stage I non-small cell lung carcinoma" in Chest 2003; 123: 1 Suppl.: 181S-187S

[3] Bach, PB, Cramer, LD, Schrag, D, Downey, RJ, Gelfand, SE, Begg, CB "The influence of hospital volume on survival after resection for lung cancer" in New England Journal of Medicine 2001; 345: : 181-188

[4] Therasse, P, Arbuck, SG, Eisenhauer, EA, Wanders, J, Kaplan, RS, Rubinstein, L, Verweij, J, Van Glabbeke, M, van Oosterom, AT, Christian, MC, Gwyther, SG "New guidelines to evaluate the response to treatment in solid tumors" in Journal of the National Cancer Institute 2000; 92: : 205-216

[5] Korst, Robert J., Gold, Heather T., Kent, Michael S., Port, Jeffrey L., Lee, Paul C., Altorki, Nasser K. "Surveillance computed tomography after complete resection for non-small cell lung cancer: Results and costs" in Journal of Thoracic and Cardiovascular Surgery 2005; 129: 3: 652-660

INCIDENTAL IMAGING COMPONENT

During each cycle spent in the general population, persons may undergo imaging exams (thoracic CT, or CXR) performed for reasons unrelated to screening for lung cancer.

We fit generalized linear models to insurance claims data from 1999. Increasing age predicted higher likelihood of both CT and CXR (p

Imaging results are compared to results of prior imaging exams, if available. Persons with no detected nodules or exclusively stable nodules return to the general population. Persons with newly detected nodules undergo follow-up and are managed according to the size of the largest nodule found. Persons with nodules large enough for biopsy start the next cycle in the Workup And Staging Component. If the new nodule is smaller than the threshold diameter, the person begins the following cycle in the Follow Up Component. A threshold diameter of 8mm was used as a proxy for clinical practice and is generally in agreement with a low (5%) biopsy rate for 4-9mm nodules in a recent trial.[1]

See Parameters Test Performance for details on test characteristics of imaging examinations.

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

REFERENCES:

[1] Pinsky, P. F., Marcus, P. M., Kramer, B. S., Freedman, M., Nath, H., Kvale, P., Reding, D. "Diagnostic procedures after a positive spiral computed tomography lung carcinoma screen" in Cancer 2005; 103: 1: 157-163

SYMPTOM DETECTION COMPONENT

Symptom detection can occur via symptoms from the primary cancer or from distant metastases.

Each month, individuals with distant metastases and/or a primary lung cancer (but not those with exclusively benign nodules) may develop symptoms that result in lung cancer detection and begin the following cycle in the Workup And Staging Component. The probability of symptom detection from primary cancers varies by location (central cancers have a greater propensity to cause symptoms, given size) and cell type (NSCLC vs. SCLC) and is a logistic function of the size of the largest cancer. We assume the minimum diameter for peripheral cancers to cause symptoms is 10mm, approximately the size at which airways are obstructed. The probability of symptom detection from metastases is a logistic function of the months since metastases developed (varied by NSCLC vs. SCLC).

Symptom detection parameters were estimated during calibration (see Calibration Details). The most relevant calibration targets included the proportion of lung cancers detected via symptomatic presentation, the stage and type distributions of incident cancers, and survival by type and stage at diagnosis. The background rate of chest imaging in the population (Incidental Imaging Component) will also influence the symptom detection rates.

See also the Assumption Overview.

BENIGN COMPONENT

The benign component occurs in every cycle of the model, so that benign lesions may develop throughout life. Benign lesions may also spontaneously resorb (e.g., due to clearing an infection).

To incorporate the costs and risks of follow-up procedures for benign lesions, the natural history model allows up to 3 benign lesions (with no regard for histological type) per person.

Overview

Using a polynomial fit to raw Mayo CT screening trial prevalence data, we estimated the average number of benign lesions per person, by age. Cumulative incidences of one or more benign lesions were converted to age-specific annual probabilities of developing new lesions (ages 35 to 68) and of existing lesions regressing (after age 68). For simulations of Mayo CT participants, therefore, prevalence of benign nodules was as observed in the study.

We estimate that 3% of benign nodules exhibit growth, and assign these lesions doubling times consistent with adenocarcinomas. Non-growing benign lesions are modeled as appearing fully formed in the previous month, consistent with a range of biological mechanisms (e.g., pleural effusion, edema, and infection).

Each benign lesion's location (i.e., specific lobe) was assigned based on a study of (n=185) nodules from the Mayo Clinic, which stratified by right/left. An indicator for central/peripheral is randomly assigned. Size (diameter) was derived from the Mayo Clinic data, expressed as a lognormal distribution (mean = 0.9, variance 0.36).

Assumptions

Solidly calcified lesions are not considered suspicious for lung cancer, and are ignored.

The probability of developing benign lesions is assumed independent of smoking history; no significant correlations were observed between any of the smoking factors and numbers of lesions in the Mayo CT data; and to our knowledge, no literature sources refute this observation.

Extrapolating from Mayo CT study data on prevalence of benign nodules

The base case LCPM incorporates no regional variation in the prevalence of benign nodules. Infection with histoplasmosis is a common cause of small (less than 3mm diameter) benign nodules.[1] Histoplasmosis rates vary geographically, with nearly 100% prevalence in persons residing in the major river valleys of the central U.S.[1] The Mayo Clinic (Rochester, MN) is not in an area of the highest histoplasmosis rates.

REFERENCES:

[1] Gurney, J., Dewey, J. "Pulmonary histoplasmosis" in Radiology 1996; 199: : 297-306

TABLE GROWTH PARAMETERS

Natural History Parameters

Distribution of alpha parameters (rate of decay of growth rate) used in the Gompertz equation for lung cancer growth, and doubling times (in days) for lung cancers at various sizes (cm diameter), by cell type.

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

Cell type	Distribution of alpha parameter	Mean (SD) DT at 0.5cm	Mean (SD) DT at 1.0cm	Mean (SD) DT at 1.5cm
Adenocarcinoma/BAC	logN(-7.765, 0.5504)	187(160)	227(194)	260(222)
Large cell	logN(-6.59942, 0.68862)	61(61)	74(74)	85(85)
Small cell	logN(-5.44357, 0.611485)	19(16)	23(20)	26(23)
Squamous cell	logN(-6.6111, 0.7935)	65(72)	79(87)	90(100)
Other	logN(-6.714, 0.6634)	67(66)	81(80)	93(92)

Notes: 'Other' refers to NSCLC not otherwise specified.

COMMENTS

The alpha parameters above are inputs for the LCPM. Thus the DTs shown are for all lung cancers at the specified size (*both diagnosed and undiagnosed*) and are **not** stratified by smoking history or stage.

See Parameters Natural History and Natural History Component for details.

PROTECTED HEALTH INFORMATION

The LCPM study's protocols for use of human subject data underwent expedited review (secondary use of medical records) and was approved by the human subjects institutional review board as compliant with HIPAA guidelines.

De-identified records (including demographics, smoking histories, and screen results) from two studies were provided to our institution for model calibration and validation.

- 1,520 current and former smokers aged 50-85 years participating in a Mayo Clinic study of annual CT screening for early detection of lung cancer[1]. Participants signed informed consent waivers approved by the Mayo Clinic institutional review board before enrollment in the screening study. Transfer of the de-identified data was approved by both institutions' human subjects review boards and was exempt from further informed consent requirements.

- 3,318 current and former smokers aged 55-74 years participating in the Lung Screening Study (LSS, a pilot study for the National Lung Screening Trial) of annual CT or CXR screening for early detection of lung cancer[2]. Transfer of the de-identified data was exempt from further informed consent requirements.

REFERENCES:

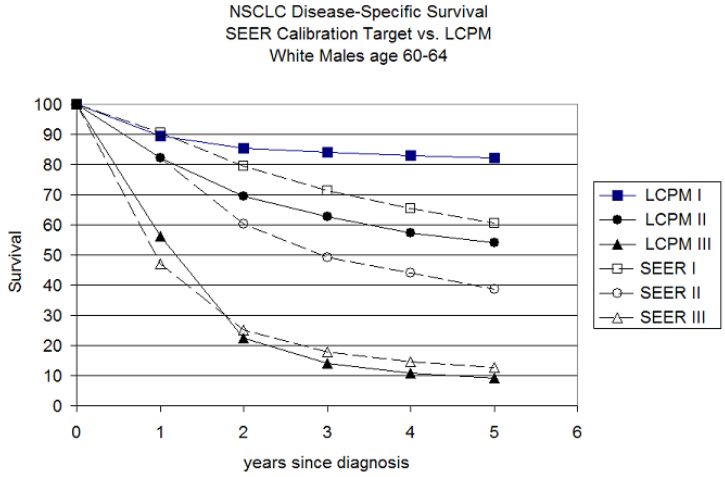
[1] Swensen, S., Jett, J., et al. "Screening for lung cancer with low-dose spiral computed tomography" in American Journal of Respiratory and Critical Care Medicine 2002; 165: : 508-513

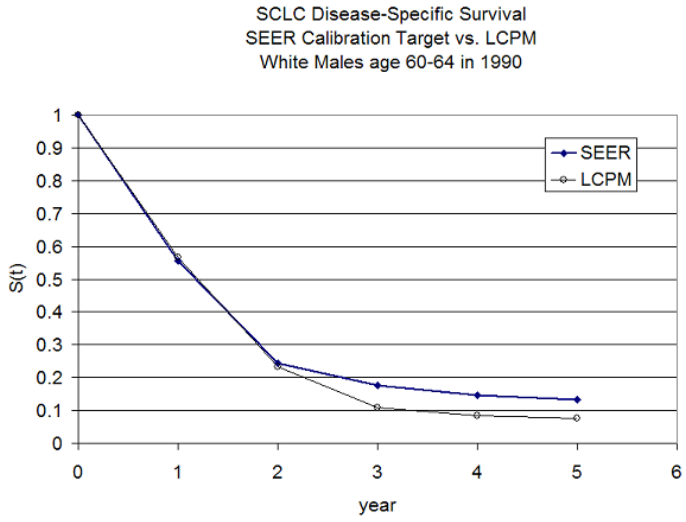
[2] Gohagan, J., Marcus, P., Fagerstrom, R., Pinsky, P., Kramer, B., Prorok, P., Writing Committee, Lung Screening Study Research Group "Baseline findings of a randomized feasibility trial of lung cancer screening with spiral CT scan vs chest radiograph: the Lung Screening Study of the National Cancer Institute" in Chest 2004; 126: 1: 114-21

CALIBRATION SURVIVAL1

CALIBRATION ENDPOINTS - SURVIVAL CURVES FOR NSCLC AND SCLC

Notes: Our use of published estimates for response rates from systemic therapies and our assumption that all patients underwent guideline staging and treatment may be irreconcilable with observed survival used for calibration targets, because many patients represented in SEER did not receive guideline treatments. Defining survival calibration targets that vary by treatment as well as stage, or by finer gradations of stage (i.e., T1N0M0 vs. T1N1M0) would address this limitation (ongoing work).

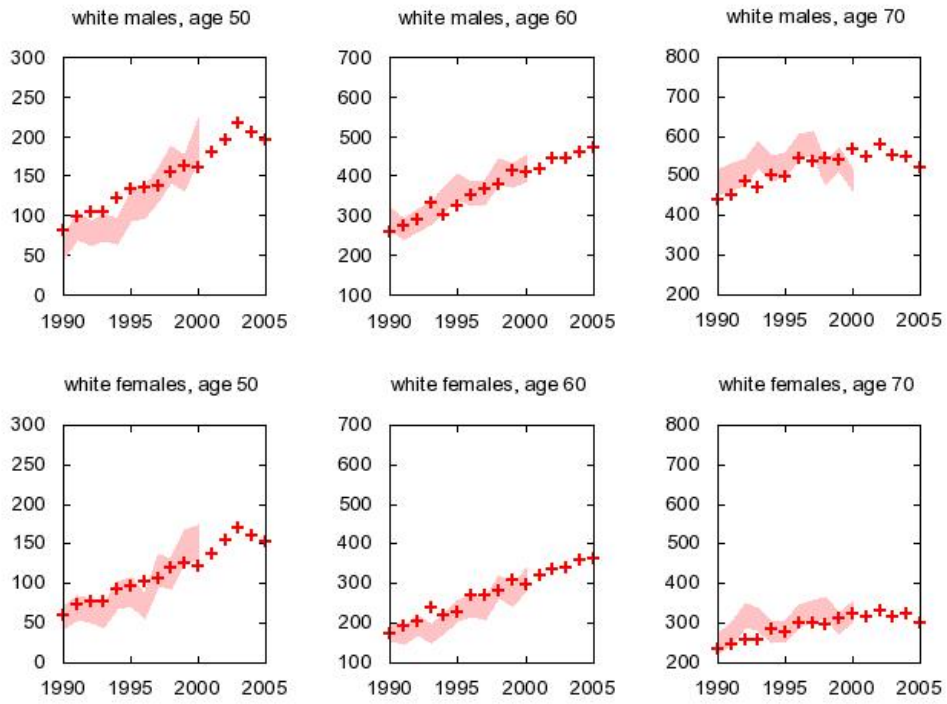




CALIBRATION INCIDENCE1

LUNG CANCER INCIDENCE FROM THE SINGLE COHORT LCPM VS. SEER

Age- (in single years), gender-, race-, and calendar year-specific lung cancer incidence rates, derived from SEER*Stat case listing files and counts of populations at risk from the NCI. Shown are incidence rates of all lung cancer types combined, for cohorts of whites.
Shaded regions are acceptance windows (95% CIs) around SEER calibration targets; crosses indicate LCPM output.



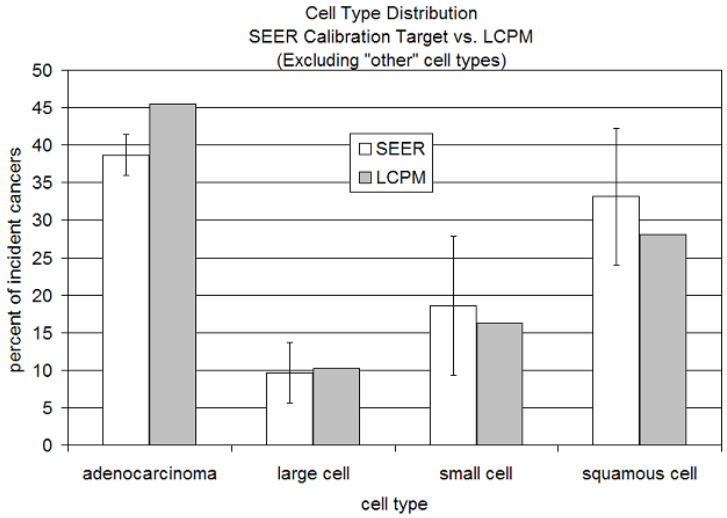
The Population LCPM (see Summary Of Versions for differences between the Population and single cohort LCPM models) was calibrated to the same reference cohort (60 year-old males).

Calibration USMortality documents calibration of period and cohort terms for the Population LCPM.

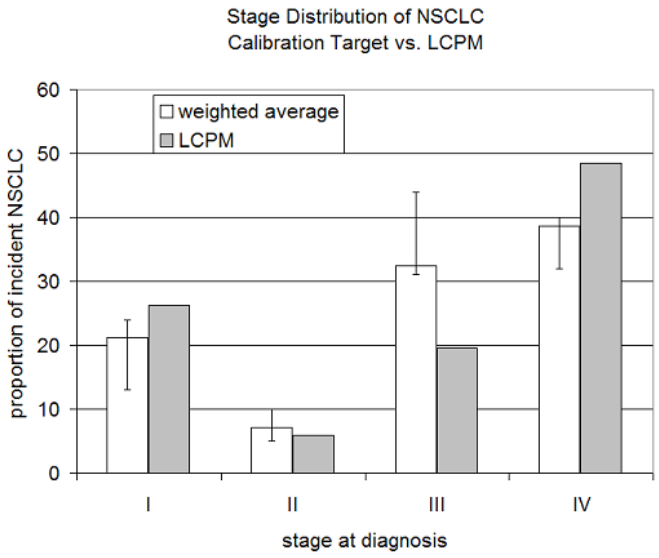
CALIBRATION SIZE TYPE STAGE1

SIZE, TYPE, AND STAGE DISTRIBUTIONS OF INCIDENT CANCERS; LCPM VS. SEER

The LCPM predicted a mean size of incident cancers of 29mm, vs. 38mm in SEER.



Calibration to type distribution of incident lung cancers, white males 60-70 years; 1990-2000.
Derived from SEER-Stat case listing files, stratified by gender, race, calendar decade, and 10-year age group.

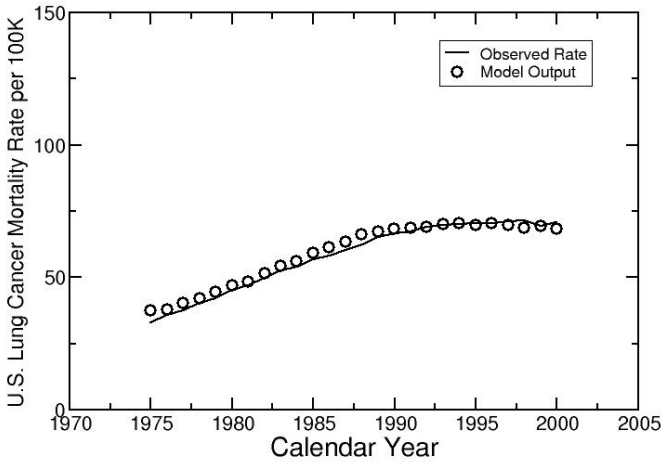


Calibration to stage distribution of incident lung cancers, white males 60-70 years; 1990-2000.

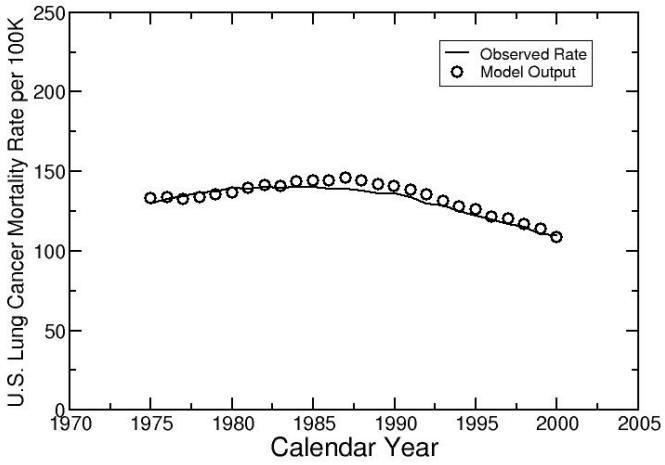
NB stratified by NSCLC and SCLC. Derived from a weighted average of 3 studies (all genders, all races) and SEER data (stratified by gender, race, decade, and age group).

CALIBRATION USMORTALITY1

Calibration plots from the Population LCPM.



Age-adjusted lung cancer mortality rates over the period 1975 to 2000. Model output vs. observed. All Races, Males



Age-adjusted lung cancer mortality rates over the period 1975 to 2000. Model output vs. observed. All Races, Females

VALIDATION LUNG SCREENING STUDY1

LCPM Populated with Lung Screening Study Population in Presence of Screening

Endpoint		Study result[2]	LCPM
participants with:	positive baseline CT screen	20.5%	20.9%
	lung cancer at baseline CT screen	1.9% (95% CI, 1.2%, 2.6)	1.2%
prevalent lung cancers that were:	adenocarcinoma	63% (n=16/30)	73.7%
	small cell	3%	3.1%
	NSCLC, NOS	10%	6.0%
prevalent lung cancers that were:	stage I	53% (n=16/30)	67.1%
	stage II	10%	8.0%
	stage III	20%	18.6%
	stage IV	10%	6.3%
	unstaged	7%	n/a
diameter of prevalent lung cancers:	median (mean, SD)	18mm (27, 23)	10mm (14.6, 8.4)
patients with prevalent lung cancers:	mean cigarettes/day (SD)	27.7 (9.3)	32.4
	mean years of smoking (SD)	47.3 (4.8)	50.3
	proportion male	0.57	0.58
	participants with lung cancer detected at screen #2	0.57% (n=8/1398)	0.29%

Notes: LCPM-predicted endpoints calculated from 250,000 simulated participants. Endpoints not provided in references were calculated directly from study data (see Protected Health Information). CI = confidence interval; SD = standard deviation; positive CT screen defined as detection of at least one non-solidly calcified pulmonary nodule at least 4mm in diameter. Retrospectively identified nodules not included in prevalence estimate. Adenocarcinoma includes bronchioloalveolar carcinoma (BAC) and mixed adenocarcinoma/BAC.

REFERENCES:

[1] Gohagan, J., Marcus, P., Fagerstrom, R., Pinsky, P., Kramer, B., Prorok, P., Writing Committee, Lung Screening Study Research Group "Baseline findings of a randomized feasibility trial of lung cancer screening with spiral CT scan vs chest radiograph: the Lung Screening Study of the National Cancer Institute" in Chest 2004; 126: 1: 114-21

[2] Gohagan, J. K., Marcus, P. M., Fagerstrom, R. M., et al., "Final results of the Lung Screening Study, a randomized feasibility study of spiral CT versus chest X-ray screening for lung cancer" in Lung Cancer 2005; 47: 1: 9-15

CALIBRATION METHODS RESEARCH

This note is for readers of *"Calibration of Disease Simulation Model Using an Engineering Approach"*,
Value in Health, Early View, February 2009
Chung Yin Kong, Pamela M. Mc Mahon, G. Scott Gazelle

Kong, et al. compare approaches for calibration of the LCPM using an earlier model version that included only 4 lung cancer cell types. Since completion of the work described in this article, we have added a 5th cell type ('Other', represented by ICD-O-2 code 80103) to both the Natural History Component and the calibration targets for age-specific incidence lung cancer incidence.

ABSTRACT

Objectives: Calibrating a disease simulation model's outputs to existing clinical data is vital to generate confidence in the model's predictive ability. Calibration involves two challenges: 1) defining a total goodness-of-fit score for multiple targets if simultaneous fitting is required; and 2) searching for the optimal parameter set that minimizes the total goodness-of-fit score (i.e., yields the best fit). To address these two prominent challenges, we have applied an engineering approach to calibrate a microsimulation model, the Lung Cancer Policy Model (LCPM).
Methods: First, eleven targets derived from clinical and epidemiological data were combined into a total goodness-of-fit score by a weighted-sum approach, accounting for the user-defined relative importance of the calibration targets. Second, two automated parameter search algorithms, Simulated Annealing (SA) and Genetic Algorithm (GA), were independently applied to a simultaneous search of 28 natural history parameters to minimize the total goodness-of-fit score. Algorithm performance metrics were defined for speed and model fit.
Results: Both search algorithms obtained total goodness-of-fit scores below 95 within 1,000 search iterations. Our results show that SA outperformed GA in locating a lower goodness-of-fit. After calibrating our LCPM, the predicted natural history of lung cancer was consistent with other mathematical models of lung cancer development.
Conclusion: An engineering-based calibration method was able to simultaneously fit LCPM output to multiple calibration targets, with the benefits of fast computational speed and reduced need for human input and its potential bias.

INDEX SUPPLEMENT CANCER

This document is intended as a guide to this Model Profiler for readers of "Adopting helical CT screening for lung cancer: Potential health consequences over a fifteen-year period" Mc Mahon PM, Kong CY, Weinstein MC, Tramontano AC, Cipriano LE, Johnson BE, Weeks JC, Gazelle GS.

C. 2008 Dec 15;113(12):3440-9.

PMID: 18988293 />

OVERVIEW

A description of the model structure, major components, and purpose is available in the Model Overview

INPUT PARAMETERS AND SOURCES

Test characteristics and mortality risks are described in Parameters Test Performance

Treatment effects and mortality risks are described in Parameters Treatment

Natural history parameters are described in Parameters Natural History

Smoking histories were derived from survey data as described in the Population Component

ASSUMPTIONS

The Assumption Overview describes the major assumptions underlying the LCPM and their possible implications.

Omissions from current LCPM that may influence the estimate of screening effectiveness:
Harms and benefits from incidental detection of other diseases (e.g., other cancers) found at screening

Increased lung cancer risks from radiation doses during screening or follow-up CT

examinations
(Brenner, D.J., Radiation risks potentially associated with low-dose CT screening of adult smokers for lung cancer. Radiology, 2004. 231(2): p. 440-5)

Tumor seeding of surgical or biopsy site, which is mainly described in case studies (eg Raja and Bessman, JCO 2003) and is thought to be a rare event

Explicit modeling of late surgical mortality
(Handy JR, Jr., Asaph JW, Skokan L, et al. Chest 2002; 122:21-30 and Toker, et al., Eur J Cardio-Thoracic Surg 2004;25:515-519)
In these studies, a small minority of 'late' (>30 day) post-resection deaths were due to late surgical mortality (1 infection/16 late deaths in Handy, et al., and 5 late surgical mortality death/51 total late deaths in Toker, et al.). A higher proportion of late mortality was due to cancer progression (7/16 in Handy and 16/51 in Toker), which the LCPM models explicitly as a cause of post-operative death. Other-cause deaths due to respiratory failure and heart disease are to a large extent captured by the increased competing mortality risks faced by smokers in the LCPM.

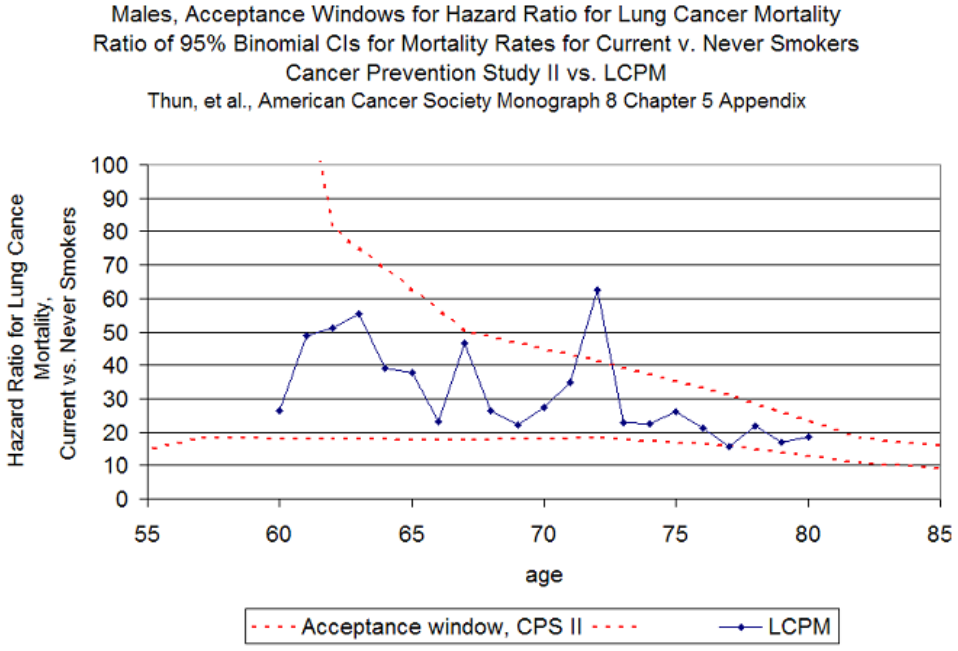
Economic consequences (costs) and influence of screening on quality of life (QALYs) were not considered in this analysis.

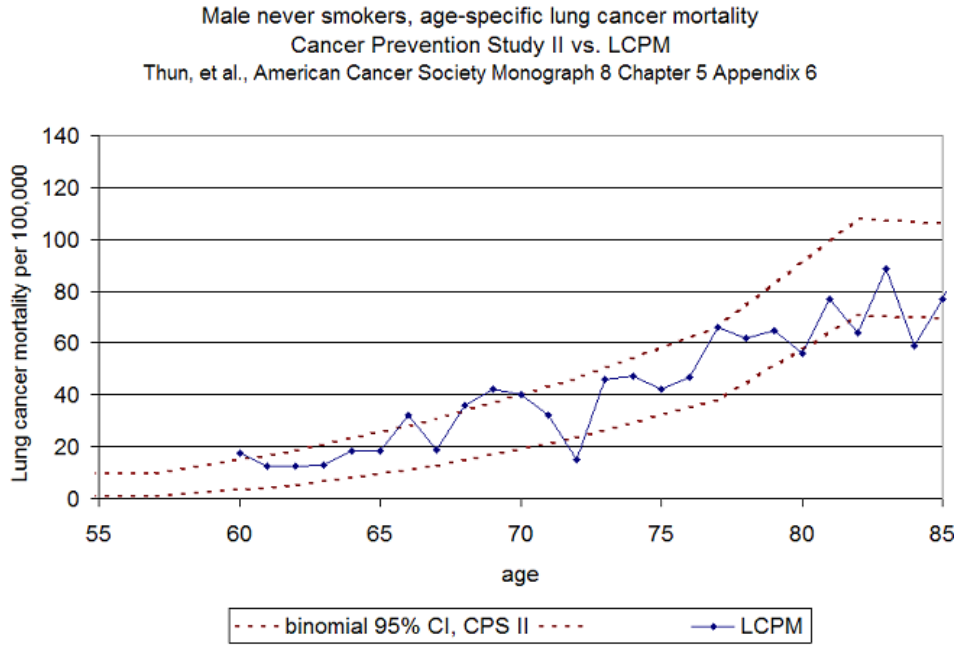
CALIBRATION COHORT STUDY1

SECONDARY CALIBRATION TARGETS - COHORT STUDY

Two targets were derived from an earlier (1980s) cohort study[1] with a highly selected population (i.e., volunteers with the American Cancer Society or their friends) that experienced only 70% of the all-cause mortality in the general U.S. population[2].

The LCPM predicted annual mortality rates per 100,000 non-smokers and lung cancer-specific mortality ratios for current (vs. never) smokers (by 5-year age group) that agree with observed data[1].





REFERENCES:

[1] Thun, M. J., Myers, D. G., et al. "Chapter 5. Age and the exposure-response relationships between cigarette smoking and premature death in Cancer Prevention Study II. National Cancer Institute, Smoking and Tobacco Control, Monograph 8:" 1997;

[2] Flanders, W. Dana, Lally, Cathy A., Zhu, Bao-Ping, Henley, S. Jane, Thun, Michael J. "Lung Cancer Mortality in Relation to Age, Duration of Smoking, and Daily Cigarette Consumption: Results from Cancer Prevention Study II" in Cancer Research 2003; 63: 19: 6556-6562

CALIBRATION NON SMOKERS AUTOPSY1

LUNG CANCER IN NON-SMOKERS

The LCPM predicted a percentage of non-smokers among lung cancer cases of 5.4%, in the range of reported values of 2% to 15%[1,2]. As expected, the model predicted a lower proportion of SCLC cases among non-smokers (4.3%) than among all lung cancer cases (18%)[3].

LUNG CANCER DETECTED AT AUTOPSY

Estimates of rates of undetected ("surprise") lung cancers at autopsy range from 0.34% to 55%[4,5,6,7,8,9,10,11,12]. Furthermore, autopsy techniques varied in unknown ways and there was no way to correct for potentially large biases due to unreported variability in age ranges and case mix (especially smoking prevalence and the selection bias inherent in autopsy series)[13]. Assuming that all lung cancers >15mm diameter would be diagnosed on autopsy, the LCPM (in the absence of screening) predicts a 3.6% autopsy surprise rate, in the reported range.

REFERENCES:

1 Capewell, S., Sankaran, R., et al. "Lung cancer in lifelong non-smokers" in Thorax 1991; 46: : 565-568

2 Beadsmoore, C. J., Screaton, N. J. "Classification, staging and prognosis of lung cancer" in European Journal of Radiology 2003; 45: 1: 8-17

3 Damber, L., Larsson, L.-G. "Smoking and lung cancer with special regard to type of smoking and type of cancer. A case-control study in north Sweden." in British Journal of Cancer 1986; 53: : 673-681

4 Gobbato, F, Vecchiet, F, Barbierato, D, Melato, M, Manconi, R "Inaccuracy of death certificate diagnosis in malignancy" in Human Pathology 1982; 13: 11: 1036-1038

5 Saracci, R. "Problems with the use of autopsy results as a yardstick in medical audit and epidemiology" in Quality Assurance in Health Care 1993; 5: 4: 339-44

6 Mollo, F, Bertoldo, E, Grandi, G, Cavallo, F "Reliability of death certificates for different types of cancer: an autopsy survey" in Pathology Research and Practice 1986; 181: : 442-447

7 Stenbäck, F, Päivärinta, H "Relation between clinical and autopsy diagnoses, especially as regards cancer" in Scandinavian Journal of Social Medicine 1980; 8: : 67-72

8 Engel, LW, Strauchen, JA, Chiazze Jr., L, Heid, M "Accuracy of death certification in an autopsied population with specific attention to malignant neoplasms and vascular disease" in American J Epidemiol 1980; 111: 1: 99-112

9 Delendi, M, Riboli, E, Peruzzo, P, Stanta, G, Cocchi, A, Gardiman, D, Sasco, AJ, Giarelli, L "Comparison of diagnoses of cancers of the respiratory system on death certificates and at autopsy" in Autopsy in Epidemiology and Medical Research, International Agency for Research on Cancer 1991;

10 Sternby, NH "The role of autopsy in cancer registration in Sweden, with particular reference to findings in Malmö" in Autopsy in Epidemiology and Medical Research, International Agency for Research on Cancer 1991;

[11] Chan, CK, Wells, CK, McFarlane, MJ, Feinstein, AR “More lung cancer but better survival: Implications of secular trends in "necropsy surprise" rates” in Chest 1989; 96: 2: 291-296

[12] Manser, R.L., Dodd, M., Byrnes, G., Irving, L. B., Campbell, D. “Incidental lung cancers identified at coronial autopsy: implications for overdiagnosis of lung cancer by screening” in Respir Med 2005; 99: : 501-507

[13] Agency for Healthcare Research and Quality, Department of Health and Human Services, “The autopsy as an outcome and performance measure” in Evidence Report/Technology Assessment No. 58 2002;

KEY REFERENCES

(1999) Marketscan Database,

Agency for Healthcare Research and Quality, Department of Health and Human Services, (2002) The autopsy as an outcome and performance measure in *Evidence Report/Technology Assessment No. 58*,

American Cancer Society (2001) Lung Cancer Resource Center in *American Cancer Society*,

American Cancer Society (2005) Cancer Facts and Figures in *American Cancer Society*,

Baba, Y., Takahashi, M., et al. (1998) Cost-effectiveness decision analysis of mass screening for lung cancer in *Academic Radiology* 5:2 Suppl, p S344-346

Bach, PB, Cramer, LD, Schrag, D, Downey, RJ, Gelfand, SE, Begg, CB (2001) The influence of hospital volume on survival after resection for lung cancer in *New England Journal of Medicine* 345, p 181-188

Bain, Chris, Feskanich, Diane, Speizer, Frank E., Thun, Michael, Hertzmark, Ellen, Rosner, Bernard A., Colditz, Graham A. (2004) Lung cancer rates in men and women with comparable histories of smoking in *J Natl Cancer Inst* 96:11, p 826-834

Barbone, F., Bovenzi, M., et al. (1997) Cigarette smoking and histologic type of lung cancer in men in *Chest* 112, p 1474-1479

Beadsmoore, C. J., Screaton, N. J. (2003) Classification, staging and prognosis of lung cancer in *European Journal of Radiology* 45:1, p 8-17

Beckles, M. A., Spiro, S. G., et al. (2003) Initial evaluation of the patient with lung cancer: symptoms, signs, laboratory tests, and paraneoplastic syndromes in *Chest* 123:1 Suppl, p 97S-104S

Benjamin, MS, Drucker, EA, McLoud, TC, Shepard, JO (2002) Small pulmonary nodules: detection at chest CT and outcome in *Radiology* 226, p 489-493

Black, W. (2000) Overdiagnosis: An underrecognized cause of confusion and harm in cancer screening [Editorials] in *Journal of the National Cancer Institute* 92:16, p 1280-1282

Capewell, S., Sankaran, R., et al. (1991) Lung cancer in lifelong non-smokers in *Thorax* 46, p 565-568

Caro, J., Klittich, W., et al. (2000) Could chest X-ray screening for lung cancer be cost-effective? in *Cancer* 89:11 Suppl, p 2502-2505

Centers for Disease Control (1993) Smoking Cessation During Previous Year Among Adults -- United States, 1990 and 1991 in *Morbidity and Mortality Weekly Report* 42:26, p 504-507

Chan, CK, Wells, CK, McFarlane, MJ, Feinstein, AR (1989) More lung cancer but better survival: Implications of secular trends in "necropsy surprise" rates in *Chest* 96:2, p 291-296

Chirikos, T., Hazelton, T., et al. (2002) Screening for lung cancer with CT: a preliminary cost-effectiveness analysis in *Chest* 121, p 1507-1514

Clare, S., Nakhlis, F., et al. (2000) Molecular biology of breast cancer metastasis: the use of mathematical models to determine relapse and to predict response to chemotherapy in breast cancer [Review] in *Breast Cancer Research* 2:6, p 430-435

Collins, M., Barry, M. (1996) Controversies in prostate cancer screening: Analogies to the early lung cancer screening debate. in *Journal of the American Medical Association* 276:24, p 1976-1979

Crestanello, J. A., Allen, M. S., Jett, J. R., Cassivi, S. D., Nichols, F. C., 3rd, Swensen, S. J., Deschamps, C., Pairolero, P. C. (2004) Thoracic surgical operations in patients enrolled in a computed tomographic screening trial in *Journal of Thoracic & Cardiovascular Surgery* 128:2, p 254-259

Damber, L., Larsson, L.-G. (1986) Smoking and lung cancer with special regard to type of smoking and type of cancer. A case-control study in north Sweden. in *British Journal of Cancer* 53, p 673-681

Delendi, M, Riboli, E, Peruzzo, P, Stanta, G, Cocchi, A, Gardiman, D, Sasco, AJ, Giarelli, L (1991) Comparison of diagnoses of cancers of the respiratory system on death certificates and at autopsy in *Autopsy in Epidemiology and Medical Research, International Agency for Research on Cancer*,

Detterbeck, FC, Rivera, MP (2001) Clinical presentation and diagnosis in *Diagnosis and Treatment of Lung Cancer*,

Diederich, S., Wormanns, D., et al. (2002) Screening for early lung cancer with low-dose spiral CT: Prevalence in 817 asymptomatic smokers. in *Radiology* 222, p 773-781

Earnest, Franklin, Ryu, Jay H., Miller, Gary M., Luetmer, Patrick H., Forstrom, Lee A., Burnett, Omer L., Rowland, Charles M., Swensen, Stephen J., Midthun, David E. (1999) Suspected non–small cell lung cancer: Incidence of occult brain and skeletal metastases and effectiveness of imaging for detection—Pilot Study in *Radiology* 211:1, p 137-145

Eddy, D. (1989) Screening for lung cancer. in *Annals of Internal Medicine* 111, p 232-237

Engel, LW, Strauchen, JA, Chiazze Jr., L, Heid, M (1980) Accuracy of death certification in an autopsied population with specific attention to malignant neoplasms and vascular disease in *American J Epidemiol* 111:1, p 99-112

Esnaola, NF, Lazarides, SN, Mentzer, SJ, Kuntz, KM (2002) Outcomes and cost-effectiveness of alternative staging strategies for non-small-cell lung cancer in *JCO* 20:1, p 263-273

Feig, S. A. (2005) Screening mammography controversies: resolved, partly resolved, and unresolved in *Breast Journal* 11:1 Suppl., p S3-6

Flanders, W. Dana, Lally, Cathy A., Zhu, Bao-Ping, Henley, S. Jane, Thun, Michael J. (2003) Lung Cancer Mortality in Relation to Age, Duration of Smoking, and Daily Cigarette Consumption: Results from Cancer Prevention Study II in *Cancer Research* 63:19, p 6556-6562

Gobbato, F, Vecchiet, F, Barbierato, D, Melato, M, Manconi, R (1982) Inaccuracy of death certificate diagnosis in malignancy in *Human Pathology* 13:11, p 1036-1038

Gohagan, J. K., Marcus, P. M., Fagerstrom, R. M., et al., (2005) Final results of the Lung Screening Study, a randomized feasibility study of spiral CT versus chest X-ray screening for lung cancer in *Lung Cancer* 47:1, p 9-15

Gohagan, J., Marcus, P., Fagerstrom, R., Pinsky, P., Kramer, B., Prorok, P., Writing Committee, Lung Screening Study Research Group (2004) Baseline findings of a randomized feasibility trial of lung cancer screening with spiral CT scan vs chest radiograph: the Lung Screening Study of the National Cancer Institute in *Chest* 126:1, p 114-21

Gurney, J., Dewey, J. (1996) Pulmonary histoplasmosis in *Radiology* 199, p 297-306

Gøtzsche, P. C., Nielsen, M. (2006) Screening for breast cancer with mammography in *Cochrane Database of Systematic Reviews*:4,

Gøtzsche, P., Olsen, O. (2000) Is screening for breast cancer with mammography justifiable? in *The Lancet* 355:9198, p 129-134

Harris, J. E. (1989) Trends in Smoking-Attributable Mortality, Chapter 3. Reducing the Health Consequences of Smoking, 25 Years of Progress: A Report of the Surgeon General., p 117-169

Hasegawa, M, Sone, S, Takashima, S, Li, F, Yang, Z-G, Maruyama, Y, Watanabe, T (2000) Growth rate of small lung cancers detected on mass CT screening in *British Journal of Radiology* 73, p 1252-1259

Health and Human Services (1997) The National Health and Nutrition Examination Survey III Data file 1988-1994. in *Public Use Data file Series 11*,

Health and Human Services (1990) The Health Benefits of Smoking Cessation,

Henschke, C., McCauley, D. , et al. (1999) Early Lung Cancer Action Project: Overall design and findings from baseline screening in *Lancet* 354:9173, p 99-105

Henschke, C., Naidich, D., et al. (2001) Early Lung Cancer Action Project: Initial findings on repeat screening in *Cancer* 92, p 153-159

Humphrey, L. L., Teutsch, S., et al. (2004) Lung cancer screening with sputum cytologic examination, chest radiography, and computed tomography: an update for the U.S. Preventive Services Task Force.[see comment] in *Annals of Internal Medicine* 140:9, p 740-53

Hunink, MGM, Goldman, L, Tosteson, ANA, Mittleman, MA, Goldman, PA, Williams, LW, Tsevat, J, Weinstein, MC (1997) The recent decline in mortality from coronary heart disease, 1980-1990 in *JAMA* 277:7, p 535-542

International Early Lung Cancer Action Program, I., Henschke, C. I., et al. (2006) Survival of patients with stage I lung cancer detected on CT screening.[see comment] in *New England Journal of Medicine* 355:17, p 1763-71

Jackman, DM, Chirieac, LR, Janne, PA (2006) Bronchioloalveolar carcinoma: A review of the epidemiology, pathology, and treatment. in *Sem Respir Crit Care Med* 26, p 342-352

Jemal, A, Chu, K, Tarone, R (2001) Cancer Surveillance Series: Recent trends in lung cancer mortality in the United States. in *J Natl Cancer Inst* 93, p 277-283

Kaneko, M., Kusumoto, M., et al. (2000) Computed tomography screening for lung carcinoma in Japan in *Cancer* 89:11 Suppl., p 2485-8

Knudsen, A. B., McMahon, P. M., Gazelle, G. S. (2007) Special Series (Cost): Use of modeling to evaluate the cost-effectiveness of cancer screening programs in *Journal of Clinical Oncology* 25:2, p 203-208

Korst, Robert J., Gold, Heather T., Kent, Michael S., Port, Jeffrey L., Lee, Paul C., Altorki, Nasser K. (2005) Surveillance computed tomography after complete resection for non-small cell lung cancer: Results and costs in *Journal of Thoracic and Cardiovascular Surgery* 129:3, p 652-660

Koscielny, S, Tubiana, M, Valleron, A-J (1985) A simulation model of the natural history of breast cancer in *British Journal of Cancer* 52, p 515-524

Kramer, B. S., J. K. Gohagan, et al. (1999) Cancer Screening: Theory and Practice. in *Basic and Clinical Oncology*,

Kubik, A. K., Parkin, D. M., et al. (2000) Czech Study on Lung Cancer Screening: post-trial follow-up of lung cancer deaths up to year 15 since enrollment in *Cancer* 89:11 Suppl., p 2363-8

Li, H, Boiselle, PM, Shepard, JO, Trotman-Dickenson, B, McLoud, TC (1996) Diagnostic accuracy and safety of CT-guided percutaneous needle aspiration biopsy of the lung: comparison of small and large pulmonary nodules in *American Journal of Radiology* 167, p 105-109

Mahadevia, P., Fleisher, L., et al. (2003) Lung cancer screening with helical computed tomography in older adult smokers in *Journal of the American Medical Association* 289:3, p 313-322

Mandel, J, Weinberger, SE (2005) Overview of non-small cell lung cancer staging in *UpToDate* version 13,

Manser, R.L., Dodd, M., Byrnes, G., Irving, L. B., Campbell, D. (2005) Incidental lung cancers identified at coronial autopsy: implications for overdiagnosis of lung cancer by screening in *Respir Med* 99, p 501-507

Marcus, P. M., Bergstralh, E. J., et al. (2006) Extended lung cancer incidence follow-up in the Mayo Lung Project and overdiagnosis.[see comment] in *Journal of the National Cancer Institute* 98:11, p 748-56

Marcus, P., Bergstralh, E., et al. (2000) Lung cancer mortality in the Mayo Lung Project: Impact of extended follow-up in *Journal of the National Cancer Institute* 92:16, p 1308-1316

Marshall, D., Simpson, K., et al. (2001) Potential cost-effectiveness of one-time screening for lung cancer (LC) in a high risk cohort in *Lung Cancer* 32:3, p 227-236

Marshall, D., Simpson, K., et al. (2001) Economic decision analysis model of screening for lung cancer in *European Journal of Cancer* 37:14, p 1759-1767

Massey, JT, Moore, TF, Parsons, VL, Tadros, W (1989) Design and Estimation for the National Health Interview Survey, 1985-94 in *Vital Health Statistics, Series 2, No. 110,*

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

McLoud, TC, Bourgouin, PM, Greenberg, RW, Kosuik, JP, Templeton, PA, Shepard, JO, Moore, EH, Wain, JC, Mathesen, DJ, Grillo, HC (1992) Bronchogenic carcinoma: Analysis of staging in the mediastinum with CT by correlative lymph node mapping and sampling in *Radiology* 182, p 319-323

McMahon, P. M., Zaslavsky, A. M., et al. (2006) Estimation of mortality rates for disease simulation models using Bayesian evidence synthesis in *Medical Decision Making* 26, p 497-511

Miettinen, O., Henschke, C. (2001) Commentary on Drs. Patz, Black, and Goodman's viewpoint in *Radiology* 221, p 598-599

Miettinen, O., Henschke, C. (2001) CT screening for lung cancer: coping with nihilistic recommendations [Viewpoint] in *Radiology* 221, p 592-596

Miller, D. P., Liu, G., et al. (2002) Combinations of the variant genotypes of GSTP1, GSTM1, and p53 are associated with an increased lung cancer risk in *Cancer Research* 62, p 2819-2823

Mizuno, T., Masaoka, A., et al. (1984) Comparison of actual survivorship after treatment with survivorship predicted by actual tumor-volume doubling time from tumor diameter at first observation in *Cancer* 53, p 2716-2720

Mollo, F, Bertoldo, E, Grandi, G, Cavallo, F (1986) Reliability of death certificates for different types of cancer: an autopsy survey in *Pathology Research and Practice* 181, p 442-447

Moolgavkar, S. H., Knudson, A. G., Jr. (1981) Mutation and cancer: a model for human carcinogenesis in *J Natl Cancer Inst* 66:6, p 1037-52

Mountain, C. F. (1997) Revisions in the International System for Staging Lung Cancer.[see comment] in *Chest* 111:6, p 1710-7

Naruke, T., Goya, T., et al. (1988) The importance of surgery to non-small cell carcinoma of lung with mediastinal lymph node metastasis in *Annals of Thoracic Surgery* 46:6, p 603-10

Nathan, M., Collins, V., et al. (1962) Differentiation of benign and malignant pulmonary nodules by growth rate in *Radiology* 79, p 221-232

National Cancer Institute (2002) National Lung Screening Trial,

National Cancer Institute (2002) Database: Incidence - SEER 9 Regs Public-Use,

National Center for Health Statistics National Health Interview Survey,

Oreskes, N., Shrader-Frechette, K., et al. (1994) Verification, validation, and confirmation of numerical models in the earth sciences in *Science* 263, p 641-646

Patz, E., Black, W., et al. (2001) Commentary on Drs. Miettinen and Henschke's viewpoint in *Radiology* 221, p 597

Patz, E., Black, W., et al. (2001) CT screening for lung cancer: Not ready for routine practice [Viewpoint] in *Radiology* 221, p 587-591

Patz, E., Goodman, P., et al. (2000) Screening for lung cancer [Review] in *New England Journal of Medicine* 343:22, p 1627-1633

Pinsky, P. F., Marcus, P. M., Kramer, B. S., Freedman, M., Nath, H., Kvale, P., Reding, D. (2005) Diagnostic procedures after a positive spiral computed tomography lung carcinoma screen in *Cancer* 103:1, p 157-163

Pope, RJE, Hansell, DM (2003) Extra-thoracic staging of lung cancer in *European Journal of Radiology* 45:1, p 31-38

Press, W. H., Teukolsky, S. A., et al. (2002) Numerical Recipes in C++: The Art of Scientific Computing,

Prorok, P., Andriole, G., et al. (2000) Design of the prostate, lung, colorectal and ovarian (PLCO) cancer screening trial in *Controlled Clinical Trials* 21, p 273S-309S

Puntoni, R., Toninelli, F., Zhankui, L., Bonassi, S. (1995) Mathematical modelling in risk/exposure assessment of tobacco related lung cancer in *Carcinogenesis* 16:7, p 1465-71

Rachet, B., Siemiatycki, J., Abrahamowicz, M., Leffondre, K. (2004) A flexible modeling approach to estimating the component effects of smoking behavior on lung cancer in *J Clin Epidemiol* 57:10, p 1076-85

Ramsey, S. D., McIntosh, M., et al. (2000) Simulation modeling of outcomes and cost effectiveness in *Hematology - Oncology Clinics of North America* 14:4, p 925-38

Ransohoff, D. F., Harris, R. P. (1997) Lessons from the mammography screening controversy: can we improve the debate? in *Annals of Internal Medicine.* 127:11, p 1029-34

Raz, D J, He, B, Rosell, R, Jablons, D M (2005) Current concepts in bronchioloalveolar carcinoma biology in *Clin Cancer Res* 12, p 3698-3704

Reich, J. M. (2002) Improved survival and higher mortality: the conundrum of lung cancer screening.[see comment]. in *Chest* 122:1, p 329-37

Ries, L., Eisner, M., et al. (2002) SEER Cancer Statistics Review, 1973-1999,

Rivera, M., Detterbeck, F., et al. (2001) Epidemiology and classification of lung cancer. Diagnosis and Treatment of Lung Cancer: An evidence-based guide for the practicing clinician
,

Rubins, JB, Rubins, HB (1996) Temporal trends in the prevalence of malignancy in resected solitary pulmonary lesions in *Chest* 109, p 100-103

Saracci, R. (1993) Problems with the use of autopsy results as a yardstick in medical audit and epidemiology in *Quality Assurance in Health Care* 5:4, p 339-44

Schilling, Freimut H., Spix, Claudia, Berthold, Frank, Erttmann, Rudolf, Fehse, Natalja, Hero, Barbara, Klein, Gisela, Sander, Johannes, Schwarz, Kerstin, Treuner, Joern, Zorn, Ulrich, Michaelis, Joerg (2002) Neuroblastoma screening at one year of age in *N Engl J Med* 346:14, p 1047-1053

Smythe, W. R., American College of Chest Physicians (2003) Treatment of stage I non-small cell lung carcinoma in *Chest* 123:1 Suppl., p 181S-187S

Sone, S., Li, F., et al. (2001) Results of three-year mass screening programme for lung cancer using mobile low-dose spiral computed tomography scanner in *British Journal of Cancer* 84:1, p 25-32

Song, P, Sekhon, HA, Jia, Y, Keller, JA, Blusztajn, JK, Mark, GP, Spindel, ER (2003) Acetylcholine is synthesized by and acts as an autocrine growth factor for small cell lung carcinoma in *Cancer Research* 63, p 214-221

Sox, H. (2002) Screening mammography for younger women: Back to basics in *Annals of Internal Medicine* 137:5 (Part 1), p 361-362

Spira, A, Ettinger, DS (2004) Multidisciplinary management of lung cancer in *New England Journal of Medicine* 350, p 379-392

Spratt, J. S., Spratt, J. A. (2002) Growth Rates in *Cancer of the Breast*,

Stenbäck, F, Päivärinta, H (1980) Relation between clinical and autopsy diagnoses, especially as regards cancer in *Scandinavian Journal of Social Medicine* 8, p 67-72

Sternby, NH (1991) The role of autopsy in cancer registration in Sweden, with particular reference to findings in Malmö in *Autopsy in Epidemiology and Medical Research, International Agency for Research on Cancer*,

Strauss, G., Gleason, R., et al. (1997) Screening for lung cancer: Another look; A different view in *Chest* 111:3, p 754-768

Surveillance Research Program, National Cancer Institute, (2005) Surveillance, Epidemiology, and End Results (SEER) Program Seer*Stat Database: Incidence - SEER 13 Regs Public-Use, Nov 2004 Sub (1973-2002 varying) in *Seer*Stat software version 6.1.4*,

Swensen, S. J., Jett, J. R., et al. (2005) CT Screening for Lung Cancer: Five-year Prospective Experience in *Radiology* 235:1, p 259-265

Swensen, S., Jett, J., et al. (2002) Screening for lung cancer with low-dose spiral computed tomography in *American Journal of Respiratory and Critical Care Medicine* 165, p 508-513

Swensen, S., Jett, J., et al. (2003) Lung cancer screening with CT: Mayo Clinic experience in *Radiology* 226:3, p 756-761

Swensen, S., Viggiano, R., et al. (2000) Lung nodule enhancement at CT: multicenter study in *Radiology* 214, p 73-80

Takashima, S., Sone, S., et al. (2003) Indeterminate solitary pulmonary nodules revealed at population-based CT screening of the lung: Using first follow-up diagnostic CT to differentiate benign and malignant lesions in *American Journal of Radiology* 180, p 1255-1263

Tan, BB, Flaherty, KR, Kazerooni, EA, Iannettoni, MD (2003) The solitary pulmonary nodule in *Chest* 123:1, p 89S-96S

Tan, W. Y., Zhang, L.-J. (2004) Stochastic modeling of carcinogenesis: state space models and estimation of parameters in *Discrete and Continuous Dynamical Systems - Series B* 4:1, p 297-322

Taylor, K. L., Shelby, R., et al. (2004) Quality of Life and Trial Adherence Among Participants in the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial in *J Natl Cancer Inst* 96:14, p 1083-1094

The Commit Research Group (1995) Community Intervention Trial for Smoking Cessation (COMMIT): II. Changes in Adult Cigarette Smoking Prevalence in *Am J Public Health* 85, p 193-200

Therasse, P, Arbuck, SG, Eisenhauer, EA, Wanders, J, Kaplan, RS, Rubinstein, L, Verweij, J, Van Glabbeke, M, van Oosterom, AT, Christian, MC, Gwyther, SG (2000) New guidelines to evaluate the response to treatment in solid tumors in *Journal of the National Cancer Institute* 92, p 205-216

Thun, M. J., Myers, D. G., et al. (1997) Chapter 5. Age and the exposure-response relationships between cigarette smoking and premature death in Cancer Prevention Study II. National Cancer Institute, Smoking and Tobacco Control, Monograph 8:,

Travis, W D, Garg, K, Franklin, W A, et al. (2005) Evolving concepts in the pathology and computed tomography imaging of lung adenocarcinoma and bronchioloalveolar carcinoma in *J Clin Oncol* 23, p 3279-3287

U.S. Department of Health and Human Services, National Center for Health Statistics, (1997) The National Health and Nutrition Examination Survey III Data file 1988-1994. in *Public Use Data file Series 11*,

Usuda, K., Saito, Y., et al. (1994) Tumor doubling time and prognostic assessment of patients with primary lung cancer in *Cancer* 74, p 2239-2244

Wakelee, HA, Chang, ET, Gomez, SL, et al. (2007) Lung cancer incidence in never smokers in *J Clin Oncol* 25, p 472-478

Wallace, MJ, Krishnamurthy, S, Broemeling, LD, Gupta, S, Ahrar, K, Morello, FA, Hicks, ME (2002) CT-guided percutaneous fine-needle aspiration biopsy of small (less than or equal to 1-cm) pulmonary lesions in *Radiology* 225, p 823-828

Weinberger, SE (2004) Differential diagnosis and evaluation of the solitary pulmonary nodule in *UpToDate* version 12,

Weinstein, M. C., O'Brien, B., et al. (2003) Principles of good practice for decision analytic modeling in health-care evaluation: Report of the ISPOR Task Force on Good Research Practices-Modeling Studies in *Value in Health* 6:1, p 9-17

Weinstein, MC, Coxson, PG, Williams, LW, Pass, TM, Stason, WB, Goldman, L (1987) Forecasting coronary heart disease incidence, mortality, and cost: The Coronary Heart Disease Policy Model in *American Journal of Public Health* 77, p 1417-1426

Winton, T., Livingston, R., et al. (2005) Vinorelbine plus cisplatin vs. observation in resected non-small-cell lung cancer in *New England Journal of Medicine* 352:25, p 2589-2597

Wisnivesky, J., Mushlin, A., et al. (2003) The cost-effectiveness of low-dose CT screening for lung cancer in *Chest* 124, p 614-621

Woods, William G., Gao, Ru-Nie, Shuster, Jonathan J., Robison, Leslie L., Bernstein, Mark, Weitzman, Sheila, Bunin, Greta, Levy, Isra, Brossard, Josee, Dougherty, Geoffrey, Tuchman, Mendel, Lemieux, Bernard (2002) Screening of infants and mortality due to neuroblastoma in *N Engl J Med* 346:14, p 1041-1046

Zheng, T, Holford, TR, Boyle, P, et al. (1994) Time trend and the age-period-cohort effect on the incidence of histologic types of lung cancer in Connecticut, 1960-1989 in *Cancer* 74, p 1556-1567

FLEXKB DOCUMENT
Version: HI.001.03302011.70021
Document generated: 03/30/2011

YALE UNIVERSITY

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.

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

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

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.

MODEL PURPOSE

SUMMARY

This document provides a brief overview of the Yale lung cancer model for population rates. A carcinogenesis model based on a mixture of never, current and former smokers is used to provide estimates of rates in a specified age and calendar year. In order to estimate the number of lung cancer deaths that would be expected to occur in a population with a specified cigarette smoking history, the model also introduces scale and temporal calibration that includes age, period and cohort effects.

PURPOSE

This population based model provides estimates of trends in lung cancer mortality rates using quantitative formulae derived from analytical epidemiology studies for the effect of cigarette smoking.

Our model expands the age-period-cohort (APC) temporal framework in order to discover the manner in which population trends in factors affecting lung cancer mortality can affect cancer rates. The model incorporates available population based data on cigarette smoking in order to quantify its effect on observed trends in lung cancer mortality and to evaluate the impact of changes in smoking patterns on lung cancer rates.

The APC framework offers a useful way of conceptualizing temporal trends. Age represents the effect of the degenerative process on disease risk that takes place over a lifetime. Period and cohort, on the other hand, are likely to reflect changes in the exposure to important risk factors or in the disease surveillance system. For lung cancer, period effects on trend are likely to be factors that affect the entire population regardless of age, e.g., air pollution, screening, tobacco control, modification in the manufacture of cigarettes, or artifactual changes in diagnostic technology. On the other hand, cohort effects would arise from generational changes in behavior, such as promotional campaigns for cigarette smoking directed at men enlisted during World War II and women baby boomers seeking equality in gender rights. Including this temporal calibration factor in the model provides additional detail into how well the model is able to describe existing trends.

While analytical epidemiologic studies offer the best way to estimate the effect of putative risk factors on disease risk, quantitative descriptions of the way in which changes in exposure can affect population rates can be much more challenging. The purpose of this model is to incorporate temporal trends in cigarette smoking, a known powerful risk factor for lung cancer incidence, into a quantitative description of the observed trends in lung cancer mortality rates. The model will then be used to estimate the effect of interventions designed to change risk factor exposure on disease rates.

MODEL OVERVIEW

SUMMARY

This document describes previous work leading to this model and the model itself in general terms.

PURPOSE

The Yale model extends the age-period-cohort model for lung cancer mortality trends by including cigarette smoking data for the population. We employ the two-stage clonal expansion (TSCE) model as fitted by FHLUNG to data from the Health Professionals Follow-up Study for men and the Nurses' Health Study for women. The primary objective of this approach is to provide quantitative estimates of the impact of smoking on the population at large, thus enabling one to estimate the impact of changes in exposure to this important risk factor on lung cancer mortality rates. Inclusion of age, period and cohort effects, allows one to determine the extent to which these temporal trends are explained by a carcinogenesis model using available data on cigarette consumption, and to determine which temporal factors are not well characterized. Partitioning the elements of goodness-of-fit into these more readily understood temporal effects, helps one to identify the limitations of a model. In addition, it provides an approach for introducing an additional calibration for the missing temporal effects.

The approach that is described here can be readily extended to include alternative carcinogenesis models. This will similarly provide an approach for calibrating aspects of the age, period and cohort effects that are not well characterized by the model, as well as giving diagnostic detail on how well the carcinogenesis model describes population trends. Comparing these summaries of goodness-of-fit can suggest models that agree more closely with observed trends. Reasons for lack of fit can be due either to limitations of the carcinogenesis model itself, or the quality of the exposure information for the population.

BACKGROUND

This model uses results from analytical epidemiology studies that quantify the effects of age, level and duration of smoking, and smoking cessation on lung cancer mortality rates. This is accomplished by extending the age-period-cohort model to include these results, and provide further adjustment for limitations that may arise from errors in survey data or the model that quantifies the relationship between smoking history and lung cancer mortality risk. This model can be easily modified to incorporate alternative carcinogenesis models.

For more detail, please see:

1. Descriptive Epidemiology Of Lung Cancer
2. Age-Period-Cohort Models
3. Models For The Effect Of Age On Lung Cancer Incidence
4. Exposure Models For The Effect Of Cigarette Smoking On Population Rates

Model Description

The Yale lung cancer mortality model considers the population to be a mixture of never, current and former smokers with known prevalences p_0, p_1 and p_2 respectively. For each of these groups, the TSCE model estimate for lung cancer mortality is determined as a function of summaries of the smoking history, which are known. An overall rate is determined by a average of the rates in each smoking category using the smoking prevalences as weights.

Summaries of smoking history for the population are estimated using cohort summaries of smoking initiation rates, smoking cessation rates and number of cigarettes smoked. These fundamental parameters were included in the SHG, which was run many times to simulate the experience of the overall population. Relevant average values provided estimates of smoking history summaries required by the TSCE model.

Calibration is used to correct for discrepancies that result from direct use of the carcinogenesis model. Let $\mathbf{t} = (a, b, c)$ represent temporal elements: age (a), period (p) and cohort ($c = p - a$), respectively. Details on exposure to cigarette smoking history in the population at a particular time is given by the vector $\mathbf{Z}(\mathbf{t})$. A carcinogenesis model provides an estimate of the mortality rate as a function of the population smoking exposure data, $\lambda\{\mathbf{Z}(\mathbf{t})\}$. We calibrate the estimated rates from a carcinogenesis model by introducing a multiplicative factor

$\lambda^*\{\mathbf{Z}(\mathbf{t}); \mathbf{t}\} = \theta(\mathbf{t})\, \lambda\{\mathbf{Z}(\mathbf{t})\}$.

Contributors

Theodore R. Holford

ASSUMPTION OVERVIEW

SUMMARY

This document discusses assumptions underlying the model as well as some of their implications.

BACKGROUND

This aggregate model for lung cancer mortality rates assumes that the population represents a mixture of individuals with different levels of risk that depends on smoking history. An underlying carcinogenesis model is employed to determine the effect of cigarette smoking history on lung cancer mortality. In addition, we include a multiplicative calibration function that depends on age, period and birth cohort that yields rates that agree well with the observed rates in the population, thus adjusting for underlying differences in overall health that may exist between individuals in an analytical study and those in the population as a whole.

The carcinogenesis model uses results from fitting the two stage clonal expansion (TSCE) model to the Health Professionals Follow-up Study (HPFS) for males and the Nurses Health Study (NHS) for females, which provides estimates of mortality rates for a given age of initiation, age of cessation and number of cigarette smoked per day. Prevalence of never smokers, current smokers and former smokers are estimated using initiation and cessation rates derived from the National Health Interview Survey. The survey also provided estimates of number of cigarettes smoked per day for the population. An overall rate for the population was determined by assuming there was a mixture of smoking levels using the estimated proportion in a given smoking category and the corresponding estimated mortality rate.

The age, period and cohort (APC) log linear model for rates provided the framework for determining the calibration factor that yielded estimated rates that correspond to those observed in the population. It is well known that an APC model provides an excellent description of temporal trends in mortality from cancer of the lung and bronchus. When these temporal elements are included as nominal factors in the calibration function, a nonparametric form for each component is implied. Thus, an analysis of the resulting estimate of each temporal component reveals which aspect of trend has not been adequately characterized when the model and the corresponding smoking history summaries are used to describe population rates. Correspondingly, by including the estimated parameters from this into the estimated population rates, we obtain estimates of rates and number of cases that correspond to the observed values.

ASSUMPTION LISTING

1. The TSCE natural history model developed by the FHLUNG group for lung cancer was assumed to apply for the population rates. Parameters used in this component of the model were obtained by fitting to HPFS and NHS data for males and females respectively. Further details on this model is provided in the section on Project 2 for the FHLUNG group.

2. A common multiplicative calibration function that applied to all smoking categories was employed. This was assumed to be a log-linear function of age, period and cohort, each of which being entered as a nominal variable, resulting in a nonparametric representation of each temporal factor.

3. The distribution of the number of lung cancer deaths in the US was assumed to have a Poisson distribution with additional random error that is proportional to the mean. A quasi-likelihood method of inference was employed. Maximum likelihood estimates of the temporal calibration factors were obtained using PROC GENMOD in SAS.

4. Estimates of the distribution of the population in the various smoking categories were obtained by running the SHG simulator many times. This provided not only estimates of prevalences among the broad categories of never, current and former smokers, but the distribution of time quit in former smokers and mean number of cigarettes per day by quintile of dose.

5. The TSCE model includes a contribution for age, but an additional term was included in the calibration to allow for limitations in the carcinogenesis model.

6. The period effect in the calibration can not only allow for limitations in the TSCE model and the available data on smoking history, but other factors that are not available. For example, cigarette manufacturing changes affecting lethality are not included in the model. In addition, a period effect could represent data artifact, which may result from changes in lung cancer mortality definitions or technology that may not represent changes in risk.

7. The cohort effect in the calibration can allow for corresponding aspects of trend that are not well characterized by the TSCE model or the corresponding smoking histories. These may include generational changes in smoking behavior, for example.

PARAMETER OVERVIEW

SUMMARY

This document provides an overview of the major parameters in the model, their sources, and general implications they have on model outputs.

BACKGROUND

The age, period and cohort framework for describing temporal trends in disease rates has provided a useful approach in the descriptive epidemiology of many cancer sites, including the lung and bronchus. A model with only these temporal elements assumes a nonparametric form for each component, thus allowing the form for the relationship to be revealed in the analysis. However, underlying causes driving trends for each of these temporal factors depend on biological processes and exposure trends. This model brings together the classical age-period-cohort (APC) model and a theoretical model for the effects of age and cigarette smoking on lung cancer mortality. The reasons for bringing together these two approaches are:

1. It provides a means for evaluating the adequacy of the theoretical model in explaining the temporal elements of age, period and cohort. An ideal model would not leave any systematic departure from trend in the temporal elements. However, if a systematic age calibration is required then this would imply that that model does not provide a good description of the aging effect on lung cancer risk.

2. It yields calibrated estimates of rates using the estimated temporal departures from the carcinogenesis model.

Data required for the implementation of this model are obtained from demographic and vital statistics summaries, as well as analysis of survey data on exposure of the population to cigarette smoking. In particular, the model relies on:

1. Population vital statistics

- Estimates for the population at risk were provided by the NCI through the SEER*Stat software that may be accessed on the web site;
- Number of lung cancer death from 1975-2000 were provided by the NCI and are generally available on the SEER*Stat web site.

2. Smoking histories

The observed smoking parameters that arose from a population which experienced some tobacco control were derived from estimates of smoking initiation and cessation rates obtain in the NCHS Health Interview Survey. Summaries were provided by five year birth cohorts, single years of age and gender for smoking initiation rates, quit rates and mean number of cigarettes smoked by quintile. These parameters were added to the smoking history generators, which was repeatedly invoked in order to simulate the experience of a population with the specified characteristics.

Parameter Listing Overview

1. Smoking history

The observed smoking parameters that arose from a population which experienced some tobacco control were derived from estimates of smoking initiation and cessation rates obtain in the NCHS National Health Interview Survey. History summaries derived from the survey were provided by five year birth cohorts, single years of age and gender. Specific details available were (a) smoking initiation rates, (b) quit rates and (c) mean number of cigarettes smoked by quintile. These parameters were added to the smoking history generator, which was repeatedly invoked in order to simulate the experience of a population with the specified characteristics.

Similar summaries were generated to represent hypothetical populations in which there was no tobacco control or complete control following publication of the Surgeon General's Report in 1964. The scenarios considered were:

- Actual Tobacco control (ATC)—the observed experience in the US;
- No tobacco control (NTC)—the experience that would have been expected if the smoking histories observed before 1955 had continued unabated in subsequent years; and,
- Complete tobacco control (CTC)—all smoking ceased in 1965.

The simulation results from the smoking history generator provided summary parameters for the model by single years of age and cohort for:

- Never smokers
  1. prevalence of individuals who never smoked;
- Current smokers
  1. prevalence of current smokers
  2. mean age of smoking initiation
  3. mean number of cigarettes smoked;
- Former smokers who had quit 1-2, 3-5, 6-10, 11-15 and 16 or more years
  1. prevalence of former smoker categories
  2. mean age of smoking initiation

3. mean number of cigarettes smoked
4. mean duration of smoking.

2. Effects on mortality

Moolgavkar et al (Moolgavkar 1979; Moolgavkar 1988; Moolgavkar and Luebeck 1990; Luebeck and Moolgavkar 2002) proposed a two-stage clonal expansion (TSCE) model for lung cancer. Estimates of the underlying parameters provided by the FHLUNG group were obtained by fitting to data from the Health Professionals Follow-up Study for men and the Nurses Health Study for women. (For details on how these model parameters were derived, see the FHCRC site.) Our model considers the population to be a mixture of never, current and former smokers.

3. Population calibration

Temporal and scalar calibration of lung cancer mortality rates derived from the TSCE model were obtained by finding quasi maximum likelihood estimates of the age, period and cohort model parameters using (a) data on the population at risk provided by the NCI through the SEER*Stat software that may be accessed on the web site, and (b) number of lung cancer deaths from 1975-2000 provided by the NCI.

COMPONENT OVERVIEW

SUMMARY

The components of this model are described in this section. The first component is used to describe the smoking history of the population under alternative scenarios. These are then used as parameters in a carcinogenesis model to determine mortality rates. The final component aligns model result with those observed in the population through calibration.

OVERVIEW

The age-period-cohort (APC) model has been used to systematically explore cancer incidence trends in Connecticut (Roush 1985; Roush 1987). Included in this effort were several attempts to model lung cancer incidence trends, first by considering separate dummy variables for the temporal effects (Zheng, Holford et al. 1994) and then developing more detailed algebraic expressions that considered specific models for the effects of age, period and cohort (Stevens and Moolgavkar 1979; Stevens and Moolgavkar 1984; Holford, Zhang et al. 1994; Holford, Zhang et al. 1996). Among the models used for the effect of age are the multistage or Armitage and Doll model (Armitage and Doll 1954; Stevens and Moolgavkar 1979) and the two stage clonal expansion model (Moolgavkar 1979; Moolgavkar 1988; Luebeck and Moolgavkar 2002). Other aspects of lung cancer trends can include exposure data gleaned from surveys along with the effect on mortality or incidence derived from relevant cohorts, such as the British Doctors' Study (Doll and Hill 1964) and the follow-up of cohorts generated by the American Cancer Society (Hammond 1966; Knoke, Shanks et al. 2004).

The population is broken into never, current and former smoking categorizes. Because of heterogeneity of the rates within these categories, they are further subdivided by level of smoking for current smokers and years quit for former smokers. The TSCE model used by the FHLUNG Group provided estimates of the rates for each smoking category and a lengthy simulation using SHG provided estimates of the distribution within each smoking category. Using the estimated proportions of the population in a smoking group as weights, the weighted sum of the corresponding rates provided an estimate of the overall rate for the population. These values are the rates estimated under the assumption that the population used to generate the model parameters corresponds to the US population, i.e., the estimated rate.

Calibration of the estimated rates is accomplished by estimating a multiplicative factor for each rate derived from the TSCE model. This was determined by fitting a Poisson regression models in which the calibration factor is a log-linear age-period-cohort model using Poisson regression with the observed number of lung cancer deaths in the US population as the response. The estimated number of lung cancer deaths was used to determine the calibrated estimates of lung cancer mortality.

COMPONENT LISTING

1. SHG was used to determine the population distribution for the various smoking categories.

2. TSCE model with parameters estimated by fitting to data from the HPFS and NHS for males and females respectively was used to determine the underlying lung cancer mortality rates.

3. A weighted sum of the rates estimated for each smoking category provided an overall estimate of the mortality rate for lung cancer. It was assumed that a log-linear age-period-cohort model was appropriate and maximum likelihood estimates of the parameters were found by fitting a model using PROC GENMOD in SAS.

4. Estimated rates under alternative tobacco control strategies was determined by first finding the estimated rate using the TSCE model, then multiplying by the calibration factors obtained under actual scenario.

SMOKING HISTORY GENERATOR COMPONENT

SUMMARY

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 lung cancer consortium.

OVERVIEW

The core SHG software was parameterized using three tobacco control scenarios to 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-

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**

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**	**CTC**
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 (1890-1984)	User Defined		
Gender (Male/Female)	User Defined		
Race (All race)	User Defined		

[1] Cigarettes smoked per day,[2]Other 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

Parameter	Valid Values
Seed value for PRNG used for Initiation, Cessation, OCD[1], Smoking intensity quintile	Integer from -1 to 2147483647 (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, Smoking intensity quintile and CPD[4] data files	As derived from NHIS depending on the scenario

[1]Other 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. [3]Key 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. [4]Cigarettes 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 History	Smoking intensity quintile (5 quintiles ranging from light to heavy smoking), Yearly smoking dose (CPD[2])

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

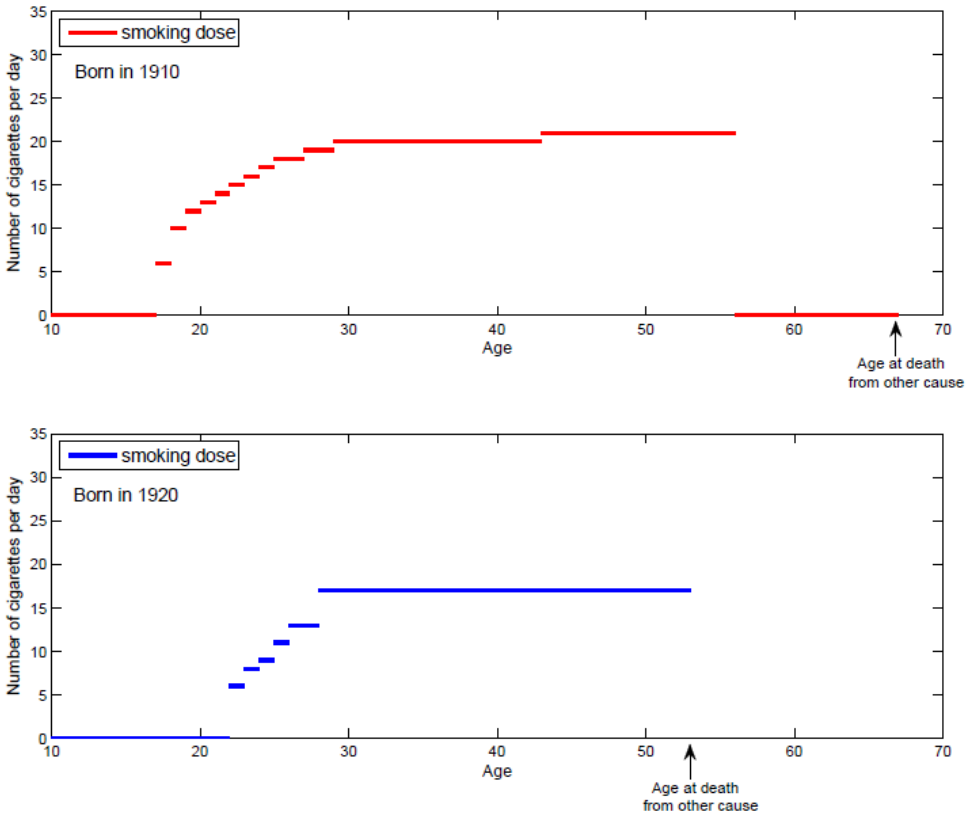


Figure SHG-1: Examples of the SHG-Generated Events

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 mortality rates are modeled.

Overview

This model regards the population as a mixture of never, current and former smokers with prevalences p_0, p_1 and p_2 respectively. The overall lung cancer mortality rate for population with this mixture of smoking histories is given by

$$\bar{\lambda} = p_0\lambda_0(t) + p_1\lambda_1(t, \bar{z}_1, \bar{t}_{1i}) + p_2\lambda_2(t, \bar{z}_2, \bar{t}_{2i}, \bar{t}_q)$$

where $\lambda_0(\cdot)$, $\lambda_1(\cdot)$ and $\lambda_2(\cdot)$ are the corresponding rates for each smoking category. Other parameters in the model are age (t), mean number of cigarettes smoked ($\bar{z}$), mean age of smoking initiation ($\bar{t}_i$), and mean age quit smoking ($\bar{t}_q$). The rate in former smokers was broken down further in order to improve accuracy in summarizing the mixture of smoking durations that were expected to occur in older age groups. The categories of smoking durations were:

1. 1-2 years
2. 3-5 years
3. 6-10 years
4. 11-15 years
5. 16 years or more

In category j, mean dose ($\bar{z}_{2j}$), mean age of initiation ($\bar{t}_{2j,i}$), mean age quit ($\bar{t}_{j,q}$) and proportion of the population who were former smokers in the category (p_{2j}) were determined, and the overall rate among former smokers was given by

$$\bar{\lambda}_2 = \sum\nolimits_j p_{2j}\lambda_{2j}(t, \bar{z}_{2j}, \bar{t}_{2j,i}, \bar{t}_{j,q})$$

The summary data on smoking history for the population were generated by running the smoking history generator many times and reporting the mean values for the parameters of interest (provided by Jihyoun Jeon and Rafael Meza).
The two stage clonal expansion (TSCE) model was employed to quantify the effects of cigarette smoking on lung cancer mortality rates. In each case, the data from HPFS and NHS were used to estimate the model parameters for males and females respectively. Temporal calibration of the model was accomplished by introducing a multiplicative factor that is a function of age (a), period (p) and cohort ($c = p - a$),

$$\bar{\lambda} = \theta(a, p, c)\left[p_0\lambda_0(t) + p_1\lambda_1(t, \bar{z}_1, \bar{t}_{1i}) + p_2\lambda_2(t, \bar{z}_2, \bar{t}_{2i}, \bar{t}_q)\right]$$

where

$$\theta(a, p, c) = \exp\{\mu + \alpha_a + \pi_p + \gamma_c\}$$

The intercept, μ, scales the rates so that the estimates from the TSCE model correspond overall with those observed in the US population. Temporal elements for age (α_a), period (π_p) and cohort (γ_c) provide correspondingly calibrated temporal elements missed by the carcinogenesis model in describing observed trends for the population as a whole. If temporal effects are all 0, then the model is in good agreement with the population. If, on the other hand, these effects are not parallel to the abscissa then that would indicate inadequacy of the carcinogenesis model in being able to characterize that particular aspect of temporal trend in population rates. Poor agreement could be the result of either a limitation in the carcinogenesis model itself or in the population estimates of exposure to relevant risk factors.

TWO STAGE CLONAL EXPANSION MODEL

Moolgavkar et al (Moolgavkar 1979; Moolgavkar 1988; Moolgavkar and Luebeck 1990; Luebeck and Moolgavkar 2002) proposed the TSCE model in which the carcinogenesis process is initiated in a cell that then multiplies to form a clone. A second hit on one of these initiated cells transform it into a cancer cell that subsequently multiplies further until it forms a tissue mass that can be clinically identified as cancer. The functional form for this model is complex and details are provided in the work of Moolgavkar et al., but it has been found to provide an excellent description of the effect of age on lung cancer incidence and mortality. Parameters required by the model were estimated using HPFS for males and NHS for females as described in the section for FHCRC.

OUTPUT OVERVIEW

SUMMARY

This document describes the output that is generated by the Yale lung cancer mortality model.

OVERVIEW

The Yale model uses the TSCE model with parameters estimated using data from HPFS and NHS for the effect of cigarette smoking in males and females respectively. It can be readily modified to consider alternative carcinogenesis models, but we shall limit the discussion in this document to the TSCE model. Estimates of the parameters are described in further detail in by the FHLUNG group.

Parameters estimated for the calibration function are produced, thus providing a diagnostic summary of the adequacy of the model in describing population rates. The intercept provides for a scale shift in the estimated rates. A perfect carcinogenesis model with completely accurate smoking history information would be expected to produce estimates of age, period and cohort effects that are zero. Effects that are not parallel to the horizontal axis suggest temporal aspects of the model that not well characterized. In addition, it provides significance tests for the departure of the carcinogenesis model from population data, and estimates of the proportion of the temporal trend explained by the model.

Estimated rates generated by the model provide calibrated estimates of the population rates. By changing the model parametrization, one can produce alternative calibration strategies, including a full APC calibration or one that only uses subsets of the temporal effects. The summaries include not only estimates of the rates, but also estimated numbers of lung cancer deaths.

Result are provided for the observed smoking experience in the US in which there was some tobacco control (ATC). We also produce estimates of age-specific rates and number of lung cancer deaths under scenarios in which there was no tobacco control (NTC) or complete tobacco control following production of the 1964 Surgeon General's Report (CTC).

OUTPUT LISTING

1. Diagnostic information about the adequacy with which the model characterizes population rates:

(a) estimates of age, period and cohort parameters;
(b) significance tests for the calibration parameters; and,
(c) summaries of the amount of temporal variation explained by the model.

2. Summaries of calibrated expected age-specific lung cancer mortality rates (i.e., lung cancer mortality hazard) for individuals with a specified smoking history. Parameters that may be specified are:

(a) year of birth;
(b) age started smoking;
(c) number of cigarettes smoked per day; and,
(d) age quit smoking.

3. Calibrated age-specific population mortality rates and estimated numbers of lung cancer deaths for a population with a mixture of smoking risks. Distributions to be specified include:

(a) age start smoking
(b) number of cigarettes smoked per day; and,
(c) time since quit smoking.

4. Calibrated age-specific population mortality rates and estimated numbers of lung cancer deaths for a population with a mixture of smoking risks resulting from a tobacco control strategy. In this case, the manner in which tobacco control affects the parameters in SHG are specified (initiation rates, quit rates and cigarettes smoked per day), and these are used to generate the smoking history distribution specified in 3. The particular scenarios presented in this analysis are:

(a) actual tobacco control (ATC);
(b) no tobacco control (NTC); and,
(c) complete tobacco control (CTC).

RESULTS OVERVIEW

SUMMARY

This document provides a summary of selected results obtained in the analysis of US lung cancer mortality rates in males and females. Three tobacco control strategies were considered: (a) actual tobacco control experience in the US (ATC); (b) no tobacco control (NTC); and, (c) complete tobacco control (CTC) following production of the Surgeon General's Report in 1964.

OVERVIEW

Scenarios

Figure 1(a and b) shows calibrated age-specific mortality rates for specified smoking history scenarios derived from the TSCE models with parameters estimated from HPFS males and NHS females. The rates for nonsmokers are considerably lower than the smokers and they increase with age. Two ages at smoking initiation of 20 cigarettes per day were considered, 14 and 25. Both age initiation groups were divided into hypothetical groups who either continued to smoke or quit at age 35. Finally, doses of 10, 20 and 40 cigarettes per day were considered for those who begin smoking at 25 and quit at 35.

Calibration and Validation

An overall summary of the calibration parameters determined by model fitting using PROC GENMOD in SAS are shown in Table 1. Deviance, G^2, is often interpreted as a likelihood goodness of fit statistics, but these data suggest the presence of random error not accounted for by the Poisson distribution that is usually employed for count data. A model with extra-Poisson variation was employed in this summary, making use of a quasi-likelihood method. This results in the use of F-tests for the significance of the individual effects. Linear trends are not estimable, so the resulting summaries only consider curvature for each temporal effect.

A comparison of the calibrated age-specific lung cancer mortality rates from this ATC model with the observed is shown in Figures 2.

A summary of the temporal calibration effects are shown in Figure 3. Figure 3(a) shows the estimated age effects for men and women using the TSCE model and the model with no carcinogenesis contribution included, using the constraint of zero slope for period in order to resolve the identifiability problem for APC models. For ages over 50 the effects are flat, suggesting that the model provides a good summary of age trends for females, although the declining trend shows the need for a correction that decreases for the older age groups, i.e., the model tends to overestimate the rates compared to younger ages. The decline is greater for males. Period effects, shown in Figure 3(b) employ the same scale as the other temporal effects to allow comparison of magnitude of calibration, and these are constrained to have zero slopes to achieve a unique set of estimates. A clear pattern is apparent, but the effects are small. Finally, the estimated cohort effects using the constraint for period are shown in Figure 3(c). It is important to recognize that the estimates for the most recent cohorts are determined from as few as a single rate in the youngest age groups, resulting in considerably less precision. It is also apparent that the TSCE model that includes smoking history data has explained much of the existing cohort trend but not all of it, especially for early cohorts.

Results List

Table 1. Summary of curvature effects and fit for models giving deviance chi-square tests (G^2), F-tests (P

		Male			Female		
Source	**df**	**G^2**	**F-test[1]**	**% explained**	**G^2**	**F-test[1]**	**% explained**
Smoking Model							
Age	53	7511.8	141.73	89.54	7598.9	143.37	73.78
Period	24	577.1	24.05	51.59	494.8	20.61	67.79
Cohort	78	4018.7	24.05	68.15	5808.0	74.46	74.61
Goodness of Fit	1272	1830.0			1554.8		
Scale estimate	1272	1.44			1.22		
No Model							
Age	53	71844.6	1355.56	-	28984.9	546.89	-
Period	24	1192.0	49.67	-	1536.0	64.00	-
Cohort	78	12618.4	161.77	-	22874.4	293.26	-
Goodness of Fit	1272	1710.8			1555.5		
Scale estimate	1272	1.35			1.22		

[1]F-tests are used because of extra-Poisson variation with numerator df shown on the row and denominator df given for the estimate of scale.

Table 2. Estimated number of lung cancer deaths under the Tobacco Control, No Tobacco Control and Complete Tobacco Control by gender.

Calibration Approach	**Actual Tobacco Control**	**No Tobacco Control**	**Complete Tobacco Control**
Constant Calibration			
Males	2,067,778	2,608,186	1,056,518
Females	1,051,980	1,250,552	480,375
APC Calibration			

Males	2,067,775	2,670,897	958,862
Females	1,051,978	1,273,151	438,857

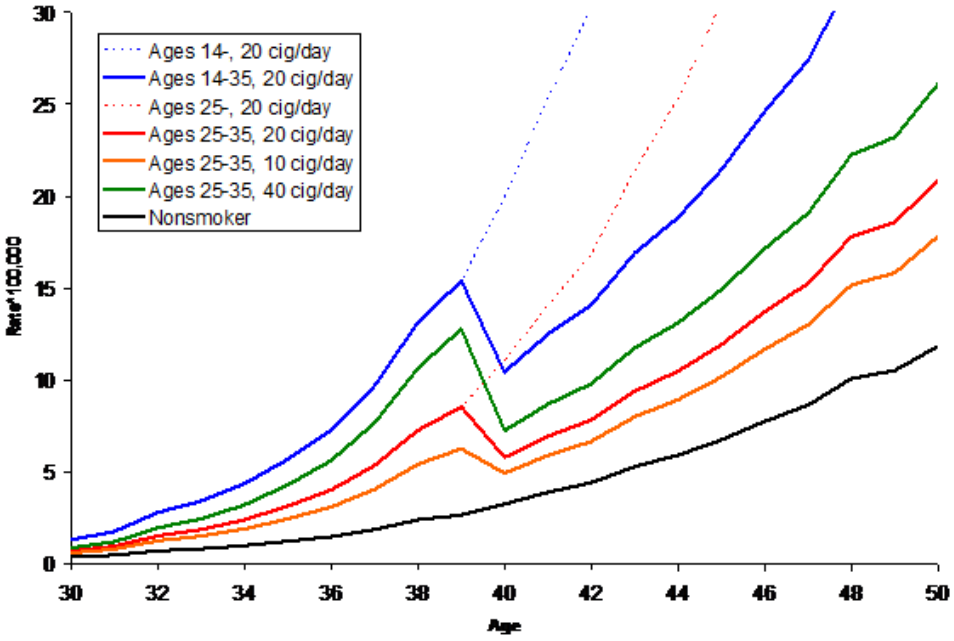


Figure 1(a). Age trends in male lung cancer rates in the HPFS TSCE model starting age 14 or 25, quitting at 35 or never, and smoking 10, 20 or 40 cigarettes/day.

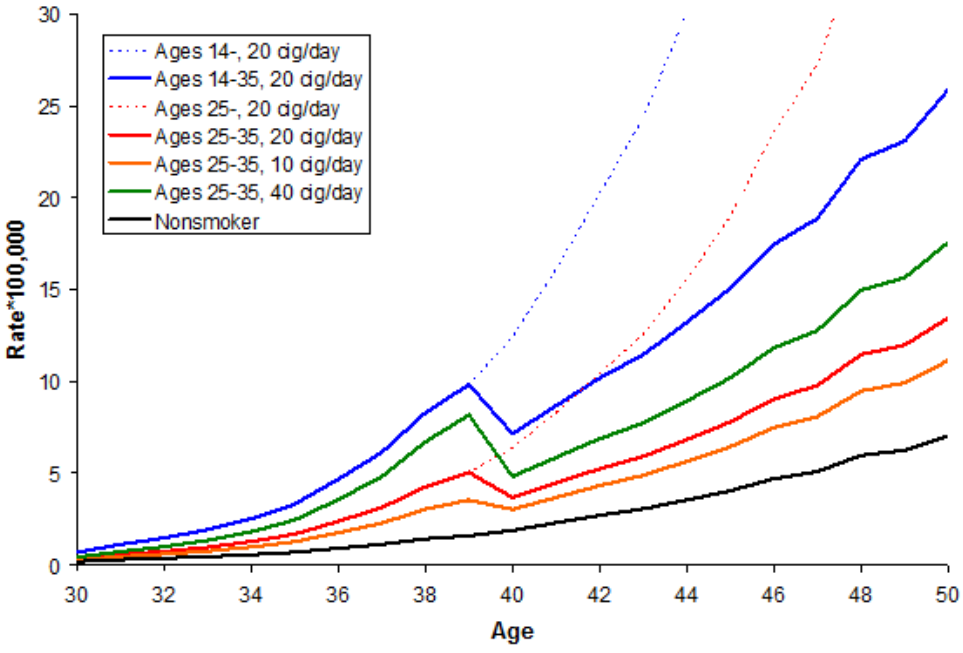


Figure 1(b). APC calibrated age trends in female lung cancer rates in the NHS TSCE model starting age 14 or 25, quitting at 35 or never, and smoking 10, 20 or 40 cigarettes/day.

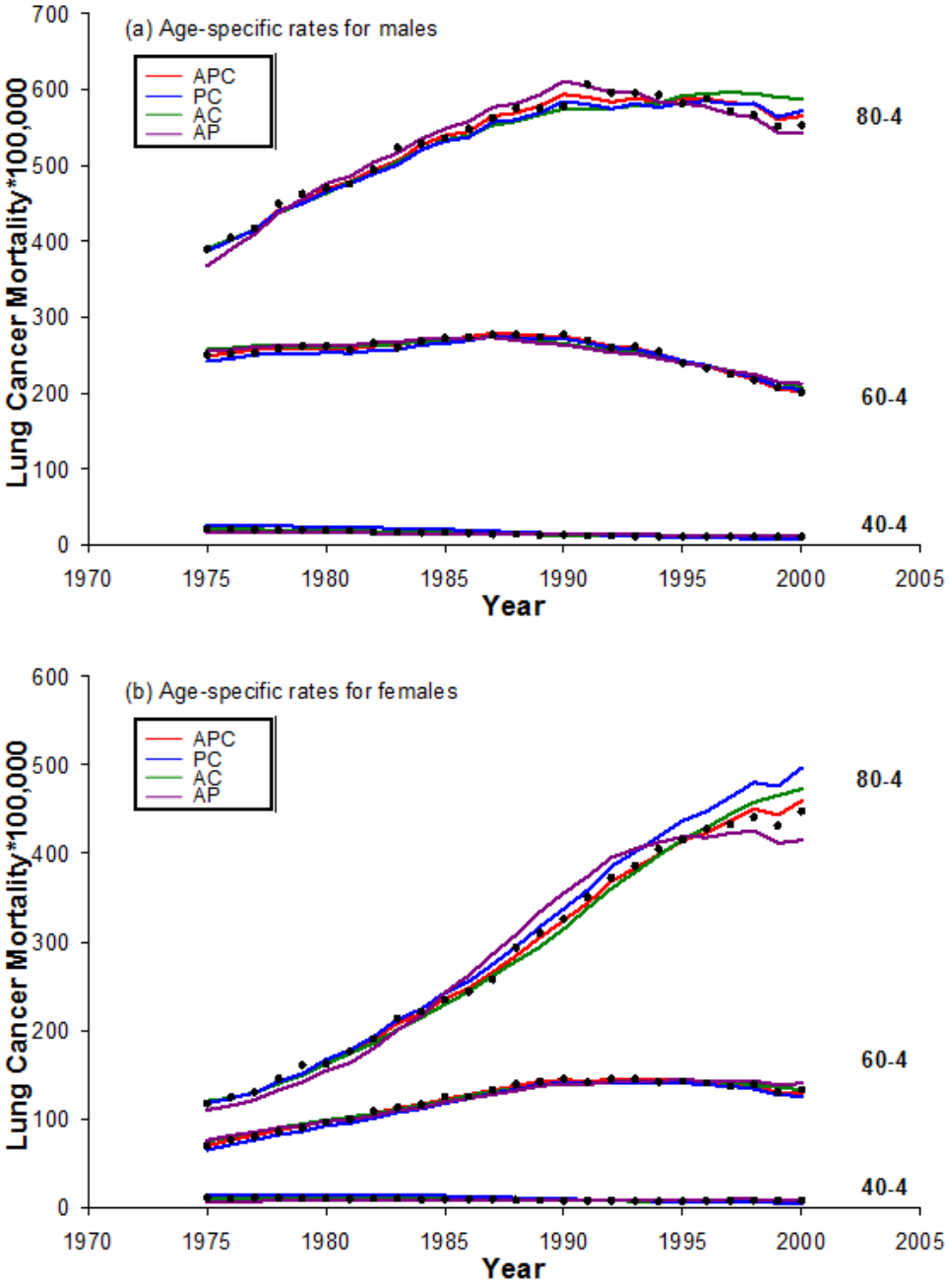


Figure 2. Observed (dots) and calibrated (APC, PC, AC, and AP) rates (solid lines) for selected age groups by gender.

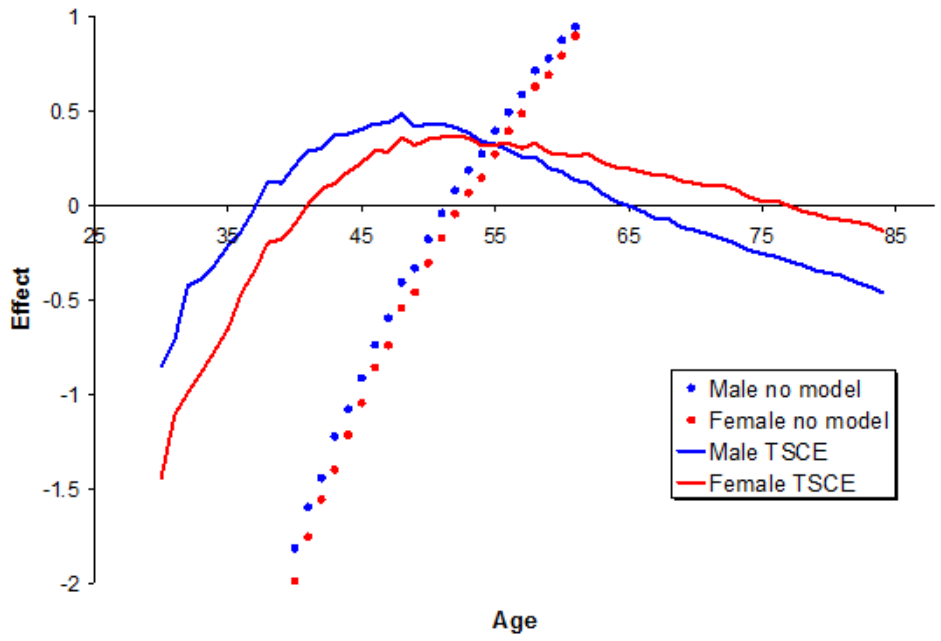


Figure 3(a). Age effects for APC calibration and no model by gender.

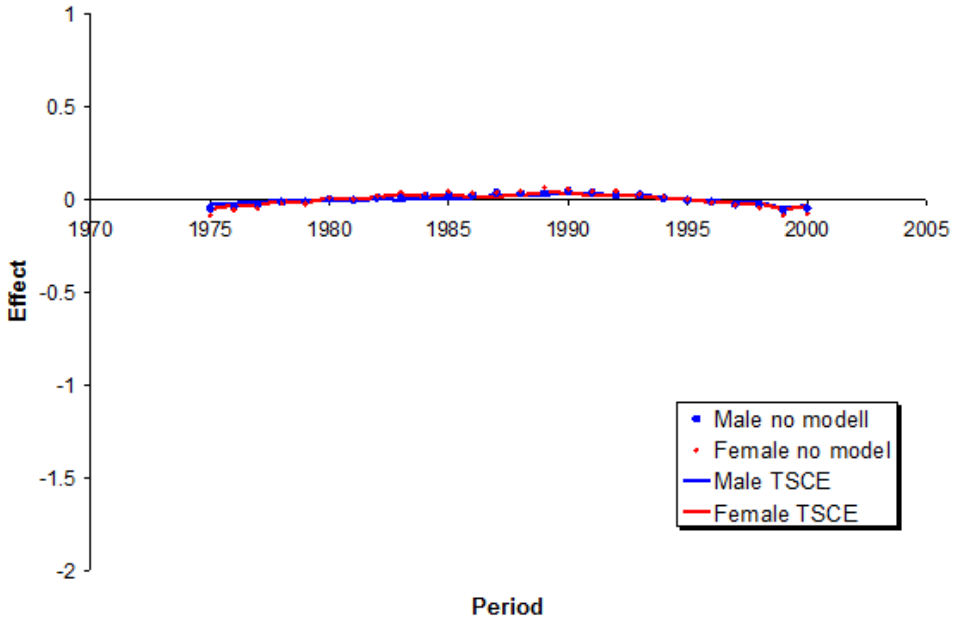


Figure 3(b). Period effects for APC calibration and no model by gender.

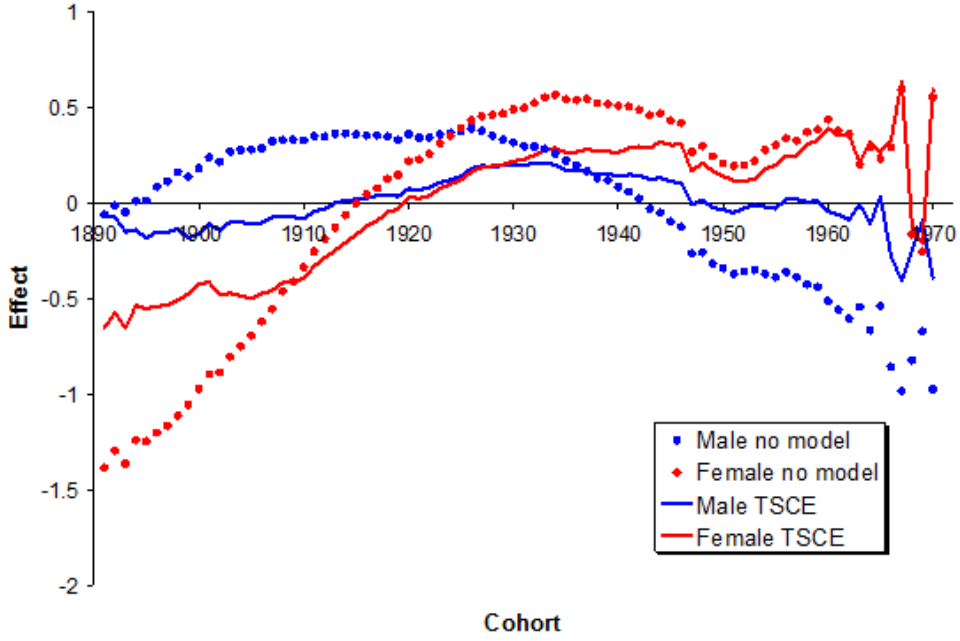


Figure 3(c). Cohort effects for APC calibration and no model by gender.

Descriptive Epidemiology Of Lung Cancer

Systematic study of cancer incidence trends for all of the major cancer sites using data from the Connecticut Tumor Registry, including lung cancer incidence, has demonstrated that cohort effects were more likely to be an important influence on the trends than period (Roush, Holford et al. 1987). This suggests that many cancer trends are likely to be the result of etiological factors rather than simply an artifact due to changes in diagnostic practice. These analyses, however, only considered the overall incidence for a particular site. Recent epidemiological studies indicate that when trying to understand time trends in disease incidence, it is not always sufficient to consider cancer of a particular site as a homogeneous entity. Instead, a further breakdown of the disease by histologic type and/or anatomic subsite may be necessary. For example, a variety of studies show that cancers with different histologic types arising in the same organ may show very different incidence patterns; and cancers with the same histologic type but arising in different organs or different parts of the same organ may have very similar time trends.(Zheng, Mayne et al. 1993; Zheng, Holford et al. 1994; Zheng, Holford et al. 1996; Zheng, Holford et al. 1997) These observations are supported by results from recent analytical epidemiologic studies which show that exposure to a particular risk factor may be capable of altering an individual's risk of one particular type of cancer without altering the risk of other forms of cancer at the same anatomic site; and exposure to a particular risk factor may cause an increased risk from one particular histologic type of cancer for one organ but reduce risk in the same type of cancer at another.

Lung cancer is one such example, in that the overall incidence and mortality rates in the US are reported to be leveling off or declining slightly in recent years, especially in younger men, which has been attributed to a decreasing smoking rate. Studies that have explored the incidence trends by histologic type include some clinical series which report that while the incidence rates for squamous cell carcinoma and small cell carcinoma have started to decrease, the rates for adenocarcinoma have continued to increase. This increase is even larger for females. In Connecticut, adenocarcinoma of the lung has been increasing since the early 1970s while squamous cell carcinoma and small cell carcinoma have started to level off (Zheng, Holford et al. 1994). We have found that adenocarcinoma has replaced squamous cell carcinoma as the leading type of lung cancer since 1991 in Connecticut males. These trends are obscured if only the overall incidence rates are examined, and the results are important since they raise the question of whether all types of lung cancer have the identical etiology.

Also see Model Overview

AGE-PERIOD-COHORT MODELS

Data available for the study of time trends often consist of age-specific incidence rates, with age and period divided into intervals of equal width. If i(=1,...,I) represents age groups, j(=1,...J) periods, and k=(1,...,K) cohorts, then a multiplicative model for the rate in one cell of the table is, $\lambda_{ijk} = \mathrm{A}_i \Pi_j \Gamma_k$, where A_i represents the effect of age on cancer incidence, and Π_j and Γ_k are period and cohort effects respectively. It is convenient to express the incidence rate as a log linear model

$$\log \lambda_{ijk} = \mu + \alpha_i + \pi_j + \gamma_k$$

(1)

where μ is an intercept term, and α_i, π_j and γ_k the corresponding log-linear effects due to age, period and cohort. This is the classical age-period-cohort model that has been discussed in considerable detail in the literature (Fienberg and Mason 1978; Holford 1983; Kupper, Janis et al. 1983; Kupper, Janis et al. 1985; Holford 1998). In this form, a model resembles the analysis of variance, and there are no restrictions on the shape of the individual parameters. The usual constraints imply that

$$\sum\nolimits_i \alpha_i = \sum\nolimits_j \pi_j = \sum\nolimits_k \gamma_k = 0$$

The linear dependence among age, period, and cohort extends to the indices for the three time effects, in that $k = j - i + I$. Hence, the design matrix for a linear model that includes all three factors is not of full rank, and a unique set of parameters for a generalized linear model including all three factors does not exist. (Fienberg and Mason 1978; Holford 1983) While not offering a solution to the estimability problem, it is possible to develop ways of understanding the source of the difficulty so that one can express estimable components that are easily interpreted. This can be accomplished by partitioning each temporal effect into two components, the slope or overall direction of the trend and curvature or deviation from linear trend.(Rogers 1982; Holford 1983) For example, we can represent the age effect by

$$\alpha_i = \left(i - \frac{I+1}{2} \right) \beta_\alpha + \alpha_{Ci}$$

(2)

where β_α is the underlying slope for the age effect, and α_{Ci} are the curvature effects. It has been shown using a similar partition of the period and cohort effects that the curvature terms (α_{Ci}, π_{Cj} and γ_{Ck}) are all estimable, but the slopes (β_α, β_π and β_γ) are not (Rogers 1982; Holford 1983). In effect, the slopes are aliased by an indeterminate constant, ν, that is hopelessly entangled with all three effects, so that any particular set of slope estimates (indicated by asterisks) is associated with a true slope by

$$\begin{aligned} \beta_\alpha^* &= \beta_\alpha + \nu \\ \beta_\pi^* &= \beta_\pi - \nu \\ \beta_\gamma^* &= \beta_\gamma + \nu \end{aligned}$$

(3)

From the rates alone, there is no way to estimate ν.

The basic APC model is primarily a tool used in descriptive epidemiology to present and analyze temporal trends in disease rates. As such, it is often used at the first step in looking for potential risk factors that may be the causal agents driving these trends. However, this model is employed at a time when there is a consensus as to the primary cause of lung cancer trends, i.e., cigarette smoking. In addition, epidemiology studies have quantified the dose response relationship between cigarette smoking and lung cancer risk, and surveys have provided estimates of exposure trends. Thus we are employing the APC model to determine the extent to which the observed trends are explained by this knowledge, and to adjust for residual temporal effect that may result from model or data limitations.

Also see: Model Overview

Models For The Effect Of Age On Lung Cancer Incidence

Age has a strong effect on lung cancer mortality, and one early observation was the apparent nearly linear relationship between the log rate and log age. Armitage and Doll provided a rationale for this relationship by introducing a multistage model for cancer in which the rate increases as a power of age, where the power corresponds to the number of stages needed to transform a normal cell to a cancerous cell (Armitage and Doll 1954). The CPS-I study provide data on nonsmokers, which enabled Knoke et al (Knoke, Shanks et al.) to estimate parameters in the multistage model for a population of white U.S. males.

While the multistage model provides a good description of the age trends, biological research on carcinogenesis has not identified four to six stages that are typically suggested by fitting this model data. Moolgavkar et al proposed an alternative set of models in which the carcinogenesis process may be initiated in a cell that then multiplies, forming a clone. A second hit on one of these initiated cell transform it into a cancer cell that subsequently develops into clinically identified cancer. While the functional form for this two stage clonal expansion (TSCE) model is more complex than the multistage model, the fit to observed data is at least as good, if not better, than the multistage model. This limited number of two or possibly three stages corresponds much more closely to what is observed in biological research on cancer, and Hazelton et al (Hazelton, Clements et al. 2005) provide estimates of the resulting parameters that arise from this model.

Also see: Model Overview

Exposure Models For The Effect Of Cigarette Smoking On Population Rates

One approach for dealing with the identifiability problem in the age-period-cohort model is to replace temporal factors with information on one or more covariates that summarize trends in known risk factors for the disease (Stevens and Moolgavkar 1979; Stevens and Moolgavkar 1984; Brown and Kessler 1987). Underlying this approach assumes that time represents one or more risk factors that are the real culprits for disease trends. If one correctly infers the underlying factors causing temporal trends, then a better analysis would include exposure trends for the factor, rather than using time as a surrogate measure.

Cigarette smoking is by far the leading cause of lung cancer, so that one has the advantage of studying the effect of, primarily, just one risk factor (US Public Health Service 1979; Doll and Peto 1981). The strength of the association between respiratory cancer mortality rates in the U.S. and the number of cigarettes consumed per capita is impressive. Kristein (Kristein 1984) reports a correlation of 0.93 for the U.S. data when a 20-year lag in the amount of smoking is related to respiratory cancer mortality. In some ways it seems remarkable that the association is so high, because potentially important details are ignored by such an analysis, including: (a) changes in cigarette consumption are not uniform over all age groups; (b) these summaries ignore consumption differences within the population; (c) the effect of smoking on lung cancer is cumulative; (d) former smokers are at different risk than either current or nonsmokers; and (e) product changes over time may have modified the effect of a cigarette.

Previous work that included population exposure to cigarettes was limited by the level detail that was available in data. Brown and Kessler analyzed U.S. lung cancer mortality from 1958-82 in order to forecast the trends through 2025(Brown and Kessler 1987). Brown and Kessler fit a model that only used data on cigarette composition over time, i.e., a measure of tar exposure, which would be expected to affect primarily the period parameters. Thus, the model was

$$\log \lambda_{ijk} = \mu + \alpha_i + \beta X_j + \gamma_k$$

where is a measure of the population's tar exposure for the j-th period, allowing for an appropriate time lag. Stevens and Moolgavkar made use of data from England and Wales which purported to give population summaries of total cigarette consumption, thus enabling them to develop a model that expressed the cohort effect as a function of the number of cigarettes smoked.(Stevens and Moolgavkar 1979; Stevens and Moolgavkar 1984) This model assumed that the log death rate is a linear function of the average cumulative number of cigarettes smoked, because the population summary was limited to aggregate information for the population yielding

$$\log \lambda_{ijk} = \mu + \alpha_i + \pi_j + \log\left(1 - X_{1,ik} + X_{1,ik}\rho^{X_{2,ik}}\right)$$

where $X_{1,ik}$ is the proportion of the population that ever smoked, $X_{2,ik}$ is the mean cumulative cigarettes consumed and ρ estimates the relative risk for the association between smoking one unit and lung cancer risk. In a study of lung cancer incidence in Connecticut, Holford et al used estimates of trends in smoking prevalence and quit ratios derived from the Health Interview Surveys conducted by the National Center for Health Statistics (Holford, Zhang et al. 1996). These models were able to account for 82% of the trends attributable to period and cohort, although the estimates of the effects of cigarette smoking did not agree well with those obtained from analytical studies.

Also see: Model Overview

KEY REFERENCES

Armitage, P. and R. Doll (1954). "The age distribution of cancer and a multi-stage theory of carcinogenesis." British Journal of Cancer 8: 1-12.

Brown, C. C. and L. G. Kessler (1987). "Projections of lung cancer mortality in the U.S.: 1985-2025."

Capocaccia, R. (1993). "Relationships between incidence and mortality in non-reversible diseases." Statistics in Medicine 12: 2395-2415.

Chiang, C. L. (1984). The Life Table and Its Applications. Malabar, FL, Krieger Publishing Co.

Doll, R. and A. B. Hill (1964). "Mortality in relation to smoking: Ten years observations of British doctors." British Medical Journal 1: 1399-1410.

Doll, R. and R. Peto (1981). "The causes of cancer." Journal of the National Cancer Institute 66: 1192-1308.

Fienberg, S. E. and W. M. Mason (1978). Identification and estimation of age-period-cohort models in the analysis of discrete archival data. Sociological Methodology 1979. K. F. Schuessler. San Francisco, Jossey-Bass, Inc.: 1-67.

Hammond, E. C. (1966). "Smoking in relation to the death rates of one million men and women." National Cancer Institute Monograph 67: 49-52.

Hazelton, W. D., M. S. Clements, et al. (2005). "Multistage carcinogenesis and lung cancer mortality in three cohorts." Cancer Epidemiology, Biomarkers & Prevention 14(5): 1171-1181.

Holford, T. R. (1983). "The estimation of age, period and cohort effects for vital rates." Biometrics 39: 311-324.

Holford, T. R. (1998). Age-period-cohort analysis. Encyclopedia of Biostatistics. P. Armitage and T. Colton. Chichester, John Wiley & Sons: 82-99.

Holford, T. R., Z. Zhang, et al. (1994). "Estimating age, period and cohort effects using the multistage model for cancer." Statistics in Medicine 13: 23-41.

Holford, T. R., Z. Zhang, et al. (1996). "A model for the effect of cigarette smoking on lung cancer incidence in Connecticut." Statistics in Medicine 15: 565-580.

Knoke, J. D., T. G. Shanks, et al. (2004). "Lung cancer mortality is related to age in addition to duration and intensity of cigarette smoking: an analysis of CPS-I data." Cancer Epidemiology, Biomarkers and Prevention 13(6): 949-957.

Kristein, M. M. (1984). "Forty years of U.S. cigarette smoking and heart disease and cancer mortality rates." Journal of Chronic Diseases 37: 317-323.

Kupper, L. L., J. M. Janis, et al. (1985). "Statistical age-period-cohort analysis: A review and critique." Journal of Chronic Diseases 38: 811-830.

Kupper, L. L., J. M. Janis, et al. (1983). "Age-period-cohort analysis: An illustration of the problems in assessing interaction in one observation per cell data." Communication in Statistics-Theory and Methods 12: 2779-2807.

Luebeck, E. G. and S. H. Moolgavkar (2002). "Multistage carcinogenesis and the incidence of colorectal cancer." Proceedings of the National Academy of Sciences, USA 99: 15095-15100.

Moolgavkar, S. H., and Venson, D.J. (1979). "A stochastic two-stage model for cancer risk assessment. I. The hazard function and the probability of tumor." Mathematical Biosciences 47: 55-77.

Moolgavkar, S. H., Dewanji, A., and Venzon, D.J. (1988). "A stochastic two-stage model

for cancer risk assessment. I. The hazard function and the probability of tumor." Risk Analysis 8: 383-392.
Moolgavkar, S. H. and E. G. Luebeck (1990). "Two-event model for carcinogenesis: biological, mathematical and statistical considerations
" Risk Analysis 10: 323-341.
Robertson, C., and Boyle, P. (1986). "Age, period and cohort models: The use of individual records." Statistics in Medicine 5: 527-538.
Rogers, W. L. (1982). "Estimable functions of age, period, and cohort effects." American Sociological Review 47: 774-796.
Roush, G. C., Holford, T.R., Schymura, M.J., and White, C. (1987). Cancer Risk and Incidence Trends: The Connecticut Perspective. New York, Hemisphere Publishing Corp.
Roush, G. C., Schymura, M.J., Holford, T.R., et al. (1985). "Time period compared to birth cohort in Connecticut incidence rates for twenty-five malignant neoplasms." Journal of the National Cancer Institute 74: 779-788.
Roush, G. D., T. R. Holford, et al. (1987). Cancer Risk and Incidence Trends: The Connecticut Perspective. New York, Hemisphere Publishing Corp.
Stevens, R. G. and S. H. Moolgavkar (1979). "Estimation of relative risk from vital data: Smoking and cancers of the lung and breast." Journal of the National Cancer Institute 63: 1351-1357.
Stevens, R. G. and S. H. Moolgavkar (1984). "A cohort analysis of lung cancer and smoking in British males." Americal Journal of Epidemiology 119: 624-641.
US Public Health Service (1979). Smoking and health: A Report of the Surgeon General. Washington, US Department of Health Education and Welfare, Public Health Service.
Verdecchia, A., R. Capocaccia, et al. (1989). "A method for the estimation of chronic disease morbidity and trends from mortality data." Statistics in Medicine 8: 201-216.
Verdecchia, A., G. De Angelis, et al. (1994). "MIAMOD: a computer package to estimate chronic disease morbidity using mortality and survival data." Computer Methods and Programs in Biomedicine 44(2): 99-107.
Zheng, T., T. R. Holford, et al. (1994). "Time trend and the age-period-cohort effect on the incidence of histologic types of lung cancer in Connecticut, 1960-1989." Cancer. 74(5): 1556-67.
Zheng, T., T. R. Holford, et al. (1994). "Time trend and the age-period-cohort effect on the incidence of histologic types of lung cancer in Connecticut, 1960-1989." Cancer 74: 1556-1567.
Zheng, T., T. R. Holford, et al. (1997). "Time trend of female breast carcinoma in situ by race and histology in Connecticut, U.S.A." European Journal of Cancer 33: 96-100.
Zheng, T., T. R. Holford, et al. (1996). "The continuing increase in adenocarcinoma of the uterine cervix: A birth cohort phenomenon." International Journal of Epidemiology 25: 252-258.
Zheng, T., S. T. Mayne, et al. (1993). "The time trend and age-period-cohort effects on incidence of adenocarcinoma of the stomach in Connecticut, 1955-1989." Cancer 72: 330-340.

RICE-MDA (TSCE)

FLEXKB DOCUMENT
Version: HI.001.03302011.70022
Document generated: 03/30/2011

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.

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

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

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.

MODEL PURPOSE

SUMMARY

This document describes the primary purpose of the model.

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 predictions for individuals based on their unique smoking histories. The Rice-MDA model is further used to simulate lung cancer mortality in individuals based on their given smoking histories.

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.

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.

PURPOSE

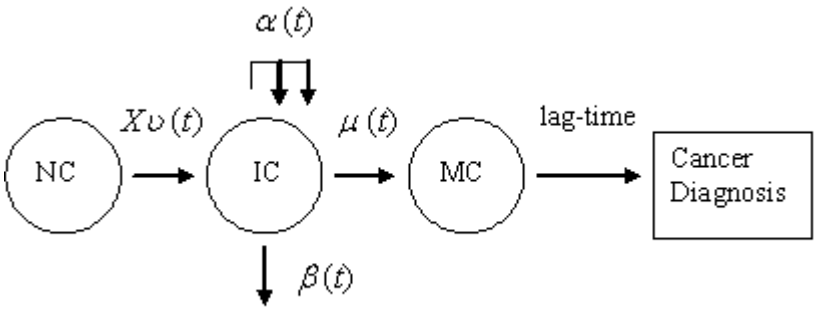
The purpose of the Rice-MDA model is to create a model for lung cancer mortality that is based on individual risk histories. The models are fit to data on risk factors collected in a case-control study combined with mortality rate data from prospective cohort studies. These models are then used to simulate 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 worldwide[2]. Smoking accounts for 90% of lung cancer cases[3]. 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 X 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 1997[5]. 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.

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

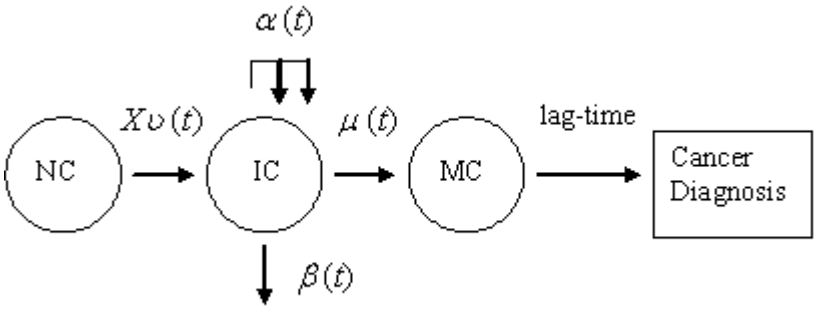
ASSUMPTION OVERVIEW

SUMMARY

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

ASSUMPTION LISTING

TSCE Model Assumptions: The TSCE model is depicted as follows and involves the following basic assumptions. More details on the model can be found in Model Overview.



- 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 k distinct 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 al[1] and Luebeck et al[2].This is assumption is justifiable since the TSCE model is insensitive to lag-time assumptions[2], 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.

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.

BACKGROUND

The TSCE model is traditionally fit to prospective cohort data. In our study, case-control data on risk factors were available. Details on how we fit the model to case-control data follow.

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	Cases	Controls
Mean Age (SD)	62.3 (11.1)	59.4 (10.9)
Race/Ethnicity	n (%)	n (%)
White	2744 (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 8[1]. 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.

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 $(P(t_i < T <= t_i +$ duration of the study (6 years)$|T > t_i) where t_i =$ 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 $P(t_i < T <= t_i + 6|T > t_i)$ 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,

$f_2 = \frac{((P(35 \leq T \leq 80|optimal DRC) - P(35 \leq T \leq 80|suboptimal DRC)) \times R_{21}) \times n)^2}{P(35 \leq T \leq 80|optimal DRC) \times n}$

where,

$R_{21} = \frac{P_2}{P_1} = \frac{P(35 \leq T \leq 80|optimal DRC)}{P(35 \leq T \leq 80|suboptimal DRC)} = \frac{P(35 \leq T \leq 80; optimal DRC)/P(optimal DRC)}{P(35 \leq T \leq 80; suboptimal DRC)/P(suboptimal DRC)}$

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 case-control study within 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

$\nu X(t) = 0.0114$

$\mu(t) = 4.845 \times 10^{-7}$

$\alpha(t) = 1.12$

$\beta(t) = 1$

Smokers while not smoking

$\nu X(t) = 0.0934$

$\mu(t) = 1.5023 \times 10^{-7}$

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

Smokers when smoking
$\nu X(t) = 0.2568$
$\mu(t) = 4.1317 \times 10^{-7}$
$\alpha(t) = 1.12 \times (1 + 0.1655 \times (log(cpd) - 1))$
$\beta(t) = 1 \times (1 + 0.1655 \times (log(cpd) - 1))$

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,

$L = \prod_j L_j$

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.

$$L_j(t_j, s_j) = \begin{cases} h(t_j - t_{lag}) S(t_j - t_{lag}) / S(s_j - t_{lag}) & \text{if diagnosed with cancer} \\ S(t_j - t_{lag}) / S(s_j - t_{lag}) & \text{otherwise} \end{cases}$$

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=$cpd/20$).

$X = 10^7$
$\nu X(t) = 1.5 \times (1 + 2.2 \times ppd)$
$\mu(t) = 1.5 \times 10^{-7} \times (1 + 2.2 \times ppd)$
$\alpha(t) = 3.2 \times (1 + 0.32 \times ppd)$
$\gamma(t) = \alpha(t) - \beta(t) - \mu(t) = 0.072 \times (1 + 0.32 \times ppd)$

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.

Overview

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

Component Listing

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 re-weighting, to match the US population. Details on the simulation of lung cancer in individuals is described in the Survival Mortality Component section.

SMOKING HISTORY GENERATOR COMPONENT

SUMMARY

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 lung cancer consortium.

OVERVIEW

The core SHG software was parameterized using three tobacco control scenarios to 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-

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**

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

Inupt	**ATC**	**NTC**	**CTC**
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 (1890-1984)	User Defined		
Gender (Male/Female)	User Defined		
Race (All race)	User Defined		

[1] Cigarettes smoked per day,[2]Other 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

Parameter	Valid Values
Seed value for PRNG used for Initiation, Cessation, OCD[1], Smoking intensity quintile	Integer from -1 to 2147483647 (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, Smoking intensity quintile and CPD[4] data files	As derived from NHIS depending on the scenario

[1]Other 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. [3]Key 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. [4]Cigarettes 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 History	Smoking intensity quintile (5 quintiles ranging from light to heavy smoking), Yearly smoking dose (CPD[2])

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

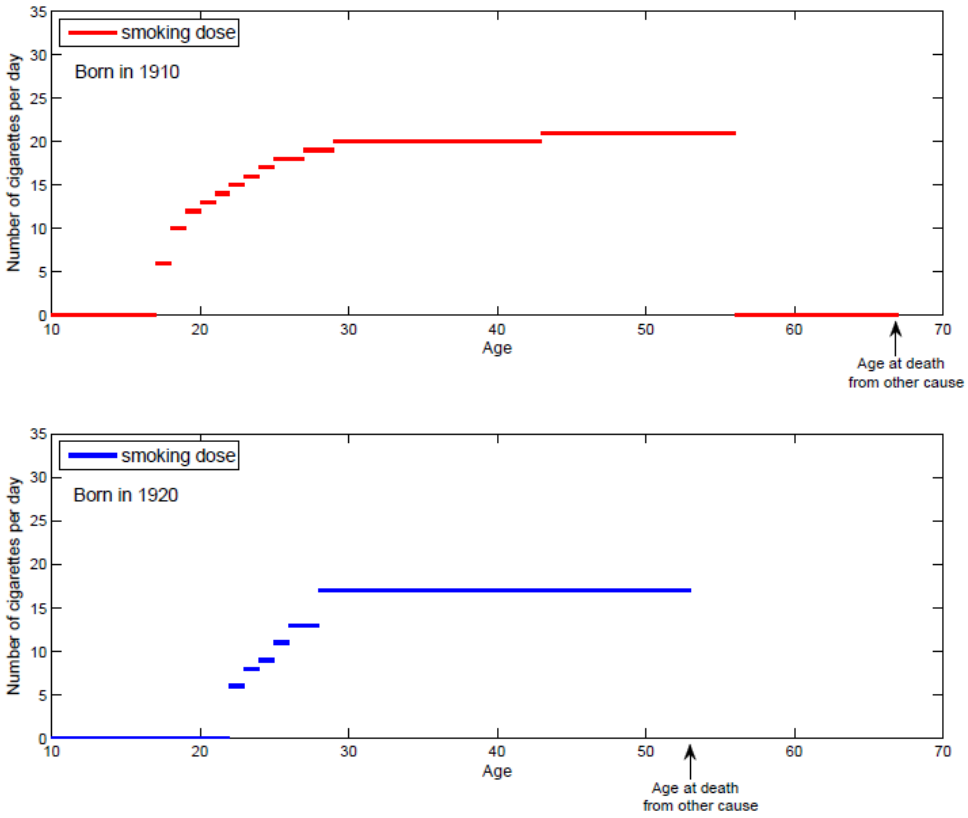


Figure SHG-1: Examples of the SHG-Generated Events

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 the Rice-MDA model.

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 by the year 2000 then their age in 2000 is considered their censoring time. Using these 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 $u <= pTSCE$ 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.

OVERVIEW

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

OUTPUT LISTING

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.

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.

OVERVIEW

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

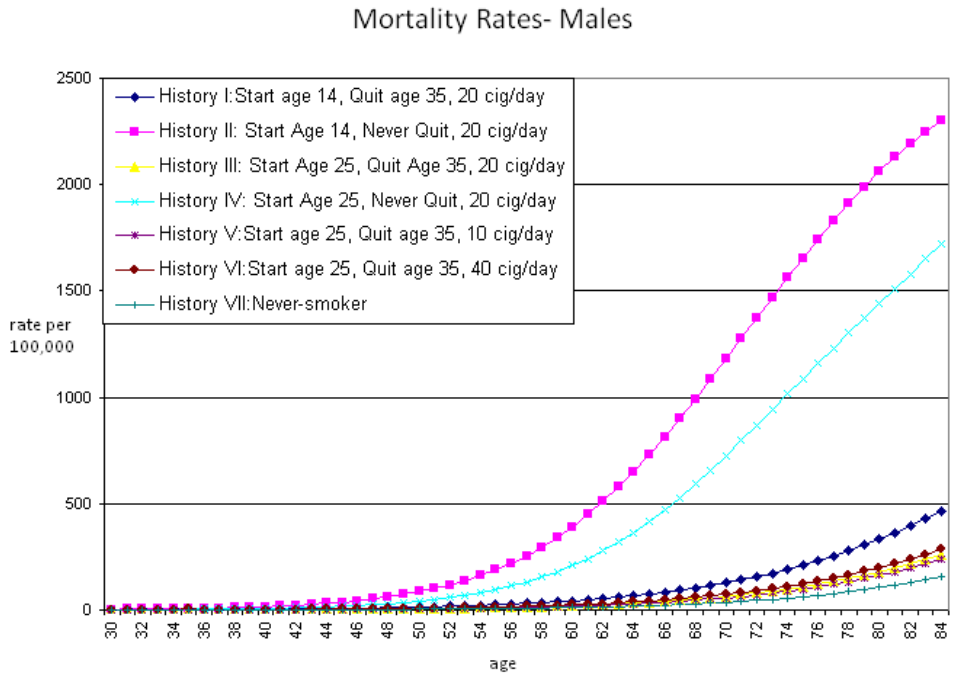
First, for the Hypothetical Scenarios the model was used to produce mortality curves for an individual based on a given hypothetical smoking history. These curves were generated directly from the model.

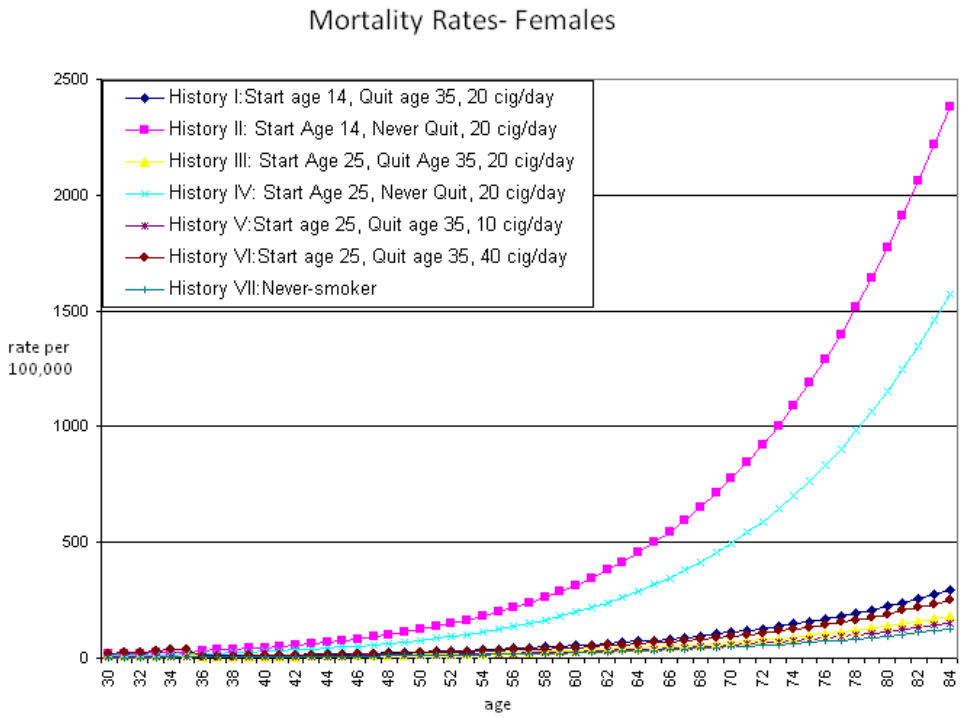
Second, the models were used to simulate lung cancer mortality in the US population. Using individuals generated from the Smoking History Generator as inputs the model was was used to simmulate whether the individuals died from lung cancer and the age at death. Fifty thousand individuals were simulated 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.

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.





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.