

YALE UNIVERSITY

FLEXKB DOCUMENT Version: HI.001.01132012.83666 Document generated: 01/13/2012



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

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



MODEL PURPOSE

SUMMARY

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

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.

CATEGORIES

Core Docs



MODEL OVERVIEW

SUMMARY

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



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

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 CATEGORIES Core Docs



ASSUMPTION OVERVIEW

SUMMARY

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

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

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.

CATEGORIES Core Docs



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.



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

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:

- 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



Yale University Parameter Overview Parameter Listing Overview 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.

CATEGORIES Core Docs



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.



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



- 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.

CATEGORIES Core Docs



Yale University Smoking History Generator Component



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

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-



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

Computational process in the usage of the SHG

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

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

1. Initialization

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



Yale University Smoking History Generator Component Relevant Parameters

RELEVANT PARAMETERS

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

Table SHG-I

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

¹ Cigarettes smoked per day,²Other Cause of Death

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



Table SHG-II

Yale University Smoking History Generator Component Dependent Outputs

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

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

DEPENDENT OUTPUTS

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

Table SHG-III

Table SHG-III

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

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



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

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

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

REFERENCES:

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





SURVIVAL MORTALITY COMPONENT

SUMMARY

This document describes how 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

$$\lambda = p_0 \lambda_0 (t) + p_1 \lambda_1 (t, \bar{z}_1, \bar{t}_{1i}) + p_2 \lambda_2 (t, \bar{z}_2, \bar{t}_{2i}, \bar{t}_q)$$

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

$$\bar{\lambda}_2 = \sum\nolimits_j p_{2j} \lambda_{2j} \left(t, \bar{z}_{2j}, \bar{t}_{2j,i}, \bar{t}_{j,q} \right)$$

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

$$\theta(a, p, c) = \exp\left\{\mu + \alpha_a + \pi_p + \gamma_c\right\}$$





Yale University Survival Mortality Component Two stage clonal expansion model 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.



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

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.
 - 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.

- 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).

CATEGORIES Core Docs



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 (C^2), F-tests (P

			Ma	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	_	

¹F-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	Actual Tobacco Control	No Tabacco Control	Complete Tobacco Control		
Approach	Tobacco Control	Tobacco control	Tobacco Control	Tobacco control	
Constant Calibratic	on				
Males	2,067,778	2,608,186	1,056,518		
Females	1,051,980	1,250,552	480,375		

APC Calibration

Ma	ales	2,067,775	2,670,897	958,862
Fer	males	1,051,978	1,273,151	438,857



Results Overview Results List



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.

CATEGORIES Core Docs



Yale University Descriptive Epidemiology Of Lung Cancer



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

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



Yale University Age-Period-Cohort Models



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

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, $\lambda_{ijk} = A_i \prod_j \Gamma_{k}$, where A_i represents the effect of age on cancer incidence, and \prod_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

$$\sum_{i} \alpha_{i} = \sum_{j} \pi_{j} = \sum_{k} \gamma_{k} = 0$$

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 ($\beta_{\alpha'}$, 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

$$\begin{aligned} \beta_{\alpha}^{*} &= \beta_{\alpha} + \nu \\ \beta_{\pi}^{*} &= \beta_{\pi} - \nu \\ \beta_{\gamma}^{*} &= \beta_{\gamma} + \nu \end{aligned}$$



Yale University Age-Period-Cohort Models

(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



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

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



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

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

$$\log \lambda_{ijk} = \mu + \alpha_i + \pi_j + \log \left(1 - X_{1,ik} + X_{1,ik} \rho^{X_{2,ik}} \right)$$



Exposure Models For The Effect Of Cigarette Smoking On Population Rates 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



Readers Guide Model Overview Assumption Overview Parameter Overview Component Overview Output Overview Results 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 nonreversible 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-periodcohort 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.