Supplemental materials for:

Maciosek MV, LaFrance A, Dehmer SP, McGree D, Xu Z, Flottemesch TJ, Solberg LI. Health benefits and cost-effectiveness of brief clinician tobacco counseling for youth and adults. *Ann Fam Med*. 2017;15(1):37-47.

Appendix: ModelHealth: Tobacco Documentation

Model Version 2.1, September 2015

Introduction

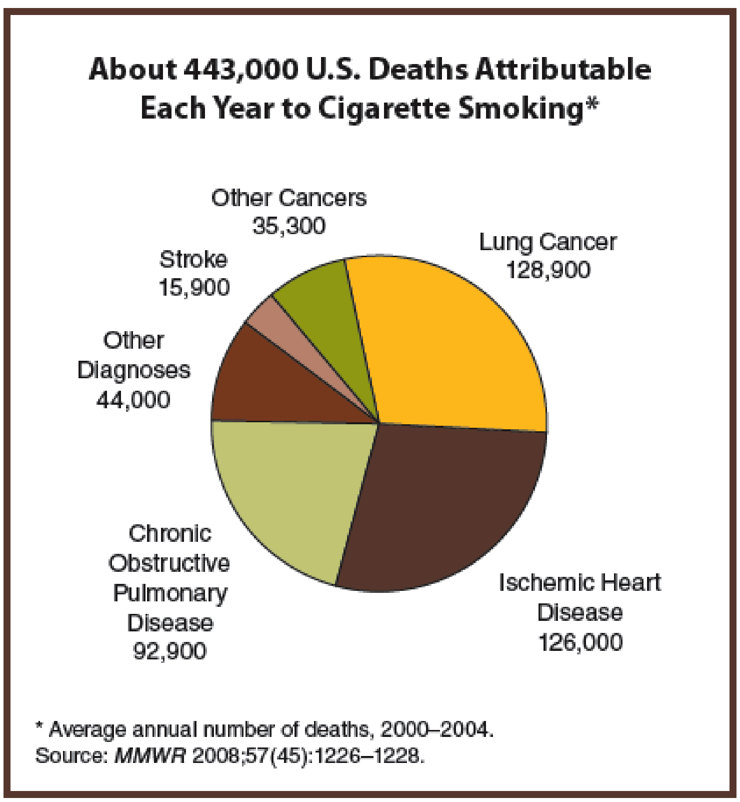
Both the United States Preventive Services Task (USPSTF) for and the *Guide to Community Preventive Services* (The Community Guide) have multiple recommendations regarding tobacco use with particular focus upon smoking.

The HealthPartners Institute for Education and Research ModelHealth™:Tobacco was developed to evaluate the health impact and cost-effectiveness of implementing these recommendations targeting diverse populations, within different environments and jurisdictions, and over varying time-frames. ModelHealth:Tobacco has been successfully used in support of multiple analyses, including state-level tax policy, and USPSTF and Community Guide recommendations.

ModelHealth:Tobacco estimates the cost-effectiveness of smoking cessation initiatives as well as the behavioral impact, health outcomes, and medical utilization impact of smoking policy. The model employs a flexible framework in which the impact of the intervention under analysis is evaluated at the individual level. These individual effects are aggregated up to the population or community level.

ModelHealth:Tobacco tracks smoking behavior of an individual and determines subsequent disease risks, health outcomes and costs. The model is able to accommodate estimation of either a single year's birth cohort or multiple birth cohorts with unique initial ages to provide the cross-sectional results. Analysis of a single year's birth cohort is common in literature of clinical preventive services. Such analyses provide insight to the average experience of patients as they age with or without studied clinical interventions. Analysis of

Figure 1. U.S. deaths attributable to smoking annually



multiple year cohorts to produce a cross-section of a population facilitates public decision-making of programs and policies that effect populations of different ages simultaneously. Such analyses are presented in the Community Health Advisor.

This document presents a description of the model, an overview of its modeling framework, the development of the inputs to the base model, and a detailed discussion of the modeling framework and embedded algorithms. Inputs specific to clinical interventions, policies and programs (counseling, tobacco taxes, media campaigns etc.) are discussed in reports specific to their analysis.

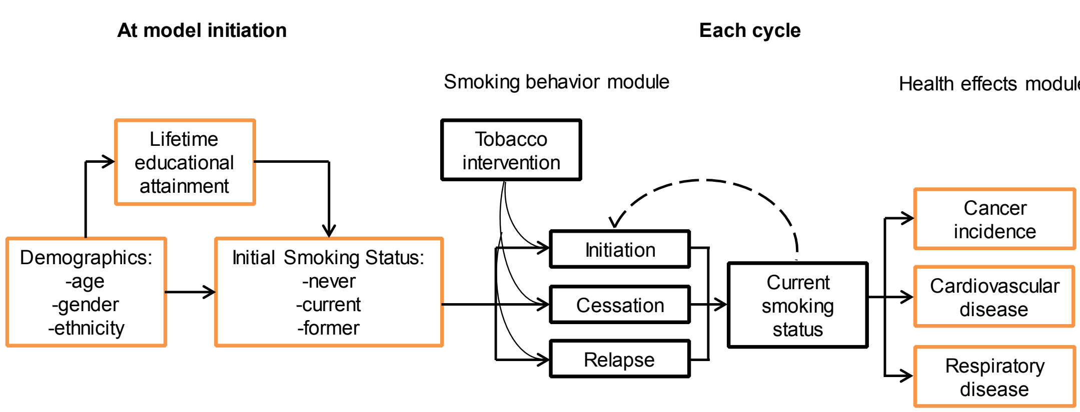
Model structure

Overview

ModelHealth:Tobacco is a Markovian individual-based model (i.e. Markov microsimulation). A Markov microsimulation is a model in which simulated individuals age over time, while facing period-specific or probabilities ('risks') of changing health behaviors and/or health outcomes ('states'). In each cycle (one year in ModelHealth:Tobacco), individuals may remain in their current state or transition to a different one. In the model, the state (age, smoking status, health, etc.) of each individual is tracked over time.

The model can be conceptualized as having three distinct parts that are shown in Figure 2 and are described further below. The first part, model initiation, allow the model to be constructed for different populations and facilitates birth cohort and cross-sectional analyses. The second and third parts, the smoking behavior module and the health effects module, determine transitions among states and outcomes in each cycle.

Figure 2. Structure of ModelHealth: Tobacco



Software

The model is constructed in TreeAge PRO 2015. The structure of the model reflects some of the capabilities and limitations of TreeAge PRO with respect to microsimulation modeling. Within TreeAge PRO, the model used multiple custom Python functions in both the base model and in implementing programs and policies for analyses. Outside of TreeAge PRO, we employ Java to import some model inputs and we frequently process model results in R. In many implementations of the model, an Excel front-end facilitates running multiple scenarios.

Initialization and population characteristics

ModelHealth: Tobacco starts by generating a population of heterogeneous simulated individuals, or agents. Four broad racial/ethnic groups are currently included: black, Hispanic, white, and other. An agent's lifetime educational achievement at age 25 is determined based on the three basic demographic factors of sex and ethnicity. Three broad levels of lifetime educational achievement are contained in the model: No high school diploma, high school degree with or without additional years of education with less than a bachelor's degree, and bachelor's degree or higher. These broad categories were used because they are consistently defined and identifiable across the multiple data sources used to parameterize the model. The likelihood of agents attaining a certain level of education at age 25 was set by sex and race/ethnicity in proportion to published data from 2010 published data from the National Center for Educational Statistics (NCES).[1]

The initial model population can vary on two important dimensions. The model can be initialized with a single-birth cohort or a cross-sectional population, and can be initialized to representative of a particular population or with population strata that can be weighted in post processing analyses.

Initialization for birth cohort vs cross-sectional analyses

Members of the population are created by assigning individual characteristics of age, sex, race-ethnicity, lifetime educational attainment, and US Census region. In model implementations that employ insurance status, characteristics of initial employment status, disability status, family poverty level and insurance type are also assigned.

However there can be differences between birth cohort and cross-sectional set-ups. Typically, in conducting analysis of a single birth cohort, the model population is assigned age zero but any age can be used. Each individual is assigned only sex, race-ethnicity, Census region, and childhood poverty status if cohort is modeled starting from an early age. When individuals reach age 18 in the model, they are also assigned employment status, disability status. Their poverty and insurance statuses may change at age 18 and each year thereafter.

In a cross-sectional set-up, individuals are assigned an age and then assigned other characteristics according to their age. A cross-section is, in effect, an analysis of multiple birth cohorts with the cohorts starting a different ages. To allow projections of population impact in future years, cohorts of individuals who are not yet alive are defined at model initiation and they are born into the model over time. It models representative of US populations, the model population grows over time as current birth cohorts that age into the model are larger than older cohorts.

At initiation these cohorts are represented with negative ages if needed and they age into the age window of analysis. For example, cohorts with an initial age of -5 represent a future birth cohort that will be born in year 5 of the simulation. In an analysis of adults ages 18 and older with an analytical horizon of 30 years, the model is initiated with a cross-section of 0 to 99 year olds, plus cohorts ages -12 to -1 years of age. Those with negative ages during any model cycle (or those at any age outside the age range of interest for a particular analysis) are excluded in post-model processing of model output.

Initialization for representative population vs weighted analysis

The simulated population in ModelHealth:Tobacco is initialized using probabilities of population characteristics. These can be set to any value. When initializing the model to represent a U.S. population, the probabilities are drawn from the Current Population Survey. They can also be set to mimic the population characteristics of participants in a randomized trial or observational study of a smoking intervention in order to predict the long-term health and economic impact of study findings.

In cross-sectional analyses of a representative population, cohorts who are initiated with negative ages (see discussion in previous section) are typically assigned the characteristics of the most recent births. In analyses for which demographic change is particularly important, projected changes in population size or characteristics can be built-in by modifying the probabilities of the cohorts that are initiated with negative ages.

The population can also be initialized assigning strata of equal size, where each strata represents a unique combination of population characteristics - for example, college-educated Hispanic women with starting age of 57 in the South census region. With strata size set to 500 and defined by sex, one of 130 ages (if initiating with negative ages to -30 years), one of three lifetime education levels, and one of four US census regions, the model start with s simulate model of 1,560,000 (500 x 2 x 130 x 3 x 4).

Conducting model runs of equal sized strata has two advantages. First, in post-model processing model results can be processed using weighted analyses to represent different populations with a single model run. Second, equally sized strata effectively over-sample smaller population groups, allowing more reliable estimates by population group in post-model processing. This is the approach taken in the Community Health Advisor.

The smoking behavior module

The model tracks smoking behavior over a lifetime and uses the smoking behavior module and the health effects module to determine the disease risk and health outcomes associated with that behavior.

The impact on smoking behavior of a clinical intervention, program or policy is determined by comparing the smoking behavior of each agent under a “baseline” scenario (a world *without* the intervention) to that agent’s smoking behavior given a world with the intervention. An individual agent’s smoking behavior may or may not change with a change in smoking behavior, and if an agent's smoking behavior does change their health outcomes may or may not be effected. Population wide effects are determined from summing the experience of all agents, those who do and do not experience change as the result of an intervention, policy and program.

For example, *the Community Guide to Preventive Services* recommends increasing the unit price of tobacco products. Some youth who would have started smoking without the increase will not start smoking with a tax increase. Of those who do not start smoking as youth with the tax increase, some will still start smoking as young adults and others will avoid a lifetime of tobacco use. Some smokers never experience significant harms of smoking; some by chance and others by quitting in time to reduce their risks. For those smokers, avoiding initiation will have no impact on health outcomes. Other would be smokers will avoid smoking-attributable disease and may have significantly longer lives. Similarly, taxes also increase the probability that smokers will quit, and whether or not a tax impacts a particular smoker's health depends on what would happened to them without a tax increase and how they respond to the tax increase. Through a series of probabilities, the microsimulation produces these heterogeneous individual experiences, and population-wide impact is determined from summing these experiences.

Smoking status

In ModelHealth:Tobacco adults may be in one of three smoking states: *never smoker*, *current smoker* and *former smoker*. Youth (younger than age 18), may be never or current smokers. Cessation and status as former smokers is not tracked for youth in the model due to the experimental nature of youth smoking and associated limitations of the data that quantify youth smoking. Adult smoking status is defined using the usual criteria of ever having smoked 100 cigarettes:

- **Never smoker:** Having smoked fewer than 100 cigarettes in their lifetime
- **Current smoker:** Having smoked at least 100 cigarettes in their lifetime and having smoked in the last week
- **Former smoker:** Having smoked at least 100 cigarettes in their lifetime and not currently a smoker

Probabilities for adult smoking status are derived from the National Health Interview Survey.[2] Youth smoking surveys ask different questions, and hence smoking prevalence rates estimated from youth surveys can yield substantially different estimates of prevalence at age 18 than do adult surveys based on the definitions above. We base our youth smoking prevalence on the Youth Risk Behaviors Survey (YRBS),[3] including self-report of age of first cigarette to estimate initiation at ages younger than the high school students who are surveyed. However, we adjust these rates to avoid discontinuity of smoking prevalence at age 19 and to calibrate the model to other's projections of smoking prevalence as described below. YRBS is limited in regard to age range and exclusion of youth who are not is school.

However its large sample size allows for more detailed estimation of smoking status by age, sex and race-ethnicity, including interaction terms.

In a birth cohort analysis that starts before age 9, all individuals are initialized as never smokers. In a birth cohort analyses that starts at an older age, and in cross-sectional analyses, individuals are initialized as being never or current smokers from aged 9-18, or never, current or formers smokers from ages 19 and older.

At model initiation, the likelihood that an agent is in any one of the three smoking states is conditioned on his/her age, gender, ethnicity, and – for those older than age 25 – the lifetime educational attainment at introduction into the model. Similarly, the likelihood that an agent who is currently in the never smoker state begins smoking within a given cycle is conditioned upon his/her age, gender, ethnicity, and – if older than age 25 – lifetime educational attainment. Given our broad educational categories, most individuals have achieved their final educational status by age 25. Our analysis intended no causal inference regarding the relationship between smoking behavior and educational attainment, merely an association.

Although the specific final multivariable risk equations vary in terms of covariates and dependent variables, several criteria were applied consistently across analyses. The statistical relationships between each covariate and other predictors were screened prior to its inclusion in a final risk equation. If the inclusion of a covariate violated assumptions (e.g., co-linearity, normality, disproportionate cell size) appropriate adjustments (e.g., center around mean, transformation, re-categorization) were made or its inclusion reconsidered. Interaction terms (e.g. differential rates of initiation between young women and young men, differential rates of cessation between African-Americans with higher education and those without a high-school diploma, etc.) were considered based on the following criteria: representing at least 10% of the larger groups (e.g. at least 10% of women *and* at least 10% of those under the age of 18, at least 10% of African Americans within each educational category), and a coefficient significant at the 10% level.

Estimating initial smoking status

A multinomial logistic regression with outcomes corresponding to the three smoking states was used to estimate the likelihood of an individual having an initial smoking status given his/her age, gender, ethnicity, and lifetime educational attainment. The estimated distribution across potential smoking states was then used to determine each agent's initial smoking status at introduction into the model.

Table 1: Results of multinomial estimation predicting initial smoking status

	Current Smoker	*95% Conf Interval*	Former Smoker	*95% Conf Interval*
Ref. Category*	-0.798	*(-0.874 , -0.722)*	-1.922	*(-2.029 , -1.816)*
Female	-0.453	*(-0.495 , -0.411)*	-0.605	*(-0.646 , -0.564)*
24-44	0.559	*(0.482 , 0.635)*	1.151	*(1.039 , 1.263)*
45-64	0.541	*(0.462 , 0.621)*	1.813	*(1.702 , 1.925)*
65+	-0.538	*(-0.632 , -0.443)*	2.203	*(2.090 , 2.315)*
Black	-0.475	*(-0.535 , -0.416)*	-0.714	*(-0.779 , -0.648)*
Hispanic	-1.249	*(-1.322 , -1.176)*	-0.723	*(-0.788 , -0.659)*
Other	-0.702	*(-0.799 , -0.604)*	-0.793	*(-0.893 , -0.694)*
High School	0.688	*(0.634 , 0.741)*	0.112	*(0.054 , 0.169)*
Post-Secondary	-1.293	*(-1.356 , -1.230)*	-0.394	*(-0.442 , -0.346)*

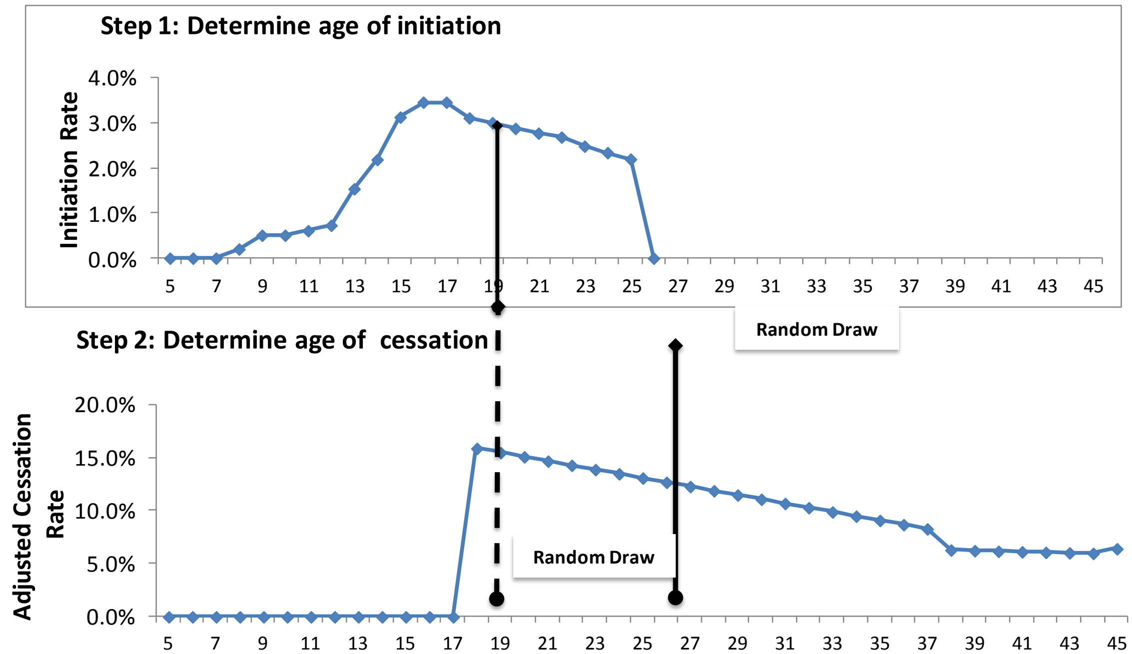
**Reference Category is Young, White, and Male, with no HS education*

Assigning age of smoking initiation and cessation

Within the behavioral module, "time in state" (i.e. the number of years spent with a smoking status) partially determines the likelihood of quitting or relapsing. For agents who start the model as never smokers and become smokers, their age at smoking initiation is determined by annual transition probabilities. Similarly, agents who are initialized as current smokers of become smokers during a model run have their age at quit determined by transition probabilities. However, an age of smoking initiation must be assigned to all agents initialized as either a current smoker or a former smoker, and an age of cessation must be assigned for those initialized as a former smoker.

Figure 3 depicts the process for an agent initialized into the model as a 45-year old former smoker. First, in Step 1, a random draw from a distribution configured to initiation rates estimated from the NHIS determines the age at which a current or former smoker first started smoking (age 19 in the example of Figure 3). Then, for those initialized as recent former smokers (Step 2), a random draw from a second distribution configured to cessation rates estimated from NHIS and truncated at the age of initiation determines the age of cessation (age 26). These two ages are then used to determine the time spent smoking and time since cessation, which are used in by the model when determining future smoking behavior.

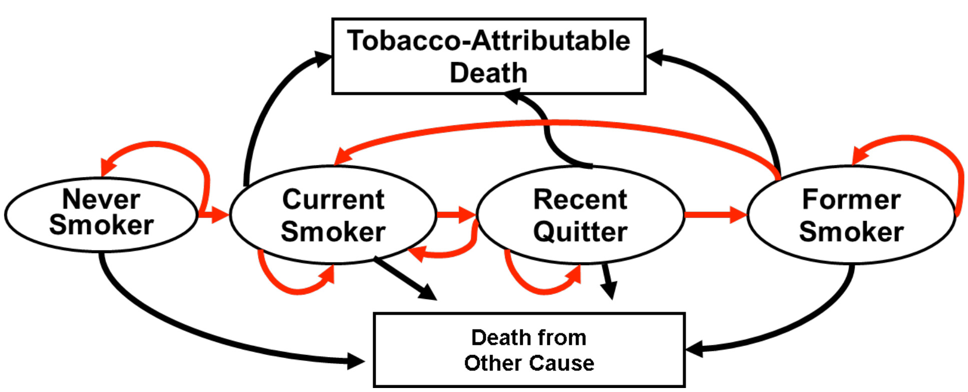
Figure 3: Determination of age of initiation and cessation



Estimating changes in smoking status

An individual's "risk" of changing smoking status (i.e. *transitioning to another smoking state*), is determined by current state, time in that state, and demographics. Individuals who have never smoked can either remain in the never smoker state or begin smoking and transition to the current smoker state. A current smoker who is in the current smoker state can remain or quit and to the former smoker state. A former smoker either relapses into the current smoker state or remains in the former smoker state. In addition, all individuals are faced with a risk of dying of either a tobacco-related illness or some other cause. Figure 4 illustrates this conceptual framework of the natural history of smoking tobacco use.

Figure 4: Natural History of Smoking Tobacco Use



Three separate logistic regressions determined the risk of smoking initiation by comparing Initiators to Never Smokers. The first, which used YBRS data,[3] applied to ages younger than 18. The second and third, which used NHIS data,[2] applied to ages 18-24 and 25 and older, respectively. Similar to the initial smoking status risk equations, the 19-24 specification was distinguished by inclusion of lifetime educational achievement.

We assumed no cessation among youth younger than age 18 due to data limitations, and instead use the available youth prevalence data to model 'net initiation' (Table 2) without explicitly modeling cessation or tracking former smoker status in youth.

Table 2: Youth tobacco smoking initiation rates*

Age	Male	Female
8	0.002	0.006
9-10	0.005	0.006
11-12	0.010	0.013
13-14	0.022	0.021
15-16	0.027	0.027
17-18	0.010	0.013

**From YBRS data*

We estimated two cessation risk equations for adults. From the NHIS data, we identified Quitters as those indicating they had ceased cigarette use within the last 12 months with no indication of relapse. Two logistic regressions (18-24 and 25 and older) compared Quitters to Current Smokers to determine the likelihood of smoking cessation. Once again, the 19-24 specification was distinguished by inclusion of lifetime educational achievement.

Table 3: Results of logistic regressions predicting adult smoking

	Tobacco Initiation	*95% Conf Interval*	**Tobacco Cessation**	*95% Conf Interval*
Ref. Category	-27.7099	*(-33.273 , -22.146)*	-1.772	*(-2.133 , -1.411)*
Female	3.5358	*(4.351 , 2.721)*	-0.046	*(-0.053 , -0.039)*
24-44	9.814	*(12.472 , 7.156)*	-0.1545	*(-0.179 , -0.130)*
xFemale	-10.0481	*(-12.656 , -7.440)*	-0.00165	*(-0.002 , -0.001)*
45-64	10.441	*(12.846 , 8.036)*	-0.1181	*(-0.139 , -0.098)*
xFemale	-5.817	*(-7.292 , -4.342)*	0.2346	*(0.294 , 0.175)*
White	-6.3501	*(-7.745 , -4.955)*	0.2966	*(0.369 , 0.224)*
xFemale	-3.8882	*(-4.893 , -2.884)*	*Not Significant*	
Black	3.4254	*(4.151 , 2.700)*	-0.0603	*(-0.073 , -0.048)*
xFemale	-3.4627	*(-4.426 , -2.499)*	*Not Significant*	
Hispanic	5.0037	*(6.435 , 3.572)*	0.0776	*(0.094 , 0.062)*
xFemale	-0.0798	*(-0.096 , -0.063)*	*Not Significant*	
No High School	6.5959	*(8.319 , 4.872)*	-0.00755	*(-0.009 , -0.006)*
xFemale	-3.8882	*(-4.791 , -2.986)*	*Not Significant*	
High School	9.2186	*(11.708 , 6.729)*	0.0191	*(0.022 , 0.016)*
xFemale	-3.4627	*(-4.365 , -2.561)*	*Not Significant*	
Post-Secondary	4.5348	*(5.593 , 3.477)*	0.3067	*(0.384 , 0.230)*
xFemale	-0.0798	*(-0.096 , -0.064)*	*Not Significant*	

**Reference Category is Young Mixed Race, Male*

Relapse rates

Relapse after quitting tobacco use is time-sensitive. The longer a person has successfully quit smoking, the less likely they are to relapse. The cross-sectional design of the NHIS surveys made estimation of relapse rates that account for time since cessation difficult. Instead, we used published estimates based upon longitudinal studies. In doing so, it is important to recognize that the probability of cessation estimated from Table 4 parameters reflect smokers who quit anytime in the year prior to the survey because relapse is very high during the first year after the quit. The estimates reflect a range of former smokers who have not relapsed from between 1 week and 51 weeks since their quit. Therefore, in applying relapse rates from the literature, we sought an estimate for the first year of relapse that reflected the probability of relapse conditional on having not relapsed for an average of 6 months.

We constructed the relapse curve represented by the conditional relapse probabilities shown in Table 4 based upon retrospective reporting from the National Health Examination and Nutrition Survey as reported in the 1990 Surgeon General's report on smoking cessation,[4] prospective probabilities of relapse reported by Wetter et al.[5] and Herd et al.[6] and a review of long term relapse rates by Hughes et al.[7] These relapse rates are applied to all quits in the model, whether they are part of the baseline model or are induced by introduction of a clinical intervention, program or policy.

Table 4: Smoking Relapse Rates

Years Since Successful Quit	*Probability of Relapse*
1	0.160
2	0.115
3	0.089
4	0.071
5	0.056
6	0.045
7	0.035
8	0.026
9	0.019
10	0.012
11	0.006

Calibration of tobacco model to CBO model

To facilitate comparison, ModelHealth: Tobacco was calibrated to reflect baseline tobacco use projections of the Congressional Budget Office (CBO).[8] These calibrated initiation and cessation rates are used as the baseline in the current model.

The CBO does not provide detail on how its tobacco baseline is parameterized. We therefore worked with Figure 1-1 in the 2010 CBO report as our guide. CBO only reports its projection of smoking prevalence among all adults in Figure 1-1. Our model determines annual smoking prevalence based upon initiation, cessation and relapse, as mediated by sex, age, race-ethnicity and educational attainment. The average adult smoking prevalence reported in Figure 1-1 could be reproduced with infinite combinations of smoking initiation, cessation, and relapse rates among males and females of different ages, race-ethnicities, and educational attainment. In addition, predictions of smoking prevalence among adults depend heavily on recent, current and near-term teen smoking initiation rates. Therefore, with only Figure 1-1 and a general description of the CBO's approach as a guide, we tested a reasonable set of parameter modifications to adjust the smoking prevalence rates produced by our model over the next 10 years to better reflect CBO's baseline.

Summary of calibration steps

The following summarizes attempts and modifications to the model to calibrate to the long-term trend incorporated into the CBO tobacco tax model.

- **Step 1:** Recalibrate relative Education and Ethnicity effects within the current model
  - **NOTE:** CBO provides only population-based temporal trends. To ease calibration, the following approach was used:
    - Create and Age/Sex initiation and cessation risk table that will be used to calibrate to CBO estimates

    - For each calibration attempt, re-apply demographic differences preserving relative differences across ethnic groups as originally estimated from the NHIS dataset.
    - Education and ethnicity effects were then re-applied to initiation and cessation rates post calibration.

- **Step 2:** Compare new ModelHealth: Tobacco baselines to CBO baselines and adjust accordingly. Here, the following additional model adjustments and reparameterizations were explored to produce overall baseline tobacco use that is more similar to that shown in Figure 1-1 of the CBO report. The following outlines the different changes to the HPIER model, identified **sources** of deviation from the CBO model, and adjustments implemented along with the effect.
  - **Baseline (Initial) attempt:**
    - Population Tobacco Prevalence was too low
    - **Adjustments**
      - Added 8 yr old initiation
      - Reduced cessation for ages>50 (prior models used TreeAge Truncation option in the cessation table, TBL_TobaccoRisk)
      - Increased 9-10 yr old initiation
      - Increased 15-18 initiation
    - **Impact**
      - Elevated prevalence in Youth and Young Adults
      - Elevated prevalence in Adults

- **Source1:** Initial Tobacco Prevalence table contained "jumps" in the prevalence because TreeAge was set to use truncation. This caused "bunching" by certain ages rather than a smooth change in prevalence.
  - **Initial Resolution Attempt:** TreeAge was set to use interpolation
    - **Impact:** Smoothed smoking prevalence across all ages 0-100
  - **Final Resolution:** Interpolation outside of the model was used to fill missing age ranges
    - **Impact:** After allow interpolation in TreeAge, more realistic initial prevalence was created. This better captures estimates of age-based relative smoking mortality
      - ***Note:** Initial prevalence was fixed after age 70 at 5%*

- **Source 2:** Tobacco relapse was elevated causing too much "churn" in initial populations due to only current and former smokers in initial population
  - **Initial Resolution Attempt:** Relapse smoothed over initial 10-yr range
    - **Impact:** improved relapse estimates, but there were too many relapses among older age groups (Aged 50 or older)

    - **Final Resolution:** The rates of relapse were reduced and smoothed over each entire 10-year age range

- **Source 3:** Inconsistent initiation patterns for eight- to 12-year-olds. The initial estimates were based on small samples and literature-based estimates that grouped estimates by two-year age groups (8-10, 11-12).
    - **Initial Resolution Attempt:** Baseline initiation rates were reduced.
        - **Impact:** The initiation pattern for the 8-12 year old age group improved but provided elevated prevalence at age 13.
    - **Final Resolution:** Initiation rates were calibrated to arrive at observed prevalence at age 13, which was estimated from a larger, more-representative YRBF sample. The calibration assumes increasing initiation rates using a relative rate (RR) increase of 30% per year. This RR increase was estimated assuming a linear growth path in initiation over the 8-12 year old age range.

- **Source 4:** Unique initiation rates for ages 15-17 were missing from abstracted articles (combined age range) and use of the truncation option within TreeAge created an "average initiation" across this entire age-range.
    - **Initial Resolution Attempt:** Interpolation in TreeAge implemented
        - **Impact:** Initiation across all age ranges was forced to 0.
    - **Final Resolution:** Initiation rates were calibrated in a manner similar to that done for the 8-12 year old age. A 30% increase in the initiation rate per age arrived at the observed 18 year old prevalence.
        - **Impact:** Initiation and prevalence for new birth cohorts reflected the observed prevalence of current 18 year old cohorts.

- **Source 5: (*Key source of deviation from CBO model*)** New birth cohorts had smoking prevalence at age 18 similar to that of current 18 year olds but 3-5% higher than the prevalence forecasted by the CB0 model. *This increased prevalence was consistent with (2010) NHIS data, but not consistent with the CBO model that shows decreasing prevalence over time.*
    - **Initial and Final Resolution Attempt:** Decrease initiation rates across ages ranges using 10 yr moving average (MA) process
        - **Impact:** Lowered prevalence among new birth cohorts that resulted in a new "steady-state" population prevalence of approximately 13-14%.
        - ***NOTE:*** *This 13-14% steady state prevalence assumes demographics (sex and ethnicity) approximate to the NHIS representative sample and that may differ when weighted to population under examination.*

- **Source 6: (*Key source of deviation from CBO model*):** Estimated initiation patterns from NHIS used age-based categories that created and stepped function and subsequent "jagged" patterns of initiation. This created elevated patterns of initiation for ages up to 24.

- **Initial and Final Resolution Attempt:** Smoothed initiation rates using a moving average process across ages ranges holding implied prevalence at end date (age 24 within birth cohort) constant
  - **Impact:** Removed "jumps" in prevalence among birth cohorts, but initiation remained relatively high

- **Source 8: (KEY *Key source of deviation from CBO model*):** Adjusted initiation among 13-17 year olds to reflect baseline patterns of prevalence at model initiation
  - **Impact:** Prevalence among adolescents initially declines and then stabilizes in a manner similar to that implied by the CBO model.

The results of the recalibration are shown in Figures 5-7.

Figure 5: Prevalence of Final Calibration by Age Group

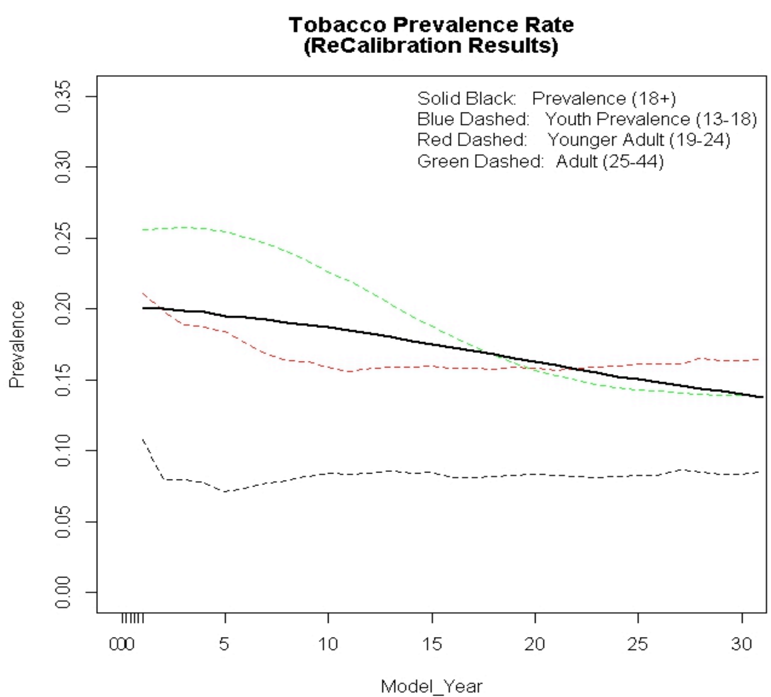


Figure 6: Initiation Rates by Age Group

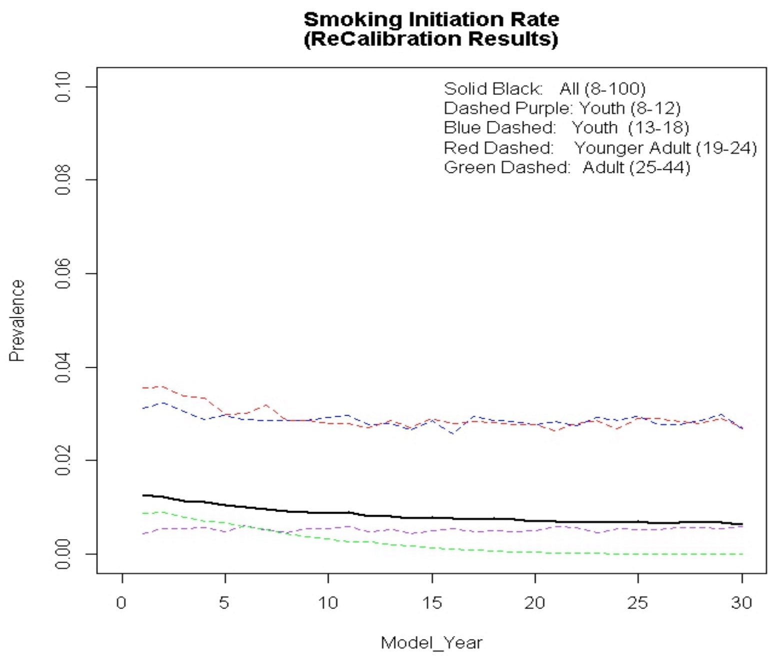
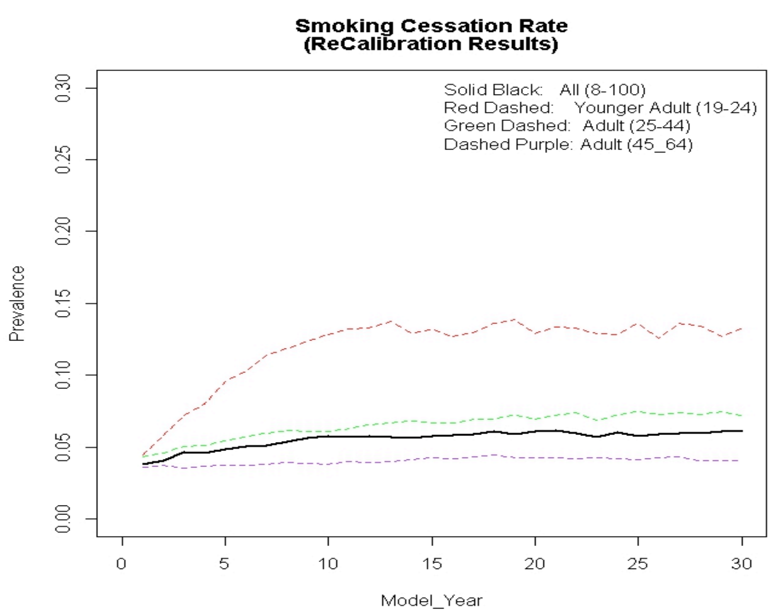


Figure 7: Cessation Rates by Age Group



Identification and specification of smoking initiation bands

There is considerable variation in tobacco prevalence across U.S. regions. The initiation equations used in the model are based on national average data. In order to obtain realistic county-specific projections of smoking prevalence, we created county-specific scalars for the initiation probabilities based on national data.

These scalars were created be first defining 'initiation bands'. Using a classification and regression tree (CART) analysis, bands were identified using County Health Rankings and Roadmaps data, which are based on the 2012 Behavioral Risk Factor Surveillance Survey (BRFSS). These county-level data were used to create evenly spaced bands of smoking prevalence with different patterns of initiation identified to arrive at the observed prevalence within each band.

The approach to identification and definition of the initiation bands is as follows:

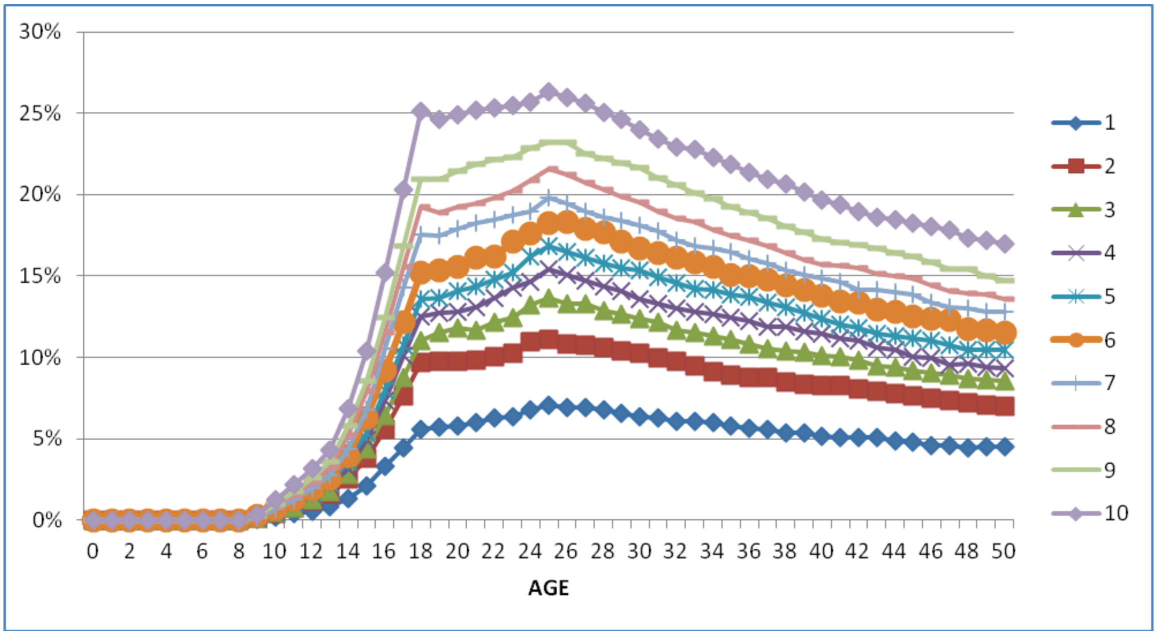
1. All counties with data within the BRFSS were grouped into categories of tobacco prevalence using a tree-classification algorithm[9,10]
2. Determine each grouping's mean and median prevalence relative to the overall U.S. prevalence:
    a. Relative Prevalence$_{\text{County i}}$ = Prevalance$_{County\ I}$ / U.S. prevalence
    b. Calculate descriptive statistics of county prevalence
    c. ***NOTE 1:*** *Groupings were determined according to ranges of prevalence not deciles of the county distribution.*
3. Adjust initiation rates in the model by applying a single scalar across all initiation rates.
    a. Four calibration runs were required
    b. The final grouping, relative initiation scalars and number of counties are listed in Table 5

Table 5: Baseline smoking tobacco relapse rates

Band	Relative initiation rate	Current median rate	Low	High	Range	N (counties)
1	0.3494	8%	0.8%	10.0%	9.2%	87
2	0.5872	11%	10.0%	13.0%	3.0%	162
3	0.7026	14%	13.0%	15.0%	2.0%	198
4	0.7833	16%	15.0%	18.0%	3.0%	413
5	0.9215	19%	18.0%	20.0%	2.0%	343
6	**1.0000**	**21%**	**20.0%**	**22.0%**	**2.0%**	**345**
7	1.1042	23%	22.0%	24.0%	2.0%	332
8	1.2709	25%	24.0%	27.0%	3.0%	330
9	1.4227	28%	27.0%	30.0%	3.0%	199
10	1.7057	33%	30.0%	47.0%	17.0%	124

Figure 9 illustrates the pattern of smoking prevalence for a series of modeled birth cohorts by initiation band.

Figure 9: Lifetime tobacco prevalence by initiation band



Determination of cigarette consumption and tax revenue

The final component of the smoking behavioral sub-module is determination of cigarette consumption among current smokers. Age, sex, ethnicity, and gender specific patterns of cigarette use were estimated using 2012 NHIS data.

Upon entering the "smoking state" (i.e. upon smoking initiation of relapse), the agent's daily cigarette consumption (CPD) is determined by a random draw from a Poisson distribution conditioned on age, sex, ethnicity, and education-based averaged estimated from the 2012 NHIS survey. Table 6 summarizes the distribution of cigarettes per day (CPD) by key demographics and consumption categories.

Table 6: Cigarettes per day (CPD)

Cigarettes per day (CPD)		1-9	10-19	20-29	30+
	Overall	37.51%	33.02%	23.61%	5.86%
Sex	**Male**	35.50%	30.99%	26.21%	7.30%
	Female	40.06%	35.59%	20.32%	4.03%
Age (M)	**0-18***	80.00%	10.00%	10.00%	0.00%
	18-24	50.83%	31.68%	15.51%	1.98%
	25-44	41.06%	31.11%	23.86%	3.96%
	45-64	27.11%	31.17%	30.52%	11.20%
	65+	30.46%	29.31%	29.31%	10.92%
Age (F)	**0-18***	80.00%	10.00%	10.00%	0.00%
	18-24	52.41%	34.48%	12.07%	1.03%
	25-44	42.48%	36.98%	17.83%	2.71%
	45-64	35.47%	34.83%	23.76%	5.94%
	65+	37.78%	34.38%	23.58%	4.26%
Ethnicity (M)	**1 (White)**	24.53%	32.32%	33.01%	10.14%
	2 (Black)	45.58%	35.77%	15.96%	2.69%
	3 (Hispanic)	64.62%	21.54%	12.75%	1.10%
	4 (Other)	54.68%	26.60%	14.29%	4.43%
Ethnicity (F)	**1 (White)**	31.94%	38.73%	24.39%	4.93%
	2 (Black)	52.05%	35.45%	11.01%	1.49%
	3 (Hispanic)	67.80%	20.34%	9.83%	2.03%
	4 (Other)	62.70%	18.25%	15.08%	3.97%
Education (M)	**1 (No HS)**	34.35%	28.12%	28.55%	8.99%
	2 (HS)	31.07%	33.59%	28.04%	7.30%
	3 (Post-Secondary)	41.84%	29.42%	22.58%	6.16%
Education (F)	**1 (No HS)**	37.29%	33.28%	23.08%	6.35%
	2 (HS)	37.18%	34.29%	23.72%	4.81%
	3 (Post-Secondary)	46.43%	33.67%	17.19%	2.71%

**Not in NHIS data and derived from YRBS*

As shown in Table 6, cigarette consumption among smokers tends to increase with age. To accommodate this trend, the agent's CPD is reset by a new random draw every five years they remain in the smoking state. For those who relapse, CPD is also determined by another random draw upon returning to the smoking state.

The total number of packs consumed per agent during the year (TPY), as well as per-pack tax revenue, is calculated by dividing the product of CPD x 365 days per year by 20 cigarettes per pack. However, self-reported cigarette consumption is under-reported in surveys. Therefore, we calibrated cigarette consumption to produce US total packs sold per year[11] by multiplying each agent's CPD by the same scalar, maintaining demographic specific consumption. In the model, the number of cigarettes smoked does not change agents' health risks or smoking-attributable medial costs. This calibration only serves to obtain accurate revenues from taxes for reporting the revenue impact of tax increases and for determining the amount of earmarked tax revenue that a tax might make available to fund additional tobacco control programs and policies.

For smokers initiating and those quitting during a year, a proportion of their TPY is applied using a random draw from a uniform distribution.

The Health Impact Module

The Health Impact Module determines the health impact of the policy, or policies, under evaluation by determining how the smoking behavior of simulated individuals impacts disease incidence, morbidity and mortality. It does this by comparing the disease outcomes of each agent that occur in the baseline scenario (i.e. a world *without* the intervention) with those that occur in the intervention scenario. Population-wide estimates of an intervention's impact are determined by aggregating individual effects.

A key benefit of tracking health outcomes at the level of individual agent is that both the incidence and timing of disease outcomes can be determined. For example, consider an individual who, without intervention, begins smoking at age 18 and suddenly dies of a cardiovascular event at age 55. Then, with counseling consistent with the USPSTF recommendation for clinician smoking cessation, that individual successfully quits at age 30 and has no cardiovascular event at age 55. However, at age 70, that same individual is diagnosed with colorectal cancer and dies at age 75. The net health impact of the USPSTF recommendation upon that individual is 15 additional disease-free years and 5 additional years with colorectal cancer. From a population perspective, this agent's adherence to the USPSTF recommendation provided additional quality adjusted life years, a decrease in cardio-vascular disease burden, but an increase in cancer incidence and burden. By tracking policy impact at the individual level, we are able to identify which events are avoided and which additional events occur during the extended lifespan resulting from the policy.

The Health Impact Module is capable to tracking health impacts in one of two ways. The primary method is to track outcome across a variety of tobacco-related diseases simultaneously using age, sex, and smoking status based risks derived from the Smoking-Attributable Mortality, Morbidity, and Economic Costs (SAMMEC) tool maintained by the Center for Disease Control (CDC).[12] This approach provides a broad accounting of all smoking attributable risks and diseases. The second approach provides a detailed examination of cardio-vascular events and disease burden by interfacing with the HealthPartners Institute ModelHealth™:CVD microsimulation model.

SAMMEC-based disease and burden estimation

Our approach to attributing events by age, sex and smoking status has been described elsewhere, but in the context of creating alternative estimates of smoking-attributable medical costs by age, sex and smoking status and using a different set of diseases and relative risks than used here.[13] The mathematics used to implement the approach described below are available in the appendix of that article.

Smoking-attributable mortality risk by age, sex and smoking status

We obtained the age-specific (5 year age groups from age 35 to 84, and 85+) and sex-specific mortality risks for smoking-attributable conditions from compressed mortality files.[14] Smoking-attributable conditions are the 10 cancers, 6 cardiovascular disease categories, and 3 respiratory disease categories identified in Smoking-Attributable Mortality, Morbidity, and Economic Costs (SAMMEC)[12] as shown in Table 7 below. To distribute mortality risk by age, sec and smoking status, we applying sex-specific smoking-attributable relative risks for each disease category that we also obtained from SAMMEC.

Smoking-attributable disease risk by age, sex and smoking status

To estimate disease events by smoking status we first assessed the number and distribution of smoking-attributable disease events in the US population by age and sex. Smoking related disease events were obtained from the Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute,[15] the National Hospital Discharge Survey[16] and compressed mortality files.[14] Hospitalizations were selected if their first-listed discharge diagnosis was for a smoking-attributable disease as defined in SAMMEC. [12] From SEER we were able to derive cancer incidence rates for the same 5-year age ranges used for morality. To approximate the distribution of CVD and respiratory disease hospitalizations in the same age categories, we distributed hospitalizations according to the distribution of fatalities with in each disease.

Neither SEER cancer data nor the NHDS contain cigarette smoking status that could be used to calculate the distribution of disease events by smoking status, and relative risks for nonfatal events are not available for a broad range of diseases from another standardized source. Therefore, mortality relative risks provided in SAMMEC were used to distribute the age- and sex-specific disease events among never, current and former smokers. Relative risks are assumed to equal1.0 for ages below 35 in SAMMEC and hence there is no smoking-attributable disease prior to age 35 in the model. The use of mortality relative risks implicitly assumes that the event-fatality rate is constant across smoking status groups. If this is not the case, then our calculations may over-state or under-state the benefits of quitting.

Smoking-attributable diseases, health utilities, and duration

The Health Impact Module independently evaluates incidence of each disease. Given incidence of a particular disease, severity, final outcome (death or recovery), and episode duration is determined. Disease specific quality of life (QofL) decrements are imposed during disease episodes to capture morbidity with the maximum decrement across all concurrent episodes of .5 quality adjusted life years (QALYs).

Table 7 lists the diseases included in the Health Impact Model only with their assumed duration and quality of life decrement.

Table 7: Summary of diseases included in ModelHealth: Tobacco

	Episode Duration		Quality Adjusted Life Year Decrements	
CANCERS	**Terminal**	**Non-Terminal**	**Initial Year of Event**	**Subsequent Years**
Lip, Oral Cavity, Pharynx	4.3	5	0.2	0.2
Esophagus	1.8	5	0.3	0.3
Stomach	3	5	0.2	0.2
Pancreas	1.24	5	0.3	0.3
Larynx	2	5	0.3	0.3
Trachea_Lung_Bronchus	2	5	0.3	0.3
Cervix Uteri	4	5	0.2	0.2
Kidney and Renal Pelvis	4.7	5	0.2	0.2
Urinary Bladder	4.7	5	0.2	0.2
Acute Myeloid Leukemia	4.6	5	0.2	0.2
CVD				
Ischemic Heart Disease	0	0.5	0.1500	
Other Heart Disease	5	0.0769	0.0231	0.3
Cerebrovascular Disease Stroke+	1	until death	0.4000	0.4
Atherosclerosis	5	0.0769	0.0231	0.3
Aortic Aneurism	0	0.0769	0.0231	
Other Arterial Disease	5	0.0769	0.0231	0.3
Respiratory Disease				
Pneumonia Influenza	0	0.0384	0.0115	
Bronchitis Emphysema+	5	until death	0.2	0.2

**Durations are rounded up to the nearest cycle. Episodes with 0 duration indicate instant death and no decrement applied.*

***For CVD and Respiratory Diseases, the initial year decrement is scaled to reflect partial year episode*

+Following initial episode, agent remains at risk for death in future cycles.

The duration of terminal cancer episodes ranges from 1 to 5 years with applicable decrements applied during the terminal episode. The duration of a non-terminal cancer episode was assumed to be 5 years across all cancers. Quality of Life decrements were the same for both terminal and non-terminal cancer episodes and ranges from .2 to .3 QALYs based upon the standardized health utilities for chronic and acute conditions used in analyses for the National Commission on Prevention Priorities.[17] Once a non-terminal cancer episode ended, the individual is at risk of another episode of that cancer with no addition or reduced risk due to relapse.

From the SAMMEC data, cardiovascular and respiratory disease were modeled as both terminal events and chronic episodes with quality of life decrements ranging from .01 (influenza) to .4 (stroke). Events resulting in death had duration of one year. Non-terminal cardiovascular and respiratory events did not end, and the corresponding quality of life decrement was imposed every year following the event. Individuals experiencing a non-terminal cardiovascular and/or respiratory event could experience a repeat event. Their risk for such a repeat event was the same as that of experiencing the initial event. For instance, a non-terminal cerebrovascular disease episode (i.e. stroke) resulted in a quality of life decrement of .4 QALYs every cycle following that event. The individual experiencing that initial stroke was at risk of another stroke in subsequent years. Similarly, a person could experience repeated cases of pneumonia and/or influenza.

Use of case fatality rates

For external validity to the US population, we obtained the age and sex-specific mortality risks for smoking-attributable conditions from compressed mortality files.[14] For internal validity such that an agent may not die from a smoking-attributable condition without having an event for that condition, we apply event-fatality rates calculated as the ratio of mortality incidence rates to event incident rates by age group and sex. These are approximate rates; events occurring in one age group may precede a death that occurs during a later age group. Therefore event-fatality rates at younger ages are likely to be somewhat understated and those at older ages overstated. However, as applied in the simulation model, the timing of events and deaths (using the durations described above) remains reasonably accurate.

Competing causes of death

During each cycle, individuals are also subject to age-specific probabilities of death from other causes. These probabilities are approximated by subtracting the combined probabilities of death from smoking-attributable conditions obtained from compressed mortality data[14] from overall mortality rates by age obtained from U.S. life tables.[18]

Health benefits of interventions

In the model, the health benefits of programs and policies which preventing tobacco initiation are realized by keeping agents in the never smoking state and avoiding the probabilities of smoking-attributable disease faced by current and former smokers. The health benefits of programs and policies that encourage cessation are realized by moving smokers from the disease risks of current smokers to the lower risks of former smokers.

'Recent quitters' and lagged change in disease risk

Recent quitters have smoking-attributable health risks that is within 25% of that of current smokers for approximately 4 years after quitting although the delay for cardiovascular disease benefits may be less.[8] Therefore, ModelHealth:Tobacco imposes a 4-year lag between the time a smoker quits to the time a smokers disease risks for cancers and respiratory disease are reduced from those of current smokers to those of former smokers.

Costs and productivity

Model health tracks both direct medical care expenditures and indirect productivity impacts of smoking, though productivity impacts are not necessarily used in all analyses.

Smoking-attributable medical costs

We estimated the medical costs of smoking from observed associations between smoking status and medical costs in the Medical Expenditure Panel Survey (MEPS), using smoking status from linked National Health Interview Survey (NHIS) responses.[19] We followed the method of Levy et al.,[20] including controlling for potentially confounding factors in a two-part model using a gamma distribution and a log-link in the second part. However we combined multiple years of data (2001-2010) to create more stable estimates for age, sex and smoking status subgroups; we also estimated separate models by primary insurer to determine smoking costs by the primary insurer type. MEPS and other claims data are complicated by higher utilization of former smokers (whose quits were likely prompted by diagnoses that lead to increased healthcare utilization in the years following their successful quits). For former smokers, we fit an exponential function to the relationship of current and former risk based on time since quit, as reported by the Congressional Budget Office (Figure 3-5 in CBO report). We applied this function to the costs for current smokers (which we estimated from MEPS data) to obtain estimates of what the medical costs of former smokers would be by age, sex and time since quit, assuming they had a proactive quit:

$$y = 0.9927 - 1.086e(-0.1171t)$$

Where y is the portion of a current smokers' smoking-attributable costs that is reduced according to years since quit (=t). Thus each former's smoker cost will be calculated as a portion of current smokers' costs with the same age, sex and insurance status as estimated from MEPS.

Table 8 provides the resulting smoking-attributable costs for current smokers by age and sex. In the model, former smoker costs will vary be age, sex and year since quit per the equation specified above. For illustrative purposes, it also provides costs of former smokers who have been quit for 5 years.

Table 8. Smoking-attributable medical costs by age, gender, smoking status ($2012)

Age categories (in years)	Male Current	Female Current	Male Former*	Female Former*
Private Insurance				
0-34	0	0	0	0
35-44	987	1,210	604	740
45-54	1,265	1,499	774	917
55-64	1,597	1,843	977	1,128
65-74	1,994	2,253	1,220	1,379
75-84	2,465	2,743	1,509	1,679
85+	2,734	3,024	1,673	1,851
Medicare Insurance				
0-34	0	0	0	0
35-44	1,301	1,531	796	937
45-54	1,639	1,879	1,003	1,150

55-64	2,040	2,296	1,248	1,405
65-74	2,518	2,795	1,541	1,710
75-84	3,089	3,391	1,890	2,075
85+	3,414	3,733	2,090	2,284
Medicaid Insurance				
0-34	0	0	0	0
35-44	1,823	2,117	1,115	1,296
45-54	2,283	2,593	1,397	1,587
55-64	2,830	3,162	1,732	1,935
65-74	3,480	3,842	2,130	2,351
75-84	4,258	4,656	2,606	2,850
85+	4,702	5,123	2,878	3,136
Uninsured				
0-34	0	0	0	0
35-44	374	548	229	335
45-54	517	710	316	435
55-64	695	906	426	554
65-74	914	1,138	559	697
75-84	1,180	1,415	722	866
85+	1,332	1,571	815	962
Other/Multiple Insurance				
0-34	0	0	0	0
35-44	1,536	1,783	940	1,091
45-54	1,922	2,184	1,177	1,337
55-64	2,384	2,664	1,459	1,630
65-74	2,932	3,236	1,795	1,980
75-84	3,587	3,922	2,195	2,400
85+	3,961	4,315	2,424	2,641

*Costs of former smokers are determined by time since quit as described in the text. Former smoker costs here depict 5 years since quit.

Smoking-attributable medical costs by payer

ModelHealth:Tobacco has the ability to apportion smoking attributable costs by payer through an insurance submodel. This feature was not used in reporting results. Details are available upon request.

Productivity

ModelHealth:Tobacco includes productivity losses. Productivity gains are not included in analyses for clinical Prevention Priorities analyses. Details are available upon request.

References

1. National Center for Education Statistics: Data Tools in Sciences. 2014. http://nces.ed.gov/datatools.
2. National Center for Health Statistics. National Health Interview Survey, 2007. Hyattsville, Maryland: National Center for Health Statistics, Centers for Disease Control and Prevention2008.
3. Centers for Disease Control and Prevention. Youth Risk Behavior Survey (YRBS). 2011. http://www.cdc.gov/healthyyouth/yrbs/index.htm. Accessed 2/27/2014.
4. *The Health Benefits of Smoking Cessation.* Rockville, MD: U.S. Department of Health and Human Services Public Health Service Centers for Disease Control Center for Chronic Disease Prevention and Health Promotion Office on Smoking and Health; 1990; No. DHHS Publication No. (CDC) 90-8416.
5. Wetter DW, Cofta-Gunn L, Fouladi RT, Cinciripini PM, Sui D, Gritz ER. Late relapse/sustained abstinence among former smokers: a longitudinal study. *Prev Med.* 2004;39(6):1156-1163.
6. Herd N, Borland R, Hyland A. Predictors of smoking relapse by duration of abstinence: findings from the International Tobacco Control (ITC) Four Country Survey. *Addiction.* 2009;104(12):2088-2099.
7. Hughes JR, Peters EN, Naud S. Relapse to smoking after 1 year of abstinence: a meta-analysis. *Addict Behav.* 2008;33(12):1516-1520.
8. Congressional Budget Office. Raising the Excise Tax on Cigarettes: Effects on Health and the Federal Budget. Washington: Congressional Budget Office; 2012.
9. *R: A language and environment for statistical computing* [computer program]. Vienna, Austria2014.
10. *Recursive partitioning and regression trees* [computer program]. 2014.
11. *The Tax Burden on Tobacco.* Arlington, VA: Orzechowski and Walker;2015.
12. Smoking-Attributable Mortality, Morbidity, and Economic Costs (SAMMEC): Adult SAMMEC and Maternal and Child Health (MCH) SAMMEC software. 2007. http://apps.nccd.cdc.gov/sammec/.
13. Maciosek MV, Xu X, Butani AL, Pechacek TF. Smoking-attributable medical expenditures by age, sex, and smoking status estimated using a relative risk approach. *Prev Med.* 2015;77:162-167.
14. Centers for Disease Control and Prevention. Compressed Mortality File 1999-2006. CDC WONDER On-line Database, compiled from Compressed Mortality File 1999-2006 Series 20 No. 2L, 2009. http://wonder.cdc.gov/cmf-icd10.html. Accessed Dec 8, 2009 12:52:14 PM.

15. Horner MJ RL, Krapcho M, Neyman N, Aminou R, Howlader N, Altekruse SF, Feuer EJ, Huang L, Mariotto A, Miller BA, Lewis DR, Eisner MP, Stinchcomb DG, Edwards BK (eds). SEER Cancer Statistics Review, 1975-2006. 2009; http://seer.cancer.gov/csr/1975_2006/. Accessed April, 12 2011.
16. DeFrances CJ, Hall MJ. 2002 National Hospital Discharge Survey. *Adv Data.* 2004(342):1-29.
17. Maciosek MV, Coffield AB, McGinnis JM, et al. Methods for priority setting among clinical preventive services. *Am J Prev Med.* 2001;21(1):10-19.
18. Arias E. *United States Life Tables, 2010.* Centers for Disease Control and Prevention, Division of Vital Statistics; November 6, 2104 2014.
19. Chowdhury S MS, Wun L. *Linking Medical Expenditure Panel Survey to the National Health Interview Survey: Weighting and Estimation. Working Paper No. 12005.* Rockville, MD: Agency for Healthcare Research and Quality;2012.
20. Levy DE NJ. Assessing the effects of tobacco policy changes on smoking-related health expenditures. In: Bearman PS NK, Wright L, ed. *After Tobacco: What Would Happen If Americans Stopped Smoking?* New York: Columbia University Press; 2011:256–289.