**S1 Appendix: Model Formulation**

# S1.1 Conceptual Model Description

This section details the conceptual formulation for the dynamic, multi-state model for assessing the effects of tobacco product use behaviors on population health. Up to any combination of *N* tobacco products are being consumed within a specified population, and the model can be used to track the number of people in subpopulations over time. Subpopulations are defined to be groups of individuals that share a common set of the following attributes:

* Age, denoted by *a =* 0, 1, 2,…, *G,G+*. *G* represents the maximum individual age tracked and *G+* represents the group of ages greater than *G*. For example, model formulations frequently set *G* = 99 and *G+* represents *a* > 99.
* Sex, denoted by *s = male*, *female*.
* Tobacco product use status. An individual’s tobacco product use status with respect to the *N* tobacco products is denoted with an *N*-tuple, , where , *k =* 1,…, *N*, is be defined to be

 

where *Mk* denotes the maximum number of years tracked for quitters of product *k*. This manuscript uses the notation to denote the tobacco product use status of never use for all *N* products.

Over time, individuals age during each modeled time step; individuals may die; and individuals may change theirtobacco product use status, e.g., they may initiate new products, quit using products, relapse and take up products they had previously quit, and so on. Changes in age and tobacco use status cause individuals to move from one subpopulation to another. The probability that an individual transitions from one tobacco use status to another is determined by the individual’s current age, sex, and tobacco use status. Fig. A and B show the set of tobacco-use-status transitions for one- and two-product model formulations, respectively. This model can be extended to include additional products.



**Figure A: Tobacco-use-status transitions for a single product model formulation.**

**Figure Note**: In the current model transition takes place among subpopulations defined by age, sex, and tobacco use status.



**Figure B: Tobacco-use-status transitions for a two-product model**

**Figure Note**: Transitions are categorized into four different behavior groups: initiation behaviors, cessation behaviors, switching behaviors, and relapse behaviors. The boxes represent the nine possible use statuses, with the first and second terms corresponding to the first and second tobacco products, respectively.

Deaths are calculated using all-cause mortality probabilities. Mortality probabilities vary by age, sex, and tobacco product use status. It is assumed that current and former tobacco product use may result in an elevated mortality risk, relative to never use of the products. Deaths decrease the overall size of the population (and subpopulations), and births contribute to the population size. Net international migration can increase or decrease the population size, depending on whether immigration or emigration is larger.

The model tracks the number of individuals in each subpopulation over time. At each time step[[1]](#footnote-1) and for each subpopulation, the model tracks the number of individuals within the subpopulation, the number of individuals that leave the subpopulation and join another subpopulation, and the number of individuals within the subpopulation who die. Key model output measures include prevalence of use for each tobacco product, mortality, and mortality attributable to tobacco product use.

# S1.2 Mathematical Model Description

Let  be the set of ages,  be the set of sexes, and *U* be the set of all possible tobacco product use status *N*-tuples considered in the model. Furthermore, let *F* denote the set of childbearing ages for females. Then Equations (1) through (9) provide a mathematical representation of the model, and Table A defines key model parameters and variables. Given that much of the data necessary to develop parameters is provided on an annual basis, time steps are generally taken to be one year increments, and parameters and variables are defined and described accordingly. A different sized time step could certainly be used, and interpretation of parameter and variable definition would need to change accordingly.

 (1)

 (2)

 (3)

 (4) (5)

 (6)

 (7)

 (8)

 (9)

**Table A: Model parameters and variables**

| **Parameter** | **Description** | **Input Parameter or Output Variable** |
| --- | --- | --- |
|  | Number of individuals of age *a*, sex *s*, and tobacco use status *u* at year . The population at year is defined to be the initial population. | output (except for the initial population which is an input parameter) |
|  | Annual proportion of individuals of age *a*, sex *s*, and tobacco use status *x* that transition to tobacco use status *u* in the time interval . When *x = u*, this parameter represents the rate at which individuals maintain and do not change their tobacco use status. | input |
|  | Annual proportion of individuals with age *a*, sex *s*, and tobacco use status *u* that die in the time interval  | input |
|  | Annual number of births per female of childbearing age during the time interval  | input |
|  | Fraction of births that are of sex *s* during the time interval  | input |
|  | Number of births of sex *s* during the time interval  | output |
|  | Annual number of net international migrants per person in the population entering/leaving during the time interval  | input |
|  | Fraction of net international migrants entering/leaving the population during the time interval  of age *a*, sex *s*, and tobacco use status *u*  | input |
|  | Number of net international migrants of age *a*, sex *s*, and tobacco use status *u* entering/leaving the population during the time interval  | output |
|  | Prevalence of individuals with tobacco use status in the subset among the population of age in the subset  and sex in the subset  at year  | output |
|  | Attributable deaths (from all products) among the population of age in the subset **and sex in the subset  during the time interval  | output |

Equations [1] – [4] describe how subpopulations are tracked. The number of newborns (age 0) is determined solely by births and migration of infants less than age 1 (Equation [1]), and all newborns are assumed to have never used any of the tobacco products (Equation [2]). For older ages (Equations [3] and [4]), the size of a subpopulation is calculated by determining the number of people from the previous year who transition into a particular subpopulation and do not die and the number of net international migrants entering or leave the subpopulation.

Equations [5] and [6] calculate the number of net international migrants and births, respectively. Births are assumed to be linearly proportional to the number of females of childbearing ages in the population, and the number of immigrants is assumed to be linearly proportional to the entire population. These assumptions are made for consistency with U.S. Census Bureau projection methods. However, it should be noted that alternative birth and immigration functions could be easily substituted for the linear functions used in Equations [5] and [6].

The model assumes that the all-cause mortality proportion for a subpopulation is the product of the all-cause mortality relative risk (*RR*) for this subpopulation and the base mortality probability of dying for never users of any of the *N* tobacco products by sex and age (Equation [7]). (The base mortality probability for never users is specified in S2 Appendix). Because use of any combination of the *N* tobacco products is assumed to have no protective effect (relative to never use), all relative risks are greater than or equal to 1. Similar to births and migration, an alternative risk formulation could be substituted for this calculation.

Equations [8] and [9] describe how to calculate product use prevalence among a population and deaths attributable to use of any combination of the *N* tobacco products, respectively. The prevalence calculation is straightforward.

1. In its most general form, the model can use any specified time step for which appropriate parameters can be determined. In practice, annual time steps are most commonly used. [↑](#footnote-ref-1)