REDUCING TOBACCO USE THROUGH TAXATION IN TRINIDAD AND TOBAGO: MODELLING THE LONG-TERM HEALTH AND ECONOMIC IMPACT
REDUCING TOBACCO USE THROUGH TAXATION IN TRINIDAD AND TOBAGO: MODELLING THE LONG-TERM HEALTH AND ECONOMIC IMPACT
Reducing Tobacco Use Through Taxation in Trinidad and Tobago: Modelling the Long-Term Health and Economic Impact

The Risk Factor Model
Relative Risks
Modelling Diseases
Methods for Approximating Missing Disease Statistics
Approximating attributable cases
Potential Years of Life Lost
Premature Mortality Rate
Costs Module
Premature Mortality Costs
Propagation of errors equation

Software architecture
Aim of the Model
Summary of the Architecture of the Existing Model
Main C++ classes used by the model
  Tperson C++ class
  Tdisease C++ class
  Tscenario C++ class

Appendix B: Cigarette Tax Scenarios Output, 2015-2017, Ukraine (Results from TaXSim Modelling)

References

LIST OF FIGURES

Figure 1: Illustration of the microsimulation model
Figure 2: Male and female smoking prevalence by year for each scenario
Figure A1: Population pyramid, 2015, Ukraine
Figure A3: Ex-smoker relative risks as a function of time after smoking cessation
Figure A2: The model structure
Figure A3: Multistage disease architecture
LIST OF TABLES

Table 1: Summary of total disease cases (epidemiological) and costs (economic) by parameter, year, and scenario, total population (values in parentheses are uncertainty values) 3
Table 2: Never, ex-smoker, and smoker prevalence (%) by age group and sex, 2011 9
Table 3: References for disease data 10
Table 4: TaXSiM Model results 15
Table 5: Smoking prevalence by year, sex and scenario (percentage) 17
Table 6: Summary table of total disease cases (epidemiological) and costs (economic) by parameter, year, and scenario, total population 19
Table 7: Cases per year, total population 20
Table 8: Cumulative cases for each disease by year, total population 20
Table 9: Cumulative cases avoided relative to scenario 0 for the total Trinidad and Tobago population by 2025 and 2035 21
Table 10: Mortality cases in the total population per year 21
Table 11: Mortality Cases avoided in the total population per year 22
Table 12: Direct cumulative healthcare costs avoided (million, TT$) 22
Table A1: Summary of the parameters representing the distribution component 40
Table A2: Parameter estimates for \( \gamma_0 \) and \( \eta \) related to each disease 43
Table A3: Survival percentage for lung cancer 47
Table A4: C++ Tperson class 57
Table A5: C++ Tdisease class 59
Table A6: The C++ Tscenario class 61
ACKNOWLEDGMENTS

This report was prepared by a team led by Patricio V. Marquez, Lead Public Health Specialist, Health, Nutrition and Population Global Practice, World Bank Group. Team members include: Lise Retat, Senior Economic and Mathematical Modeller, UK Health Forum; Abbygail Jaccard, Chief Technology Officer, UK Health Forum; Laura Webber, Deputy CEO, UK Health Forum; Karl Theodore, Director, HEU, Centre for Health Economics, The University of the West Indies, St. Augustine, Trinidad; Althea La Foucade, Coordinator, HEU, Centre for Health Economics, The University of the West Indies, St. Augustine, Trinidad; Samuel Gabriel, Researcher, HEU, Centre for Health Economics, The University of the West Indies, St. Augustine, Trinidad; Christine Laptiste, Research Fellow, HEU, Centre for Health Economics, The University of the West Indies, St. Augustine, Trinidad.

The comments and advice provided by the following peer reviewers were incorporated in the final version of this assessment: Sheila Dutta, Senior Health Specialist, World Bank Group; Santiago Herrera, Lead Economist, World Bank Group; Alberto Gonima, Consultant, World Bank Group.

Washington, DC

August 31, 2018
ABSTRACT

Background:
Tobacco is a major contributor to the rise in Non-Communicable Diseases (NCDs) and is often linked to the increase in cardiovascular and respiratory diseases and various forms of cancer. Trinidad and Tobago’s existing prevention and control interventions are in urgent need of strengthening if the country is to reduce its tobacco use. Tobacco taxation has been shown to be very effective. This study quantifies the impact of increasing tobacco tax in Trinidad and Tobago on the future burden of smoking-related diseases.

Methods:
The UK Health Forum microsimulation model (McPherson and others 2007) was used to simulate a virtual ‘Trinidad and Tobago’ population and quantify the impact of different tobacco taxation scenarios on the future burden of smoking-related disease.

Results and conclusions:
The results showed that the higher tax increase scenario yielded the most significant results. If tobacco tax is increased by 100% in each of the next three years it is estimated that 2,537 new cases of smoking-related disease will be avoided by 2035, saving TT$254.7 million to the health system. These findings support the “go big and go fast” approach outlined in the World Bank report Tobacco Tax Reform – A Multisectoral Perspective: At the Crossroads of Health And Development (World Bank 2018).
Trinidad and Tobago’s existing prevention and control interventions are in urgent need of strengthening if the country is to reduce its tobacco use.
INTRODUCTION

The rising prevalence of noncommunicable diseases (NCDs) in the Caribbean represents one of the region’s biggest challenges. If not confronted head-on, the epidemic threatens to rapidly reverse the substantial health and economic gains that have been realized across the Caribbean Community (CARICOM) in the last three decades. It is estimated that in the Region of the Americas, which includes CARICOM, 80 percent of all deaths and 77 percent of premature deaths among persons ages 30 to 70 can be attributed to NCDs.

Tobacco is a major contributor to this rise in NCDs, and is often linked to the increase in cardiovascular and respiratory diseases and various forms of cancer. In the Region of the Americas, 14 percent of deaths among adults ages 30 years and under are linked to tobacco consumption, as are 16 percent of deaths from cardiovascular diseases, 52 percent of deaths from chronic respiratory diseases (PAHO 2016), and 25 percent of deaths from cancer. Associations between cigarette smoking and the risk of developing diabetes (Śliwińska-Mossoń and Milnerowicz 2017; Will and others 2001; Rimm and others 1995) and cerebrovascular diseases (Indira, Muralidhar, and Munisekhar 2014; Molgaard and others 1986) have also been found.

In 2011, the STEPS NCD Risk Factor Survey in Trinidad and Tobago indicated that approximately 21 percent of the population smoked cigarettes, with an average daily usage of 11.5 cigarettes (PAHO 2012). To curb tobacco use and institute penalties for breaches of laws and regulations, tobacco control measures including the Tobacco Control Act of 2009 and the Tobacco Control Regulations of 2013 have been implemented. Further, the country signed the World Health Organization’s Framework Convention on Tobacco Control (WHO FCTC) in August 2003, ratifying it one year later.

Tobacco Use and its Costs to Health and the Economy in Trinidad and Tobago

According to the 2006 United Nations Common Country Assessment report for Trinidad and Tobago (Government of Trinidad and Tobago, Ministry of Health 2017), tobacco was linked to 7 percent of all NCD deaths, 9 percent of ischemic heart disease and 61 percent of lung cancer in 2004. Furthermore, in 2016, ischemic heart disease, diabetes and cerebrovascular disease were the top three causes of premature deaths (in terms of years of life lost) in the country (IHME 2015).

1 CARICOM members are: Antigua and Barbuda, Bahamas, Barbados, Belize, Dominica, Grenada, Guyana, Haiti, Jamaica, Montserrat, Saint Lucia, St Kitts and Nevis, St Vincent and the Grenadines, Suriname, Trinidad and Tobago.

2 This is the World Health Organization’s STEPwise approach to Surveillance – a simple, standardized method for collecting, analysing and disseminating data in WHO member countries.
The economic costs of NCDs are also substantial. A 2016 RTI International study (Government of Trinidad and Tobago, Ministry of Health 2017) in Trinidad and Tobago estimated that the economic burden of diabetes, cancer and hypertension was roughly TT$8.7 billion annually, or approximately 5 percent of GDP. Of that total, cancer, which has tobacco use as its most important risk factor (WHO 2018) and diabetes, cost the Government of Trinidad and Tobago TT$2 billion and TT$3.5 billion, respectively, each year.

**Taxation is Key to Strengthening Prevention and Control**

Trinidad and Tobago’s existing prevention and control interventions are in urgent need of strengthening if the country is to reduce its tobacco use. Tobacco taxation has been shown to be most effective way to do this (Blecher and others 2014). For instance, WHO-recommended best practices for tobacco taxation suggest that to substantially reduce consumption, excise taxes as a percentage of the final consumer price of tobacco products should be no less than 70 percent (WHO 2014a). However, in 2016 Trinidad and Tobago’s excise taxes share was substantially lower, at 14.7 percent3 (WHO 2017). Evidently, there is much work to do to achieve this goal.

This report uses selected scenarios for increasing excise taxes on tobacco products in Trinidad and Tobago, with the aim of reducing smoking prevalence across the population. These scenarios provide inputs for modeling the long-term health and cost benefits to the population of proposed excise tax increases. The report provides evidence from the modeling exercise. Impacts are calculated relative to the status quo before the tax hike and are modeled beginning in 2017 for 2025 and 2035.

A microsimulation model was employed to simulate the long-term impact of tobacco taxations on the future burden of a range of NCDs. Specifically, the disease outcomes quantified were for coronary heart disease (CHD), stroke, chronic obstructive pulmonary disease (COPD), and lung cancer. The microsimulation model has been deemed by the Organisation for Economic Cooperation and Development (OECD) as the most relevant method for NCD modeling based on risk-factor data (Oderkirk and others 2012). This report complements modeling work done to estimate the fiscal-revenue impact and expected reduction in consumption that might stem from proposed additional tobacco excise tax increases in Trinidad and Tobago. This work has been carried out by the World Bank, using a model based on the Tobacco Tax Simulation Model (TaXSiM) developed by WHO.

Table 1 presents a summary of total disease cases (epidemiological) and costs (economic) avoided by parameter, year, and scenario, for the Trinidad and Tobago population.

---

3 For the most-sold brand.
The model estimated that by 2035, the specified tax increase would result in the avoidance of 1,633 and 2,537 new cases of smoking-related disease for the two scenarios modelled. These reductions in disease will result in TT$2.09 million and TT$19.18 million in healthcare costs avoided for the two scenarios respectively. Consequently, the results showed that scenario two (the higher tax increase) yielded the most significant results, supporting the “go big and go fast” approach outlined in the World Bank report Tobacco Tax Reform – A Multisectoral Perspective: At the Crossroads of Health And Development (World Bank 2018). This figure is conservative because: (a) only a subset of smoking-related diseases has been included (for instance, diabetes has not been included in the model in this project); (b) indirect, social care and productivity costs have not been estimated due to a lack of input data.

Finally, it is important to note that a non-statistically significant impact on premature deaths avoided was derived in the study. However, these results are based on preliminary analysis using a subset of tobacco-related diseases and limited availability of country-specific data inputs.

Table 1: Summary of total disease cases (epidemiological) and costs (economic) by parameter, year, and scenario, total population (values in parentheses are uncertainty values)

<table>
<thead>
<tr>
<th>EPIDEMIOLOGICAL OUTPUTS</th>
<th>YEAR</th>
<th>BASELINE</th>
<th>SCENARIO 1</th>
<th>SCENARIO 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cumulative cases</td>
<td>2025</td>
<td>88,645 [±95]</td>
<td>88,366 [±95]</td>
<td>88,194 [95]</td>
</tr>
<tr>
<td></td>
<td>2035</td>
<td>210,146 [±135]</td>
<td>208,512 [±135]</td>
<td>207,609 [135]</td>
</tr>
<tr>
<td>Cumulative cases avoided</td>
<td>2025</td>
<td>NA</td>
<td>279 [±134]</td>
<td>452 [±134]</td>
</tr>
<tr>
<td></td>
<td>2035</td>
<td>NA</td>
<td>1,633 [±191]</td>
<td>2,537 [±191]</td>
</tr>
<tr>
<td>Cases per year</td>
<td>2025</td>
<td>10,943 [±33]</td>
<td>10,863 [±33]</td>
<td>10,823 [±33]</td>
</tr>
<tr>
<td></td>
<td>2035</td>
<td>14,090 [±33]</td>
<td>13,891 [±33]</td>
<td>13,785 [±33]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ECONOMIC OUTPUTS</th>
<th>YEAR</th>
<th>BASELINE</th>
<th>SCENARIO 1</th>
<th>SCENARIO 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cumulative direct costs avoided (millions, TT$)</td>
<td>2025</td>
<td>NA</td>
<td>15.46 [±24.1]</td>
<td>27.81 [±24.1]</td>
</tr>
<tr>
<td></td>
<td>2035</td>
<td>NA</td>
<td>155.46 [±44.41]</td>
<td>254.73 [±44.39]</td>
</tr>
</tbody>
</table>
TRINIDAD AND TOBAGO’S EXISTING PREVENTION AND CONTROL INTERVENTIONS ARE IN URGENT NEED OF STRENGTHENING IF THE COUNTRY IS TO REDUCE ITS TOBACCO USE.
SUMMARY OF METHODOLOGY

Methodology

- The model simulates a virtual population of Trinidad and Tobago, based on latest population statistics.4
- Data on initial smoking prevalence by age and sex are extracted from the Trinidad and Tobago Chronic Non-Communicable Disease Risk Factor Survey (Pan American STEPS).
- Scenarios take account of two different tax increases on cigarette prices, and the impact of these tax increases on smoking prevalence and subsequent disease burden. 5
- Individual smokers included in the model have a specified smoking status, and a probability of contracting, dying from, or surviving a disease.
- Future prevalence of smoking is calculated based on the numbers of smokers and non-smokers who are still alive in a particular year.
- Data for disease incidence and mortality are extracted from the Institute for Health Metrics and Evaluation, Global Burden of Disease database.
- Relative risks of contracting diseases in smokers compared to never-smokers are extracted from DYNAMO-HIA.
- A five-module microsimulation model is used to predict the future health and economic impacts of tobacco taxes by 2025 and 2035.
- The model quantifies the future impact on health and related costs of different levels of tax increase relative to a “no change” scenario.

Assumptions

- Smoking prevalence follows a static trend from 2011 smoking prevalence data.
- A specified percentage of smokers who are affected by the tax increase move to the “ex-smoker” category in 2018, 2019, and 2020 in order to account for reductions in uptake due to price increases.
- If a smoker quits as a result of the intervention, he/she becomes an ex-smoker for the rest of the simulation.

---

Time since cessation is included in the model to account for changes in disease risk for an ex-smoker.

Smokers react quickly to tax changes so immediate effects are modelled in the year following the year of implementation of the tax rise.

Limitations

- The model does not take account of future changes in policy or technology.
- No change in secondhand smoke exposure is modeled.
- The baseline is static over time.
- The simulation only includes four smoking-related diseases, so results are likely to underestimate the true effects.
- No data on non-healthcare costs, for example lost productivity due to disease, were available.
- No data were available to explore differences by social groups.
- The simulation did not model a possible relapse in smoking among smokers who gave up as a result of tax-induced price increases.
- No uncertainty analysis was conducted.
TRINIDAD AND TOBAGO’S EXISTING PREVENTION AND CONTROL INTERVENTIONS ARE IN URGENT NEED OF STRENGTHENING IF THE COUNTRY IS TO REDUCE ITS TOBACCO USE.
FULL METHODOLOGY

Data Collection

Smoking Prevalence Data

Table 2 sets out the baseline prevalence data used (never, ex-smoker and smoker). The data are from 2011.

<table>
<thead>
<tr>
<th>AGE</th>
<th>SEX</th>
<th>YEAR</th>
<th>SAMPLE SIZE</th>
<th>NEVER</th>
<th>EX-SMOKER</th>
<th>SMOKER</th>
</tr>
</thead>
<tbody>
<tr>
<td>20–24</td>
<td>M</td>
<td>2011</td>
<td>116</td>
<td>68.5</td>
<td>8.6</td>
<td>22.9</td>
</tr>
<tr>
<td>25–29</td>
<td>M</td>
<td>2011</td>
<td>121.5</td>
<td>39.8</td>
<td>17.0</td>
<td>43.2</td>
</tr>
<tr>
<td>30–34</td>
<td>M</td>
<td>2011</td>
<td>121.5</td>
<td>39.8</td>
<td>17.0</td>
<td>43.2</td>
</tr>
<tr>
<td>35–39</td>
<td>M</td>
<td>2011</td>
<td>118.5</td>
<td>52.0</td>
<td>14.6</td>
<td>33.4</td>
</tr>
<tr>
<td>40–44</td>
<td>M</td>
<td>2011</td>
<td>118.5</td>
<td>52.0</td>
<td>14.6</td>
<td>33.4</td>
</tr>
<tr>
<td>45–49</td>
<td>M</td>
<td>2011</td>
<td>105.5</td>
<td>27.4</td>
<td>35.8</td>
<td>36.8</td>
</tr>
<tr>
<td>50–54</td>
<td>M</td>
<td>2011</td>
<td>105.5</td>
<td>27.4</td>
<td>35.8</td>
<td>36.8</td>
</tr>
<tr>
<td>55–59</td>
<td>M</td>
<td>2011</td>
<td>95</td>
<td>9.8</td>
<td>55.4</td>
<td>34.8</td>
</tr>
<tr>
<td>60–64</td>
<td>M</td>
<td>2011</td>
<td>95</td>
<td>9.8</td>
<td>55.4</td>
<td>34.8</td>
</tr>
<tr>
<td>20–24</td>
<td>F</td>
<td>2011</td>
<td>134</td>
<td>84.9</td>
<td>6.2</td>
<td>8.9</td>
</tr>
<tr>
<td>25–29</td>
<td>F</td>
<td>2011</td>
<td>125.5</td>
<td>69.2</td>
<td>16.5</td>
<td>14.3</td>
</tr>
<tr>
<td>30–34</td>
<td>F</td>
<td>2011</td>
<td>125.5</td>
<td>69.2</td>
<td>16.5</td>
<td>14.3</td>
</tr>
<tr>
<td>35–39</td>
<td>F</td>
<td>2011</td>
<td>150</td>
<td>79.2</td>
<td>13.5</td>
<td>7.3</td>
</tr>
<tr>
<td>40–44</td>
<td>F</td>
<td>2011</td>
<td>150</td>
<td>79.2</td>
<td>13.5</td>
<td>7.3</td>
</tr>
<tr>
<td>45–49</td>
<td>F</td>
<td>2011</td>
<td>180</td>
<td>78.6</td>
<td>13.5</td>
<td>7.9</td>
</tr>
<tr>
<td>50–54</td>
<td>F</td>
<td>2011</td>
<td>180</td>
<td>78.6</td>
<td>13.5</td>
<td>7.9</td>
</tr>
<tr>
<td>55–59</td>
<td>F</td>
<td>2011</td>
<td>165.5</td>
<td>74.4</td>
<td>19.4</td>
<td>6.2</td>
</tr>
<tr>
<td>60–64</td>
<td>F</td>
<td>2011</td>
<td>165.5</td>
<td>74.4</td>
<td>19.4</td>
<td>6.2</td>
</tr>
</tbody>
</table>

Note: Children were assumed not to smoke.

Disease Data

The following smoking-related NCDs were modeled for this study: coronary heart disease (CHD), stroke, lung cancer, and chronic obstructive pulmonary disease (COPD). Incidence
and mortality data by age and sex were extracted from the Institute for Health Metrics and Evaluation, Global Burden of Disease, and the International Agency for Research on Cancer databases. Lung cancer data were grouped with trachea and bronchus data in the database, so these may have been overestimated for lung cancer only (IARC 2012). No survival data were available for these diseases in Trinidad and Tobago, therefore survival data was calculated from incidence and mortality using WHO DISMOD II equations (WHO 2014b).

Relative risks for smokers compared to non-smokers were extracted from Dynamo-HIA for CHD (Song and others 2008; Baba and others 2006; Tolstrup and others 2014; Burns 2003; Cronin and others 2012; U.S. Department of Health and Human Services 2014), COPD (U.S. Department of Health and Human Services 2014; Prescott and others 1997; Johannessen and others 2005; Terzikhan and others 2016; Thun and others 2013), lung cancer (U.S. Department of Health and Human Services 2014; Thun and others 2013; Freedman and others 2008; Bae and others 2007), and stroke (Mannami and others 2004; Shinton and others 1989; Wannamethee and others 1995). As various cohort studies usually observed participants of different age groups, their estimates were compared and combined to cover the modeled population: Thus, relative risks for various age groups may derive from different studies.

Ex-smokers’ relative risk was assumed to decrease post-cessation and was computed using a decay function method developed by Hoogenveen and others (Hoogenveen and others 2008). This function uses current smoker relative risk for each disease as the starting point, and then models the decline in relative risk of disease for an ex-smoker over time (see appendix A).

Table 3: References for disease data

<table>
<thead>
<tr>
<th>INCIDENCE</th>
<th>MORTALITY</th>
<th>DIRECT HEALTHCARE COSTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHD (Institute for Health Metrics and Evaluation, Global Burden of Disease)</td>
<td>Institute for Health Metrics and Evaluation, Global Burden of Disease</td>
<td>Calculations based on data from The Cost of Health Services in Trinidad and Tobago. 2013 HEU, Centre for Health Economics (unpublished)</td>
</tr>
<tr>
<td>Stroke (Institute for Health Metrics and Evaluation, Global Burden of Disease)</td>
<td>Institute for Health Metrics and Evaluation, Global Burden of Disease</td>
<td>Calculations based on data from The Cost of Health Services in Trinidad and Tobago. 2013 HEU, Centre for Health Economics (unpublished)</td>
</tr>
<tr>
<td>COPD (Institute for Health Metrics and Evaluation, Global Burden of Disease)</td>
<td>Institute for Health Metrics and Evaluation, Global Burden of Disease</td>
<td>Calculations based on data from The Cost of Health Services in Trinidad and Tobago. 2013 HEU, Centre for Health Economics (unpublished)</td>
</tr>
<tr>
<td>Lung cancer (International Agency for Research on Cancer databases)</td>
<td>International Agency for Research on Cancer databases</td>
<td>Calculations based on data from The Cost of Health Services in Trinidad and Tobago. 2013 HEU, Centre for Health Economics (unpublished)</td>
</tr>
</tbody>
</table>
Health Economic Data

Calculation of Direct Healthcare Costs

Disease cost estimations were conducted using estimates of the cost of disease and health services provided in the *Cost of Health Services in Trinidad and Tobago* report carried out for the Ministry of Health in 2013 (HEU, Centre for Health Economics, 2013). This report was used in conjunction with consultations with medical professionals in respect to the various inputs required for treatment of the four health conditions studied. The cost per case per year includes costs of:

- diagnostics;
- pharmaceuticals and or medical supplies;
- rehabilitation, where applicable; and
- inpatient and outpatient care.

Diagnostic costs include the average costs of laboratory tests and imaging. Pharmaceutical costs were estimated using the most probable drugs to be prescribed for each of the diseases, together with information on the average frequency of use, length of use and the average price per drug. The cost of medical supplies/equipment was included where appropriate. The cost of inpatient care was calculated based on the average number of inpatient days per disease coupled with the average cost per inpatient day. Outpatient costs were estimated based on the average number of visits per disease condition multiplied by the average cost per outpatient visit. Where applicable, the cost of rehabilitation includes physiotherapy sessions, as well as other general rehabilitation techniques.

Population Data

To simulate the population of Trinidad and Tobago, the population by age and sex, births by mother’s age, and total fertility rate statistics were taken from the 2011 population prospects database. Total mortality rates were taken from the 2017 population prospects database.

The Microsimulation Model

The UK Health Forum (UKHF) microsimulation model was originally developed for the UK government’s Foresight inquiry (McPherson and others 2007; Wang and others 2011) and has been developed over the past decade to incorporate a number of additional interacting risk factors, including smoking (methods are described in greater detail in (UK Health Forum and CRUK 2016; UK Health Forum) and in appendix A). The model simulates a virtual population that reproduces the characteristics and behavior of a large sample of individuals (50 million). These characteristics (age, sex, smoker status) can evolve over the
life course based on known population statistics and risk factor data. Individuals can be born and die in the model, which is modular in nature (see figure 1).

- Module 1 uses cross-sectional data on the prevalence of the risk factor – cigarette smoking in this case. For the current study, 2011 smoking prevalence data for Trinidad and Tobago were extrapolated forward to 2035. It was assumed that the proportions of the population within each smoking category as calculated in 2011 remained constant until 2035.

- Module 2 is a microsimulation model which uses the prevalence of the risk factor over time, along with the specified data on the risks of developing diseases, to make projections of future disease burden.

A wide range of different outputs is produced, including cumulative incidence. To the authors’ knowledge, no other studies have used a microsimulation model to quantify the future costs and health impacts of tobacco taxation policy scenarios in Trinidad and Tobago.

**Figure 1: Illustration of the microsimulation model**

![Figure 1: Illustration of the microsimulation model](source: UK Health Forum 2017)

**Development of Scenarios**

An initial modeling study was carried out by the World Bank Group (Marquez and others 2018) using a version of WHO’s TaXSiM model.6 Within this model, a scenario that reflects tobacco excise tax changes in 2017 was simulated to calculate the revenue impact as a result of this tax increase.

---

6 WHO tobacco tax simulation model (TaXSiM) [http://who.int/tobacco/economics/taxsim/en/](http://who.int/tobacco/economics/taxsim/en/)
The modified TaXSiM also calculated the percentage reduction in total cigarette consumption due to the suggested tax changes. These taxation changes result in non-smokers (predominantly young people) not initiating smoking; smokers quitting, and smokers reducing the number of cigarettes smoked.

There was one baseline and two intervention scenarios:

- **Baseline:** A baseline “static” trend
  This assumed that smoking prevalence stays constant at 2011 rates.

- **Scenario 1:**
  The 2017 specific excise tax rate on cigarettes is increased by 50 percent in 2018 to TT$6.57 per 20 cigarettes; by 100 percent in 2019 (TT$13.14 per 20 cigarettes), and 100 percent in 2020 (TT$26.28 per 20 cigarettes).
  The consumption as a result of the previously stated tax applied to cigarettes is estimated to reduce by (only cessation included) a relative reduction of:
  - 1.95 percent in 2018
  - 5.15 percent in 2019 and
  - 7.00 percent in 2020.

- **Scenario 2:**
  The 2017 specific excise tax rate on cigarettes is increased by 150 percent in 2018 to TT$10.95 per 20 cigarettes; by 100 percent in 2019 to TT$21.90 per 20 cigarettes; and by 100 percent in 2020 to TT$43.80 per 20 cigarettes.
  As a result of the previously stated tax applied to cigarettes, consumption is estimated to reduce by (only cessation included) a relative reduction of:
  - 5.65 percent in 2018
  - 6.55 percent in 2019, and
  - 7.95 percent in 2020.

Assumptions for implementation of the scenarios in the UKHF microsimulation are as follows:

- Several studies suggest that around 50 percent of the effect of price increases on overall cigarette consumption results from participation changes (Farelly and others 2001; Centers for Disease Control 1998). Therefore, it is assumed that 50 percent of smokers would quit and 50 percent would cut down their tobacco intake. An estimated 50 percent reduction in cigarette consumption was used as an estimate of the reduction in the total prevalence of smoking. While taxation that raises the real prices of tobacco might reduce the intensity of smoking, research suggests that people who cut down may actually inhale more, as measured by serum cotinine.
levels (Fidler and others 2011). Further, the WHO target is focused on a total reduction in smoking prevalence. Therefore, modeling proceeded with a focus on current smoking prevalence, as opposed to the number of cigarettes smoked.

- Future uptake of smoking was not included in the current scenarios.

- While these average changes were not the same for each group, and usually people under 30 years of age initiate smoking, the model did not take age differences into account, and the relative decline in percentages of current smokers was applied to all age groups.

- A baseline “static” trend was included. This assumed that smoking prevalence remains constant at 2011 rates. The tax increase scenario was compared to this baseline.

- The tax increase scenarios represent the tax change adopted in 2018, 2019 and 2020.

- The scenarios are based on Monte Carlo simulations (individuals were sampled from the population and simulated over time).

- The specified percentage of smokers who are affected by the tax increase move to the ex-smoker category in 2018, 2019 and 2020.

- If an individual’s smoking status is changed by the scenario, their smoking status will remain fixed for the entire simulation.

- An immediate reduction in smoking prevalence due to the tax increases in 2018, 2019 and 2020 was assumed. We learned via personal communication with Professor Joy Townsend that there are different views on the temporal impact of a tax: econometricians follow Becker’s model, assuming that, as tobacco is very addictive, the reaction to price increases is slow and greater in the long run. Becker, therefore, uses a lagged variable of y (t-1) (Becker and Murphy 1998). Townsend and Atkinson take the opposite view (Atkinson and Skegg 1973) – that smokers tend to react quickly to a price change. A model similar to theirs was used, with an immediate effect and then a linear trend, and in line with the modified TaXSIM model outputs (table 4).
Table 4: TaXSiM Model results

<table>
<thead>
<tr>
<th>Scenario</th>
<th>2017 Specific Excise Tax Rate (TT$)</th>
<th>2018 Specific Excise Tax Rate (TT$)</th>
<th>2019 Specific Excise Tax Rate (TT$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario 0</td>
<td>TT$4.38 per 20 cigarettes</td>
<td>TT$6.57 per 20 cigarettes</td>
<td>TT$13.14 per 20 cigarettes</td>
</tr>
<tr>
<td>Scenario 1</td>
<td>TT$10.95 per 20 cigarettes</td>
<td>TT$21.90 per 20 cigarettes</td>
<td>TT$43.80 per 20 cigarettes</td>
</tr>
</tbody>
</table>

- **Total cigarettes taxed (billion pieces)**: 1.16, 1.12, 1.00, 0.86, 1.03, 0.90, 0.76
- **Average cigarette price per pack (TT$)**: 23.45, 26.48, 36.33, 56.49, 32.82, 49.73, 83.37
- **Average excise tax burden (excise tax as percentage of price)**: 18.7, 24.8, 36.2, 46.5, 33.4, 44, 52.5
- **Average excise tax (per 1,000 pieces) (TT$)**: 219, 328.5, 657, 1,314, 547.5, 1,095, 2,190
- **Average tax burden (total tax – import excise and VAT as percentage of price)**: 29.9, 36.0, 47.4, 51.7, 44.6, 55.2, 63.7
- **Percentage change in total cigarette taxed**: -1.9, -3.9, -10.3, -14.0, -11.3, -13.1, -15.9

Source: WBG Staff estimates.

Note: *Based on assumptions for elasticity price and elasticity income for high-income countries (Marquez and others 2018).
TRINIDAD AND TOBAGO’S EXISTING PREVENTION AND CONTROL INTERVENTIONS ARE IN URGENT NEED OF STRENGTHENING IF THE COUNTRY IS TO REDUCE ITS TOBACCO USE.
RESULTS

Smoking Prevalence (Percentage)

Table 5 shows smoking prevalence for males, females, and both males and females combined for the baseline scenario, and scenarios 1 and 2. By 2035, smoking prevalence decreases to 12.74 and 10.64 for scenario 1 and 2 respectively. More specifically, smoking prevalence in men reduces to 20.80 and 17.33 for scenarios 1 and 2 respectively, with the smoking prevalence among women falling to 4.80 and 4.05 respectively.

Table 5: Smoking prevalence by year, sex and scenario (percentage)

<table>
<thead>
<tr>
<th>YEAR</th>
<th>SCENARIO 0 (BASELINE)</th>
<th>SCENARIO 1</th>
<th>SCENARIO 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MALE</td>
<td>FEMALE</td>
<td>TOTAL</td>
</tr>
<tr>
<td>2017</td>
<td>25.81</td>
<td>6.48</td>
<td>16.18</td>
</tr>
<tr>
<td>2020</td>
<td>26.26</td>
<td>6.52</td>
<td>16.41</td>
</tr>
<tr>
<td>2025</td>
<td>26.58</td>
<td>6.44</td>
<td>16.51</td>
</tr>
<tr>
<td>2035</td>
<td>27.46</td>
<td>6.27</td>
<td>16.78</td>
</tr>
</tbody>
</table>

Figure 2: Male and female smoking prevalence by year for each scenario
Figure 2: Male and female smoking prevalence by year for each scenario, Cont.

Summary
There are a number of outputs from the microsimulation.

Epidemiological Indicators
Results from the microsimulation are presented as rates per the Trinidad and Tobago population, 2011.

Incidence
The total number of new cases of disease, divided by the total number of susceptible people in a given year presented as a rate per population.

Cumulative incidence rate per year, per Trinidad and Tobago population
To calculate the cumulative incidence rate per year, the total number of new cases of disease was divided by the total number of susceptible people in a given year and accumulated over a specified period of the simulation from the year 2016. Therefore, the cumulative number of cases represents a sum of all of the cases from the start of the simulation.

Cumulative incidence avoided per Trinidad and Tobago population over the simulation period
The total number of cases of disease avoided or gained as compared to baseline (i.e., scenario 0) was estimated. A positive value represents the number of cases avoided, whereas a negative value represents the number of cases gained.

Mortality per Trinidad and Tobago population over the simulation period
The number of deaths from a disease was estimated.
Mortality cases avoided per Trinidad and Tobago population over the simulation period

The number of deaths from a disease avoided or gained as compared to baseline (i.e., scenario 0) was estimated.

Economic outputs

Direct costs avoided

These are cumulative direct costs across the period of the simulation. The result for 2020 represents the cumulative costs avoided for the period 2016 to 2020. These costs are scaled to the total population of Trinidad and Tobago. Table 6 presents a summary table of total disease cases (epidemiological) and costs (economic) by parameter, year, and scenario as rates per Trinidad and Tobago population.

The model estimated that by 2035 the specified tax increase would result in the avoidance of 1,633 and 2,537 new cases of smoking-related diseases respectively for the two scenarios modelled. These reductions in disease will result in the TT$2.09 million and TT$19.18 million in healthcare costs respectively avoided for the two scenarios.

Table 6: Summary table of total disease cases (epidemiological) and costs (economic) by parameter, year, and scenario, total population

<table>
<thead>
<tr>
<th>EPIDEMIOLOGICAL OUTPUTS</th>
<th>YEAR</th>
<th>SCENARIO 0–BASELINE</th>
<th>SCENARIO 1</th>
<th>SCENARIO 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cumulative cases</td>
<td>2025</td>
<td>88,645 [±95]</td>
<td>88,366 [±95]</td>
<td>88,194 [95]</td>
</tr>
<tr>
<td></td>
<td>2035</td>
<td>210,146 [±135]</td>
<td>208,512 [±135]</td>
<td>207,609 [135]</td>
</tr>
<tr>
<td>Cumulative cases avoided</td>
<td>2025</td>
<td>NA</td>
<td>279 [±134]</td>
<td>452 [±134]</td>
</tr>
<tr>
<td></td>
<td>2035</td>
<td>NA</td>
<td>1,633 [±191]</td>
<td>2,537 [±191]</td>
</tr>
<tr>
<td>Cases per year</td>
<td>2025</td>
<td>10,943 [±33]</td>
<td>10,863 [±33]</td>
<td>10,823 [±33]</td>
</tr>
<tr>
<td></td>
<td>2035</td>
<td>14,090 [±33]</td>
<td>13,891 [±33]</td>
<td>13,785 [±33]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ECONOMIC OUTPUTS</th>
<th>YEAR</th>
<th>SCENARIO 0–BASELINE</th>
<th>SCENARIO 1</th>
<th>SCENARIO 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cumulative direct costs avoided (millions, TT$)</td>
<td>2025</td>
<td>NA</td>
<td>15.46 [±24.1]</td>
<td>27.81 [±24.1]</td>
</tr>
<tr>
<td></td>
<td>2035</td>
<td>NA</td>
<td>155.46 [±44.41]</td>
<td>254.73 [±44.39]</td>
</tr>
</tbody>
</table>
Cases of Disease per Year

Table 7 presents the annual cases for each disease by year.

Table 7: Cases per year, total population

<table>
<thead>
<tr>
<th>YEAR</th>
<th>CHD</th>
<th>COPD</th>
<th>LUNG CANCER</th>
<th>STROKE</th>
<th>TOTAL</th>
</tr>
</thead>
</table>

Cumulative Cases of Disease

Table 8 presents the cumulative cases for each disease by year, and table 9 presents the cumulative cases avoided. By 2035, the cumulative cases avoided for scenario 1 compared to scenario 0 are 239, 452, 120,823 respectively for CHD, COPD, lung cancer and stroke.

Similarly, by 2035, the cumulative cases avoided for scenario 1 compared to scenario 0 are 359, 691, 199, and 1,288 respectively for CHD, COPD, lung cancer and stroke.

Table 8: Cumulative cases for each disease by year, total population

<table>
<thead>
<tr>
<th>YEAR</th>
<th>CHD</th>
<th>COPD</th>
<th>LUNG CANCER</th>
<th>STROKE</th>
<th>TOTAL</th>
</tr>
</thead>
</table>
Mortality

Table 10 and table 11 present the mortality cases for scenarios 1 and 2, relative to scenario 0. Note that it was not possible to derive premature mortality costs due to the non-significance of the results (this is explained in the discussion section). Relative to scenario 0, by 2035, 53 and 93 mortality cases are avoided for scenarios 1 and 2 respectively.

Table 9: Cumulative cases avoided relative to scenario 0 for the total Trinidad and Tobago population by 2025 and 2035

<table>
<thead>
<tr>
<th>SCENARIO 1</th>
<th>CHD</th>
<th>COPD</th>
<th>LUNG CANCER</th>
<th>STROKE</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>2035</td>
<td>239 [±133]</td>
<td>452 [±93]</td>
<td>120 [±40]</td>
<td>823 [±93]</td>
<td>1,634 [±191]</td>
</tr>
<tr>
<td>SCENARIO 2</td>
<td>CHD</td>
<td>COPD</td>
<td>LUNG CANCER</td>
<td>STROKE</td>
<td>TOTAL</td>
</tr>
</tbody>
</table>

Table 10: Mortality cases in the total population per year

<table>
<thead>
<tr>
<th>YEARS</th>
<th>SCENARIOS</th>
<th>TOTALS</th>
</tr>
</thead>
<tbody>
<tr>
<td>2025</td>
<td>Scenario 0</td>
<td>9,894 [±53]</td>
</tr>
<tr>
<td></td>
<td>Scenario 1</td>
<td>9,867 [±53]</td>
</tr>
<tr>
<td></td>
<td>Scenario 2</td>
<td>9,880 [±53]</td>
</tr>
<tr>
<td>2035</td>
<td>Scenario 0</td>
<td>11,872 [±66]</td>
</tr>
<tr>
<td></td>
<td>Scenario 1</td>
<td>11,819 [±66]</td>
</tr>
<tr>
<td></td>
<td>Scenario 2</td>
<td>11,780 [±53]</td>
</tr>
</tbody>
</table>
Mortality cases avoided

Table 11: Mortality Cases avoided in the total population per year

<table>
<thead>
<tr>
<th>YEARS</th>
<th>SCENARIOS</th>
<th>TOTALS</th>
</tr>
</thead>
<tbody>
<tr>
<td>2025</td>
<td>Scenario 1 – Scenario 0</td>
<td>27 [±53.12]</td>
</tr>
<tr>
<td>2025</td>
<td>Scenario 2 – Scenario 0</td>
<td>13 [±53.12]</td>
</tr>
<tr>
<td>2035</td>
<td>Scenario 1 – Scenario 0</td>
<td>53 [±66.4]</td>
</tr>
<tr>
<td>2035</td>
<td>Scenario 2 – Scenario 0</td>
<td>93 [±66.4]</td>
</tr>
</tbody>
</table>

Direct Cumulative Costs Avoided

Table 12 presents the cumulative direct healthcare costs avoided for scenarios 1 and 2, relative to scenario 0. By 2035, scenario 1 results in costs avoided of TT$155 million compared to scenario 0; similarly, by 2035, scenario 2 results in TT$255 million in direct costs avoided compared to scenario 0.

Table 12: Direct cumulative healthcare costs avoided (million, TT$)

<table>
<thead>
<tr>
<th>YEAR</th>
<th>CHD</th>
<th>COPD</th>
<th>LUNG CANCER</th>
<th>STROKE</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>2025</td>
<td>Scenario 1 relative to Scenario 0</td>
<td>-0.97 [±21.15]</td>
<td>0.78 [±7.29]</td>
<td>0.82 [±1.6]</td>
<td>14.82 [±8.83]</td>
</tr>
<tr>
<td>2035</td>
<td>Scenario 1 relative to Scenario 0</td>
<td>24.01 [±38.75]</td>
<td>25.09 [±13.52]</td>
<td>4.3 [±2.93]</td>
<td>102.07 [±16.69]</td>
</tr>
<tr>
<td>2035</td>
<td>Scenario 2 relative to Scenario 0</td>
<td>49.05 [±38.75]</td>
<td>40.17 [±13.51]</td>
<td>8.86 [±2.92]</td>
<td>156.65 [±16.67]</td>
</tr>
</tbody>
</table>
TRINIDAD AND TOBAGO’S EXISTING PREVENTION AND CONTROL INTERVENTIONS ARE IN URGENT NEED OF STRENGTHENING IF THE COUNTRY IS TO REDUCE ITS TOBACCO USE. TOBACCO TAXATION HAS BEEN
DISCUSSION

This study explored the impact of two tobacco tax increase scenarios in Trinidad and Tobago on the future burden of four smoking-related diseases up to 2035. The results showed that small changes in smoking prevalence in one year can have relatively large impacts in terms of disease into the future. The results showed that scenario 2 (the higher tax increase) yielded more significant results, supporting the “go big and go fast” approach outlined in World Bank report Tobacco Tax Reform – A Multisectoral Perspective: At the Crossroads of Health And Development (World Bank 2018). This publication suggested that “tax strategies should focus on health gains first, then on fiscal benefits. This means going for big tobacco excise tax rate increases starting early in the process.”

The study included just four smoking-related diseases (CHD, COPD, stroke, lung cancer). However, we know that smoking is responsible for many more diseases, and harms almost every organ in the body (Centers for Disease Control 2016). Therefore, we are likely to see much wider epidemiological benefits than those observed here. Future work could update this study by including additional smoking-related diseases that would also increase the impact of the scenarios on epidemiologic outputs such as cumulative incidence, incidence, and mortality. The relatively small number of premature deaths avoided can be explained by a Danish study (with 15 years follow-up of a large cohort of smokers) that found no evidence that heavy smokers who reduced their number of cigarettes had a lower risk of death from CHD or from all causes (Gotfredsen and others 2003).

It was not possible to derive premature mortality costs due to the non-significance of the results. There are three possible explanations for this: first, the threshold age of premature mortality was assumed to be 65 years, which is very close to life expectancy in Trinidad and Tobago. Second, the number of mortality cases avoided (by diseases studied) is small (see table 10). Consequently, given that premature mortality is a subset of mortality, premature mortality costs would be expected to be small and not significantly different. Third, the relatively small change in smoking prevalence due to the intervention and the relatively short time span of the simulation will play a part.

While the microsimulation method is advantageous in NCD modeling, a key disadvantage is that the model is data intensive and required detailed and up-to-date data. Unfortunately, it was not possible to include data on indirect costs (such as productivity losses) because the data by disease were not available (as is the case in many countries). However, we know from studies of total costs (Rezaei and others 2016) in Trinidad and Tobago that the relative non-health cost is colossal due to premature deaths and lost productivity. For example, the American Cancer Society’s Tobacco Atlas indicates that the
Reducing Tobacco Use Through Taxation in Trinidad and Tobago: Modelling the Long-Term Health and Economic Impact

The total direct and indirect cost of smoking in Trinidad and Tobago totals TT$1,858 million (US$275 million) (WHO 2002). One systematic review estimated the direct costs of smoking to equal around 1.5–6.8 percent of national health system expenditures and 0.22–0.88 percent of GDP in the country studied (Rezaei and others 2016).

In a study in Ukraine, using the same model, the premature mortality costs avoided of increasing tobacco tax was estimated at Hrv 16.5 billion (US$695 million) (Webber and others 2017), and in the UK, increasing the tobacco duty escalator to 5 percent (from an annual tobacco tax increase of 2%) would avoid £192 million in indirect costs by 2035 alone (Knuchel-Takano and others 2017). Therefore, wider societal costs such as losses in productivity are likely to be higher than direct costs, making a stronger case for the implementation of regular tax hikes for tobacco control (Action on Smoking and Health 2015). In fact, whereas the indirect cost of tobacco use has not been assessed in Trinidad and Tobago, in 2016 RTI International (RTI 2016) estimated annual indirect costs (in terms of productivity losses) for smoking-related cancers and diabetes to be TT$1.18 billion and TT$2.32 billion respectively. According to the study, the estimated costs represent 90 percent and 58 percent, respectively of the total economic burden of cancer and diabetes.

Furthermore, it was projected that sustained prevention and control efforts can result in significant savings in terms of productivity losses. These costs are hardly comparable to most other studies, since they only include productivity losses while excluding other aspects such as morbidity costs and costs of premature retirement. Nevertheless, as noted by Rezaei et al. (Rezaei and others 2016), indirect costs exceeding direct costs is not an uncommon occurrence. Of the 14 studies reviewed Rezaei et al., seven reported substantially higher indirect costs, ranging from 53.3 percent to 81 percent of the total costs of smoking. If indirect cost data by disease becomes available, then the model can once again easily be updated in the future.

Another data limitation was the lack of trend data on smoking prevalence. Therefore, only a static trend could be included. This may overestimate the impacts if smoking prevalence is actually falling, or underestimate the impact if smoking prevalence is actually increasing.

We know that social groups react differently to tax increases (Krasovsky 2013). Due to small sample sizes, it was not possible to model the long-term health impacts on different social groups within the microsimulation.

One specific limitation of any predictive model is that it does not take account of major future changes in circumstances, such as the behavior of the tobacco industry, or the introduction of new drugs or technologies. In theory, their effects can be estimated by altering parameters in the model, but these will significantly increase the degrees of
uncertainty. However, they could be simulated as different scenarios in the future relative to a "no change" scenario.

At present, the model does not take account of multi-morbidity and the joint effect of several risk factors on disease occurrence and related mortality. However, individuals can get more than one smoking-related disease in their lifetime. Future work could expand the scope of the model to take account of technological and economic changes and their potential effects, and also to model the clustering of risk factors and diseases in the same individuals.

The model did not take account of passive smoking/secondhand smoke. Understanding the combined risk of smoking and passive smoking on later disease outcomes will enable us to model the combined impact of these risk factors on later disease outcomes.

It was beyond the scope of this study, given the time constraints, to carry out an in-depth uncertainty and sensitivity analysis. We are aware that this is good practice; however, there is a lack of validated datasets with which to compare our outputs. Furthermore, the microsimulation is complex, relative to spreadsheet models, for example. It involves many thousands of calculations which are completed during the simulation of 50 million individuals. Given this complexity, local uncertainty analysis would demand many thousands of consecutive runs and would require a supercomputer to complete the exercise in a realistic timescale.

This study complements modelling work done by the World Bank Group and the HEU, Centre for Health Economics of The University of the West Indies, and shows the health and related economic benefits of increasing tobacco taxes in Trinidad and Tobago. Even small reductions in smoking prevalence in one year will have long-term impacts on disease incidence and subsequent health costs.
REFERENCES


Team UNC. Common Country Assessment: Republic of Trinidad and Tobago. 2006.


UK Health Forum and Cancer Research UK (CRUK). 2016. *Aiming High: Why the UK should aim to be tobacco-free.* London: UK Health Forum and CRUK.


References


TRINIDAD AND TOBAGO’S EXISTING PREVENTION AND CONTROL INTERVENTIONS ARE IN URGENT NEED OF STRENGTHENING IF THE COUNTRY IS TO REDUCE ITS TOBACCO USE. TOBACCO TAXATION HAS BEEN
APPENDIX A: MICROSIMULATION MODEL

The microsimulation consists of two modules. The first module calculates the predictions of risk factor trends over time based on data from rolling cross-sectional studies. The second module performs the microsimulation of a virtual population, generated with demographic characteristics matching those of the observed data. The health trajectory of each individual from the population is simulated over time allowing them to contract, survive or die from a set of diseases or injuries related to the analysed risk factors. The detailed description of the two modules is presented below.

**Module One: Predictions of Smoking Prevalence Over Time**

For the risk factor (RF), let \( N \) be the number of categories for a given risk factor, e.g. \( N = 3 \) for smoking. Let \( k = 1, 2, \ldots, N \) number these categories and \( p_k(t) \) denote the prevalence of the RF that corresponds to the category \( k \) at time \( t \). We estimate \( p_k(t) \) using multinomial logistic regression model with prevalence of RF category \( k \) as the outcome, and time \( t \) as a single explanatory variable. For \( k\leq N \), we have

\[
\ln \left( \frac{p_k(t)}{p_i(t)} \right) = \beta_k + \beta_i^t
\]  

(0.1)

The prevalence of the first category is obtained by using the normalization constraint \( \sum_{k=1}^{N} p_k(t) = 1 \). Solving equation (0.0) for \( p_k(t) \), we obtain

\[
p_k(t) = \frac{\exp(\beta_k^t + \beta_i^t)}{1 + \sum_{i=1}^{N-1} \exp(\beta_i^t + \beta_i^t)},
\]

(0.2)

which respects all constraints on the prevalence values, i.e. normalization and \([0, 1]\) bounds.

**Multinomial Logistic Regression for Smoking Prevalence**

Measured data consist of sets of probabilities, with their variances, at specific time values (typically the year of the survey). For any particular time the sum of these probabilities is unity. Typically such data might be the probabilities of smoker, ex-smoker, never smokers as they are extracted from the survey data set. Each data point is treated as a normally
The regression equations are most easily derived from a familiar least square minimization. In the interpretable as a probability – a real number lying between 0 and 1. of the probabilities is equal to 1 and a Gaussian distribution is assumed for all functions.

The regression consists of fitting a set of logistic functions \( \{p_k(a, b, t) | k \in [0, K-1] \} \) to these data – one function for each \( k \)-value. At each time value the sum of these functions is unity. Thus, for example, when measuring smoking in the three states already mentioned, the \( k = 0 \) regression function represents the probability of being a never smoker over time, \( k = 1 \) the probability of being an ex-smoker and \( k = 2 \) the probability of being a smoker.

The regression equations are most easily derived from a familiar least square minimization. In the following equation set the weighted difference between the measured and predicted probabilities is written as \( S \); the logistic regression functions \( p_k(a, b, t) \) are chosen to be ratios of sums of exponentials (this is equivalent to modelling the log probability ratios, \( p_i/p_b \) as linear functions of time).

\[
S(a, b) = \frac{1}{2} \sum_{k=0}^{K-1} \sum_{i=0}^{N-1} \left( \frac{p_k(a, b, t_i) - \mu_k}{\sigma^2_k} \right)^2
\]

\[
p_k(a, b, t) = \frac{e^{tb} \exp(a)}{1 + e^{tb} \exp(a)}
\]

\[
a = (a_0, a_1, ..., a_{K-1}), \quad b = (b_0, b_1, ..., b_{K-1})
\]

\[
A_0 = 0, \quad A_i = a_i + b_it
\]

The parameters \( A \), \( a_0 \) and \( b_0 \) are all zero and are used merely to preserve the symmetry of the expressions and their manipulation. For a \( K \)-dimensional set of probabilities there will be \( 2(K-1) \) regression parameters to be determined.

For a given dimension \( K \) there are \( K-1 \) independent functions \( p_k \) with the following criteria: the sum of the probabilities is equal to 1 and a Gaussian distribution is assumed for all functions.

Note that the parameterization ensures that the necessary requirement that each \( p_k \) be interpretable as a probability – a real number lying between 0 and 1.

---

4 Depending on the circumstances this assumption will be more or less accurate and more or less necessary. In general, it is both extremely useful and accurate. For simple surveys the individual Bayesian prior and posterior probabilities are Beta distributions – the likelihood being binomial. For reasonably large samples, the approximation of the beta distributions by normal distributions is both legitimate and a practical necessity. For complex, multi-PSU, stratified surveys, it is again assumed that these base probabilities are approximately normally distributed and, again, it is an assumption that makes the analysis tractable. Depending on the nature of the raw data set it may be possible to use non-parametric statistical methods for this analysis.
The minimum of the function $S$ is determined from the equations

$$\frac{\partial S}{\partial a_j} = \frac{\partial S}{\partial b_j} = 0 \quad \text{for } j=1,2,\ldots,k-1$$

(0.5)

noting the relations

$$\frac{\partial p_j}{\partial A_j} = \frac{\partial}{\partial A_j} \left( \frac{e^k}{1 + e^k + \ldots + e^{k+1}} \right) = p_j \delta_{ij} - p_i p_j$$

$$\frac{\partial}{\partial a_j} = \frac{\partial}{\partial A_j}$$

$$\frac{\partial}{\partial b_j} = I$$

(0.6)

The values of the vectors $a, b$ that satisfy these equations are denoted $\hat{a}, \hat{b}$. They provide the trend lines $p_k(\hat{a}, \hat{b}, t)$ for the separate probabilities. The confidence intervals for the trend lines are derived most easily from the underlying Bayesian analysis of the problem.

**Bayesian Interpretation**

The $2K-2$ regression parameters $(a, b)$ are regarded as random variables whose posterior distribution is proportional to the function $\text{exp}(-S(a, b))$. The maximum likelihood estimate of this probability distribution function, the minimum of the function $S$, is obtained at the values $\hat{a}, \hat{b}$. Other properties of the $(2K-2)$-dimensional probability distribution function are obtained by first approximating it as a $(2K-2)$-dimensional normal distribution whose mean is the maximum likelihood estimate. This amounts to expanding the function $S(a, b)$ in a Taylor series as far as terms quadratic in the differences $(a - \hat{a}), (b - \hat{b})$ about the maximum likelihood estimate $\hat{S} = S(\hat{a}, \hat{b})$. Hence

$$S(a, b) = \frac{1}{2} \sum_{i=0}^{2K-2} \sum_{j=0}^{2K-2} \left( \frac{p_i(a, b; t) - \hat{p}_i}{\sigma_i^2} \right)^2$$

$$= S(\hat{a}, \hat{b}) + \frac{1}{2} (a - \hat{a}, b - \hat{b}) P^{-1} (a - \hat{a}, b - \hat{b}) + ...$$

$$\approx S(\hat{a}, \hat{b}) + \frac{1}{2} \sum_{i,j} (a_i - \hat{a}_i) \frac{\partial^2 S}{\partial \hat{a}_i \partial \hat{a}_j} (a_j - \hat{a}_j) + \frac{1}{2} \sum_{i,j} (a_i - \hat{a}_i) \frac{\partial^2 S}{\partial \hat{a}_i \partial \hat{b}_j} (b_j - \hat{b}_j) +$$

$$+ \frac{1}{2} \sum_{i,j} (b_i - \hat{b}_i) \frac{\partial^2 S}{\partial \hat{b}_i \partial \hat{a}_j} (a_j - \hat{a}_j) + \frac{1}{2} \sum_{i,j} (b_i - \hat{b}_i) \frac{\partial^2 S}{\partial \hat{b}_i \partial \hat{b}_j} (b_j - \hat{b}_j)$$

(0.7)

The $(2K-2)$-dimensional covariance matrix $P$ is the inverse of the appropriate expansion coefficients. This matrix is central to the construction of the confidence limits for the trend lines.
Estimation of Confidence Intervals

The logistic regression functions $p_k(t)$ can be approximated as a normally distributed time-varying random variable $N(\hat{p}_k(t), \sigma^2_k(t))$ by expanding $p_k$ about its maximum likelihood estimate (the trend line)

$$p_k(a, b, t) = p(a, b, t) \approx \hat{p}_k(t) + \left( \nabla_2, \nabla_1 \right) \hat{p}_k(t) \begin{pmatrix} a - \hat{a} \\ b - \hat{b} \end{pmatrix} + \ldots$$

(0.8)

Denoting mean values by angled brackets, the variance of $p_k$ is thereby approximated as

$$\sigma^2_k(t) = \left\{ (p_k(a, b, t) - \hat{p}_k(t))^2 \right\} = \left( \nabla_2 \hat{p}_k(t), \nabla_1 \hat{p}_k(t) \right) \begin{pmatrix} a - \hat{a} \\ b - \hat{b} \end{pmatrix}^T \begin{pmatrix} a - \hat{a} \\ b - \hat{b} \end{pmatrix} \times \left( \nabla_2 \hat{p}_k(t), \nabla_1 \hat{p}_k(t) \right)^T$$

(0.9)

When $K=3$ this equation can be written as the 4-dimensional inner product

$$\sigma^2_k(t) = \left( \frac{\hat{p}_k(t)}{\partial a_k} \frac{\hat{p}_k(t)}{\partial b_k} \frac{\hat{p}_k(t)}{\partial a_k} \frac{\hat{p}_k(t)}{\partial b_k} \right) \begin{pmatrix} P_{a11} & P_{a12} & P_{a13} & P_{a14} \\ P_{a21} & P_{a22} & P_{a23} & P_{a24} \\ P_{a31} & P_{a32} & P_{a33} & P_{a34} \\ P_{a41} & P_{a42} & P_{a43} & P_{a44} \end{pmatrix} \begin{pmatrix} \hat{p}_k(t) \\ \hat{p}_k(t) \\ \hat{p}_k(t) \\ \hat{p}_k(t) \end{pmatrix} \begin{pmatrix} \partial a_k \\ \partial b_k \\ \partial a_k \\ \partial b_k \end{pmatrix}$$

(0.10)

where $P_{a(i)} = \left\{ (c_i, -\hat{c}_i, \hat{c}_i, -\hat{c}_i) \right\}$. The 95% confidence interval for $p_k(t)$ is centred given as

$$[\hat{p}_k(t) - 1.96\sigma_k(t), p_k(t) + 1.96\sigma(t)].$$

Module Two: Microsimulation Initialization – Birth, Disease and Death Models

Simulated people are generated with the correct demographic statistics in the simulation's start-year. In this year women are stochastically allocated the number and years of birth of their children – these are generated from known fertility and mother's age at birth statistics (valid in the start-year). If a woman has children then those children are generated as members of the simulation in the appropriate birth year.

The microsimulation is provided with a list of relevant diseases. These diseases used the best available incidence, mortality, survival, relative risk and prevalence statistics (by age
and sex). Individuals in the model are simulated from their year of birth (which may be before the start year of the simulation). In the course of their lives, simulated people can die from one of the diseases caused by smoking that they might have acquired or from some other cause. The probability that a person of a given age and sex dies from a cause other than the disease are calculated in terms of known death and disease statistics valid in the start-year. It is constant over the course of the simulation. The survival rates from tobacco-related diseases will change as a consequence of the changing distribution of smoking level in the population.

The microsimulation incorporates a sophisticated economic module. The module employs Markov-type simulation of long-term health benefits, health care costs and cost-effectiveness of specified interventions. It synthesizes and estimates evidence on cost-effectiveness analysis and cost-utility analysis. The model can be used to project the differences in quality-adjusted life years, direct and indirect lifetime health care costs, and as a consequence of interventions, incremental cost effectiveness ratios over a specified time scale. Outputs can be discounted for any specific discount rate. This section provides an overview of the initialization of the microsimulation model and will be expanded upon in the next sections.

**Population Models**

Populations are implemented as instances of the TPopulation C++ class. The TPopulation class is created from a population (*.ppl) file. Usually a simulation will use only one population but it can simultaneously process multiple populations (for example, different ethnicities within a national population).

**Population Editor**

The Population Editor allows editing and testing of TPopulation objects. The population is created in the start-year and propagated forwards in time by allowing females to give birth. An example population pyramid which can be used when initializing the model is shown in Figure A1 source not found.. It shows the Ukrainian population distribution (2015) used in the initialization of the model.
People within the model can die from specific diseases or from other causes. A disease file is created within the program to represent deaths from other causes. The following distributions are required by the population editor (table A1).

### Table A1: Summary of the parameters representing the distribution component

<table>
<thead>
<tr>
<th>DISTRIBUTION NAME</th>
<th>SYMBOL</th>
<th>NOTE</th>
</tr>
</thead>
<tbody>
<tr>
<td>MalesByAgeByYear</td>
<td>$p_m(a)$</td>
<td>Input in year$_0$ – probability of a male having age $a$</td>
</tr>
<tr>
<td>FemalesByAgeByYear</td>
<td>$p_f(a)$</td>
<td>Input in year$_0$ – probability of a female having age $a$</td>
</tr>
<tr>
<td>BirthsByAgeofMother</td>
<td>$p_b(a)$</td>
<td>Input in year$_0$ – conditional probability of a birth at age $a$ if the mother gives birth.</td>
</tr>
<tr>
<td>NumberOfBirths</td>
<td>$p_\lambda(n)$</td>
<td>Poisson distribution, probability of giving birth to $n$ children</td>
</tr>
</tbody>
</table>

### Birth Model

Any female in the child bearing years ($\text{AgeAtChild.lo}$, $\text{AgeAtChild.hi}$) is deemed capable of giving birth. The number of children, $n$, that she has in her life is dictated by the Poisson distribution $p_\lambda(n)$, where the mean of the Poisson distribution is the Total Fertility Rate (TFR) parameter.

The probability that a mother (who does give birth) gives birth to a child at age $a$ is determined from the BirthsByAgeOfMother distribution as $p_b(a)$. For any particular mother, the births of multiple children are treated as independent events, so that the

---

*This could be made to be time dependent; in the baseline model it is constant.*
probability that a mother who produces N children produces n of them at age a is given as the Binomially distributed variable,

\[ p_n(a | N) = \frac{N!}{n!(N-n)!} \left(p_m(a)\right)^n \left(1 - p_m(a)\right)^{N-n} \tag{0.11} \]

The probability that the mother gives birth to n children at age a is

\[ p_n(a) = e^{-\lambda} \sum_{N=n}^{\infty} \frac{\lambda^N}{N!} p_m(a)^n \left(1 - p_m(a)\right)^{N-n} \tag{0.12} \]

Performing the summation in this equation gives the simplifying result that the probability \( p_n(a) \) is itself Poisson distributed with mean parameter \( \lambda p_m(a) \),

\[ p_n(a) = e^{-\lambda p_m(a)} \frac{(\lambda p_m(a))^n}{n!} = p_{\lambda p_m(a)}(n) \tag{0.13} \]

Thus, on average, a mother at age a will produce \( \lambda p_m(a) \) children in that year.

The gender of the children\(^{10}\) is determined by the probability, \( p_m = \frac{1}{2} \cdot p_{male} + \frac{1}{2} \cdot p_{female} \). In the baseline model this is taken to be the probability \( N_{male}/(N_{male}+N_{female}) \).

The Population editor menu item Population Editor\Tools\Births\Show random birthList creates an instance of the TPopulation class and uses it to generate and list a (selectable) sample of mothers and the years in which they give birth.

**Deaths from Modelled Diseases**

The simulation models any number of specified diseases – some of which may be fatal. In the start year the simulation’s death model uses the diseases’ own mortality statistics to adjust the probabilities of death by age and gender. In the start year the net effect is to maintain the same probability of death by age and gender as before; in subsequent years, however, the rates at which people die from modelled diseases will change as modelled risk factors change. The population dynamics sketched above will be only an approximation to the simulated population’s dynamics. The latter will be known only on completion of the simulation.

---

\(^{10}\) The probability of the sex of a child can be made time dependent.
The Risk Factor Model

The distribution of risk factors (RF) in the population is estimated using regression analysis stratified by both sex $S = \{\text{male}, \text{female}\}$ and age group $A = (0-9, 10-19, ..., 70-79, 80+)$. The fitted trends are extrapolated to forecast the distribution of each RF category in the future. For each sex-and-age-group stratum, the set of cross-sectional, time-dependent, discrete distributions $D = \{p_k(t) | k = 1, ..., N; t > 0\}$, is used to manufacture RF trends for individual members of the population.

We model different risk factors, some of which are continuous (such as BMI) and some are categorical (smoking status).

Categorical Risk Factors

Smoking is the categorical risk factor. Each individual in the population may belong to one of the three possible smoking categories: \{never smoked, ex-smoker, smoker\} with their probabilities \{$p'_n, p_e, p_s$\}. These states are updated on receipt of the information that the person is either a smoker or a non-smoker. They will be a never-smoker or an ex-smoker depending on their original state (an ex-smoker can never become a never-smoker).

\[
\begin{align*}
\text{(never smoked)} & \rightarrow \text{(never smoked, smoker)} \\
\text{(ex-smoker)} & \rightarrow \text{(ex-smoker, smoker)} \\
\text{(smoker)} & \rightarrow \text{(ex-smoker, smoker)}
\end{align*}
\]

When the probability of being a smoker is $p$ the allowed transitions are summarized in the state update equation:

\[
\begin{bmatrix}
    p_0 \\
    p_1 \\
    p_2
\end{bmatrix} =
\begin{bmatrix}
    1-p & 0 & 0 \\
    0 & 1-p & 1-p \\
    p & p & p
\end{bmatrix}
\begin{bmatrix}
    p_0 \\
    p_1 \\
    p_2
\end{bmatrix} \tag{0.14}
\]

After the final simulation year the smoking trajectories are completed until the person’s maximum possible age of 110 by supposing that their smoking state stays fixed. The life expectancy calculation is equal to the sum of the probabilities of being alive in each possible year of life.

In the initial year of the simulation, a person may be in one of the three smoking categories; after $N$ updates there will be $3 \times 2^N$ possible trajectories. These trajectories will each have a calculated probability of occurring; the sum of these probabilities is 1.

In each year the probability of being a smoker or a non-smoker will depend on the forecast smoking scenario which provides exactly that information. Note that these states are two-dimensional and cross-sectional \{non-smoking, smoking\}, and they are turned into three-dimensional states \{never smoked, ex-smoker, smoker\} as described above. The time...
The evolution of the three dimensional states are the smoking trajectories necessary for the computation of disease table disease and death probabilities.

**Smoking**

The microsimulation model applied to smoking enables us to measure the future health impact of changes in rates of tobacco consumption. This includes the impact of giving up smoking on the following diseases: a) COPD; b) CHD (or acute myocardial infarction (AMI) if CHD data are not available); c) stroke; and d) lung cancer. In the simulation each person is categorized into one of the three smoking groups: smokers, ex-smokers and people who have never smoked. Their initial distribution is based on the distribution of smokers, ex-smokers and never smokers from published data.

During the simulation a person may change smoking states and their relative risk will change accordingly. Relative risks associated with smokers and people who have never smoked have been collected from published data. The relative risks (RR) associated with ex-smokers \( \text{RR}_{\text{ex-smoker}} \) are related to the relative risk of smokers \( \text{RR}_{\text{smoker}} \). The ex-smoker relative risks are assumed to decrease over time with the number of years since smoking cessation \( T_{\text{cessation}} \). These relative risks are computed in the model using equations 1.19 and 1.20 (Hoogenveen and others 2008).

\[
\text{RR}_{\text{ex-smoker}}(A,S,T_{\text{cessation}}) = 1 + (\text{RR}_{\text{smoker}}(A,S) - 1) \exp(-\gamma(A)T_{\text{cessation}}) \\
\gamma(A) = \gamma_0 \exp(-\eta A)
\]

where \( \gamma \) is the regression coefficient of time dependency. The constants \( \gamma_0 \) and \( \eta \) are intercept and regression coefficient of age dependency, respectively, which are related to the specified disease (see Table A2).

**Table A2: Parameter estimates for \( \gamma_0 \) and \( \eta \) related to each disease**

<table>
<thead>
<tr>
<th>DISEASE</th>
<th>( \gamma_0 )</th>
<th>( \eta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMI</td>
<td>0.24228</td>
<td>0.05822</td>
</tr>
<tr>
<td>Stroke</td>
<td>0.31947</td>
<td>0.01648</td>
</tr>
<tr>
<td>COPD</td>
<td>0.20333</td>
<td>0.03087</td>
</tr>
<tr>
<td>Lung cancer</td>
<td>0.15637</td>
<td>0.02065</td>
</tr>
</tbody>
</table>

Source: Hoogenveen and others 2008.
However, a minimum exists when the cessation time is equal to $\eta^{-1}$. The minimum value was calculated by the method detailed in equations (1.17), (1.18) and (1.19). Where time, $t$ is equal to the age, $A$ of an individual.

$$\rho_{\text{ex-smok}}(t) = 1 + (\rho_{\text{smok}} - 1) f(t)$$

$$f(t) = \exp(-\gamma_0(t-t_0)\exp(-\eta t))$$

$$f'(t) = -\gamma_0 f(t)e^{-\eta t}(\eta(t-t_0)+1)$$

The function $f(t)$ has the following properties:

$$f(t_0) = 1$$

$$f'(t_0) = -\gamma_0 e^{\eta t_0}$$

$$f(t) \text{ has a minimum at } t = t_0 + \eta^{-1}$$

$$f(\infty) = A$$

In order to avoid the $RR_{\text{ex-smoker}}$ from increasing, the cessation time was set equal to $\eta^{-1}$ when the cessation time was greater than $\eta^{-1}$ (see equation (1.20)).

$$RR_{\text{ex-smoker}}(A, S, T_{\text{cessation}}) = \begin{cases} 1 + (RR_{\text{smoker}}(A, S) - 1) \exp(-\gamma(A)T_{\text{cessation}}) & T_{\text{cessation}} < \eta^{-1} \\ 1 + (RR_{\text{smoker}}(A, S) - 1) \exp(-\gamma(A)\eta^{-1}) & T_{\text{cessation}} \geq \eta^{-1} \end{cases}$$

$$\gamma(A) = \gamma_0 \exp(-\eta A)$$

The ex-smoker relative risks as a function of time after smoking cessation were plotted in Figure A3 for AMI, stroke, COPD, and lung cancer.

**Figure A3: Ex-smoker relative risks as a function of time after smoking cessation**
Relative Risks

The reported incidence risks for any disease make no reference to any underlying risk factor. The microsimulation requires this dependence to be made clear. The risk factor dependence of disease incidence has to be inferred from the distribution of the risk factor in the population (here denoted as \( \pi \)); it is a disaggregation process:

Suppose that \( \alpha \) is a risk factor state of risk factor \( A \) and denote by \( p_d(d|\alpha,a,s) \) the incidence probability for the disease \( d \) given the risk state \( \alpha \), the person’s age, \( a \), and gender, \( s \). The relative risk \( \rho_{\alpha} \) is defined by equation (1.22).

\[
\rho_{\alpha}(d|\alpha,a,s) = \frac{p_d(d|\alpha,a,s)}{p_d(d|\alpha_0,a,s)}
\]

Where \( \alpha_0 \) is the zero risk state (for example, the moderate state for alcohol consumption).

The incidence probabilities, as reported, can be expressed in terms of the equation,

\[
p(d|a,s) = \sum_{\alpha} p_d(d|\alpha,a,s)\pi_{\alpha}(\alpha|a,s)
\]

Combining these equations allows the conditional incidence probabilities to be written in terms of known quantities

\[
p(d|\alpha,a,s) = \rho_{\alpha}(\alpha|a,s)\frac{p(d|\alpha,s)}{\sum_{\beta} \rho_{\beta}(\beta|a,s)\pi_{\beta}(\beta|a,s)}
\]

Previous to any series of Monte Carlo trials, the microsimulation program pre-processes the set of diseases and stores the calibrated incidence statistics \( p_d(d|\alpha_0, a, s) \)

Modelling Diseases

Disease modelling relies heavily on the sets of incidence, mortality, survival, relative risk and prevalence statistics.

In the simulation, individuals are assigned a risk factor trajectory giving their personal risk factor history for each year of their lives. Their probability of getting a particular risk factor-related disease in a particular year will depend on their risk factor state in that year.

Once a person has a fatal disease (or diseases), their probability of survival will be controlled by a combination of the disease-survival statistics and the probabilities of dying from other causes. Disease survival statistics are modelled as age- and gender-dependent exponential distributions.
Methods for Approximating Missing Disease Statistics

A large amount data are required for modelling these diseases. Where possible these datasets have been collected from published sources or analysed from either cross-sectional or longitudinal datasets. Another limitation is that often this data needs to be in a specific format. For example, the model updates individuals’ disease status every year so the relative risks used in the model need to be annual relative risks.

This section contains the methods used in this project in cases where data for a particular disease were unavailable.

Terminal and non-terminal single state disease incidence from prevalence

For terminal diseases, to estimate incidence (knowing prevalence and mortality rates) one can proceed by finding those incidence probabilities that minimize the distance between the known \( \bar{p}_{\text{pre}} \) and computed prevalence \( p^*_k \).

Non-terminal diseases are treated in a similar way – although, obviously, the mortality probabilities are zero.

Mortality statistics

In any year, in a sample of \( N \) people who have the disease, a subset \( N_w \) will die from the disease. Mortality statistics record the cross-sectional probabilities of death as a result of the disease – possibly stratifying by age

\[
p_w = \frac{N_w}{N}
\]  

(0.25)

Within such a subset \( N_w \) of people that die in that year from the disease, the distribution by year-of-disease is not usually recorded. This distribution would be most useful. Consider two important idealized, special cases.

Suppose the true probabilities of dying in the years after some age \( a_0 \) are \( \{p_{a0}, p_{a1}, p_{a2}, p_{a3}, p_{a4}\} \)

The probability of being alive after \( N \) years is simply that you do not die in each year

\[
p_{\text{survive}}(a_0+N) = (1-p_{a0})(1-p_{a1})(1-p_{a2})\cdots(1-p_{aN-1})
\]  

(0.26)

Survival models

There are three in use (they are easily extended if the data merit):

Survival model 0: a single probability of dying \( \{p_{a0}\} \)

\( p_{a0} \) is valid for all years.

Survival model 1: two different probabilities of dying \( \{p_{a0}, p_{a1}\} \)

\( p_{a0} \) is valid for the first year; \( p_{a1} \) thereafter.
Survival model 2: three different probabilities of dying \( \{P_{\text{a0}}, P_{\text{a1}}, P_{\text{a5}}\} \)

\( P_{\text{a0}} \) is valid for the first year; \( P_{\text{a1}} \) for the second to the fifth year; \( P_{\text{a5}} \) thereafter.

Remember that different probabilities will apply to different age and gender groups. Typically the data might be divided into 10-year age groups.

**Calculating survival from incidence and mortality**

When a person (of a given gender) dies from a disease, they must have contracted it at some earlier age. For survival model 2, this is expressed

\[
\hat{p}_{\text{mortality}}(a) = P_{\text{a0}}(a-1)P_{\text{a0}} + \\
+ P_{\text{a0}}(a-2)(1-P_{\text{a0}})P_{\text{a0}} + \\
+ P_{\text{a0}}(a-3)(1-P_{\text{a0}})(1-P_{\text{a1}})P_{\text{a1}} + \\
+ P_{\text{a0}}(a-4)(1-P_{\text{a0}})(1-P_{\text{a1}})^2P_{\text{a1}} + \\
+ P_{\text{a0}}(a-5)(1-P_{\text{a0}})(1-P_{\text{a1}})^3P_{\text{a1}} + \\
+ P_{\text{a0}}(a-6)(1-P_{\text{a0}})(1-P_{\text{a1}})^4P_{\text{a1}} + \\
+ P_{\text{a0}}(a-7)(1-P_{\text{a0}})(1-P_{\text{a1}})^5P_{\text{a1}} + \\
+ \ldots
\]

(0.27)

The three probabilities \( \{P_{\text{a0}}, P_{\text{a1}}, P_{\text{a5}}\} \) are estimated by minimizing

\[
S = \sum_{a, \text{ageGroup}} \left( \hat{p}_{\text{mortality}}(a) - \hat{p}_{\text{mortality}}(a) \right)^2
\]

(0.28)

When the longitudinal probability of the disease incidence at age \( a \) satisfies the recursion relation

\[
p_{\text{inc}}(a) = (1 - p_i(0))(1 - p_i(1))(1 - p_i(a - 1))p_i(a)
\]

(0.29)

Table A3 is taken from the Cancer Research UK website as an example. It gives 1, 5 and 10 year survival percentages for lung cancer.

**Table A3: Survival percentage for lung cancer**

<table>
<thead>
<tr>
<th>Cancer</th>
<th>1 year</th>
<th>5 year</th>
<th>10 year</th>
<th>1 - ( P_{\text{a0}} )</th>
<th>1 - ( P_{\text{a1}} )</th>
<th>1 - ( P_{\text{a5}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lung</td>
<td>32</td>
<td>10</td>
<td>5</td>
<td>0.32</td>
<td>0.75</td>
<td>0.71</td>
</tr>
</tbody>
</table>


47
The probabilities of being alive after 1, 5 and 10 years are

\[ p_{\text{survival}}(a_0 + 1) = (1 - p_w) \]
\[ p_{\text{survival}}(a_0 + 5) = (1 - p_w)(1 - p_w)^5 \]  
\[ p_{\text{survival}}(a_0 + 10) = (1 - p_w)(1 - p_w)^5(1 - p_w)^5 \]  

(0.30)

Survival rates

It is common practice to describe survival in terms of a survival rate \( R \), supposing an exponential death-distribution. In this formulation the probability of surviving \( t \) years from some time \( t_0 \) is given as

\[ p_{\text{survival}}(t) = 1 - R^{-1} \int_0^t dt e^{-Rt} = e^{-Rt} \]  

(0.31)

For a time period of 1 year

\[ p_{\text{survival}}(1) = e^{-R} \]  

(0.32)

\[ R = -\ln(p_{\text{survival}}(1)) = -\ln(1 - p_w) \]

For a time period of, for example, 4 years,

\[ p_{\text{survival}}(t = 4) = 1 - R^{-1} \int_0^4 dt e^{-Rt} = e^{-4R} = (1 - p_w)^4 \]  

(0.33)

In short, the survival rate is minus the natural log of the 1-year survival probability.

Survival models 0, 1 and 2

For any potentially terminal disease, the model can use any of three survival models, numbered {0, 1, 2}. The parameters describing these models are given below.

Survival model 0

Given the 1-year survival probability \( p_{\text{survival}}(1) \)

The model uses 1 parameter \( R \)

\[ R = -\ln(p_{\text{survival}}(1)) \]  

(0.34)

Survival model 1

The model uses two parameters \( (p_w, R) \)

Given the 1-year survival probability \( p_{\text{survival}}(1) \) and the 5-year survival probability \( p_{\text{survival}}(5) \)

\[ p_1 = 1 - p_{\text{survival}}(1) \]
\[ R = -\frac{1}{4} \ln\left( \frac{p_{\text{survival}}(5)}{p_{\text{survival}}(1)} \right) \]  

(0.35)
Survival model 2

The model uses three parameters \((p_1, R, R_5)\)

Given the 1-year survival probability \(p_{\text{survival}}(1)\) and the 5-year survival probability \(p_{\text{survival}}(5)\)

\[
P_1 = 1 - p_{\text{survival}}(1)
\]

\[
R = -\frac{1}{4} \ln \left( \frac{p_{\text{survival}}(5)}{p_{\text{survival}}(1)} \right)
\]

\[
R_5 = -\frac{1}{5} \ln \left( \frac{p_{\text{survival}}(10)}{p_{\text{survival}}(5)} \right)
\]

Approximating single-state disease survival data from mortality and prevalence

An example is provided here with a standard life-table analysis for a disease \(d\).

Consider the four following states:

<table>
<thead>
<tr>
<th>STATE</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>alive without disease (d)</td>
</tr>
<tr>
<td>1</td>
<td>alive with disease (d)</td>
</tr>
<tr>
<td>2</td>
<td>dead from disease (d)</td>
</tr>
<tr>
<td>3</td>
<td>dead from another disease</td>
</tr>
</tbody>
</table>

\(p_k\) is the probability of disease \(d\) incidence, aged \(k\)

\(p_{ik}\) is the probability of dying from the disease \(d\), aged \(k\)

\(p_{aik}\) is the probability of dying other than from disease \(d\), aged \(k\)

The state transition matrix is constructed as follows

\[
\begin{bmatrix}
p_0(k+1) \\
p_1(k+1) \\
p_2(k+1) \\
p_3(k+1)
\end{bmatrix} =
\begin{bmatrix}
(1 - p_{ik})(1 - p_{aik}) & (1 - p_{ik} - p_{aik})p_{ik} & 0 & 0 \\
(1 - p_{ik})p_{ik} & (1 - p_{ik} - p_{aik})(1 - p_{ik}) & 0 & 0 \\
0 & p_{aik} & 1 & 0 \\
p_{ik} & p_{ik} & 0 & 1
\end{bmatrix}
\begin{bmatrix}
p_0(k) \\
p_1(k) \\
p_2(k) \\
p_3(k)
\end{bmatrix}
\]

\(0.37\)

It is worth noting that the separate columns correctly sum to unity.

The disease mortality equation is that for state 2,

\[
p_2(k+1) = p_{aik}p_1(k) + p_2(k)
\]

\(0.38\)
The probability of dying from the disease in the age interval \([k, k+1]\) is \(p_{\text{dk}}(k)\), which is not to be confused with the (cross-sectional) disease mortality, called \(p_{\text{mor}}(k)\).

\(p_{\text{dk}}(k)\) is known as the disease prevalence, \(p_{\text{pre}}(k)\). Hence the relation

\[
p_{\text{dk}} = \frac{p_{\text{mor}}(k)}{p_{\text{pre}}(k)} \tag{0.39}
\]

For exponential survival probabilities, the probability of dying from the disease in the age-interval \([k, k+1]\) is denoted as \(p_{\text{dk}}\) and is given by the formula

\[
p_{\text{dk}} = 1 - e^{-R_k} \Rightarrow R_k = -\ln(1 - p_{\text{dk}}) \tag{0.40}
\]

When, as is the case for most cancers, these survival probabilities are known, the microsimulation will use them; when they are not known or are too old to be of any use, the microsimulation uses survival statistics inferred from the prevalence and mortality statistics (equation (0.39)).

An alternative derivation equation (0.39) is as follows. Let \(N_k\) be the number of people in the population aged \(k\) and let \(n_k\) be the number of people in the population aged \(k\) with the disease. Then, the number of deaths from the disease of people aged \(k\) can be given in two ways: as \(p_{\text{dk}}n_k\) and, equivalently, as \(p_{\text{mor}}(k)N_k\). Observing that the disease prevalence is \(n_k/N_k\) leads to the equation

\[
p_{\text{dk}} n_k = p_{\text{mor}}(k) N_k
\]

\[
p_{\text{pre}}(k) = \frac{n_k}{N_k} \Rightarrow p_{\text{dk}} = \frac{p_{\text{mor}}(k)}{p_{\text{pre}}(k)} \tag{0.41}
\]

Approximating multi-state disease survival data from incidence and mortality, assuming no remission

Disease mortality statistics give the probability that a person will die from the disease in a given year of life. They make no reference to when the disease from which the person dies was contracted.

Disease survival statistics give the probability that a person will die from the disease in a given year of life given that they contracted the disease in an earlier year.

The connection between the two is provided by the equation of the form

\[
p_{\text{mor}}(a) = \sum_{\alpha_0/a} p_{\text{mor}}(a|\alpha_0) p_{\text{mor}}(\alpha_0) \tag{0.42}
\]
This equation can be used to infer survival statistics when only the incidence and mortality statistics are known – essentially by choosing the survival statistics so as to get the mortality statistics calculated from equation (0.42) as close as possible to the known set.

Multi-state diseases have mortality, survival and incidence statistics that are state-dependent. Aside from this additional level of complexity, the determination of disease survival proceeds in the same way.

**Setup**

For each sex, consider an $N$-stage, terminal disease for which both the inter-stage transition probabilities, $p_{i,j}(a)$ is the probability to go from stage $i$ to stage $j$, and the state-dependent mortality probabilities $p_{mk}(a)$ (K denotes the stage number and a the age) are known. The following algorithm allows for optimal determination of the stage-dependent survival probabilities. In the special case of a single-state disease it reduces to the previously developed single-stage determination of survival.

**Definitions**

- $p_k(a_{o},K_a)$ The probability of not having died from the disease and being in stage $K$ at age $a$, given that the disease was contracted at $a_{o}$ in state $K_o$
- $p_{w}(a_{o},K_a)$ The probability of being dead (from the disease) at age $a$, given that the disease was contracted at $a_{o}$ in state $K_0$
- $p_{m}(a_{o},K_a)$ The probability of first getting the disease in $a_{o}$ in state $K_{a}$ given no disease at age 0
- $p_{s}^{K}(a | a_{o})$ The probability of dying from the disease in stage $K$ at age $a$ given that the disease was contracted at age $a_{o}$ and that the person was alive at age $a-1$

**Disease incidence**

The probability that a person, who at age 0 does not have the disease, first gets the disease at age $a_{o}$ in state $K_{0}$ is given as

$$p_{m}(a_{o},K_a) = \prod_{a=0}^{a_{o}-1} \left(1 - \sum_{k=1}^{K} p_{mk}(a)\right) p_{0K_a} \tag{0.43}$$

**Disease survival**

Once a person has the disease they can possibly change disease stage or they can die from the disease. (This analysis focuses only on the identified disease and does not allow for the possibility that they die from other causes.) Suppose they acquire the disease at age $a_{o}$ in stage $K$, then the initial state vector is determined from the initial conditions $p_{i}(a_{o}) = \delta_{K_{a}} \ (K_{a} > 0)$, $p_{w}(a_{o}) = 0$. 
At subsequent ages, the state probabilities are given by the recursion equation

\[
\begin{bmatrix}
    p_0(a + 1 | a_o, K_0) \\
p_1(a + 1 | a_o, K_0) \\
\vdots \\
p_{N-1}(a + 1 | a_o, K_0) \\
p_n(a + 1 | a_o, K_0)
\end{bmatrix}
= T(a, a_o)
\begin{bmatrix}
p_0(a | a_o, K_0) \\
p_1(a | a_o, K_0) \\
\vdots \\
p_{N-1}(a | a_o, K_0) \\
p_n(a | a_o, K_0)
\end{bmatrix}
\]

\[
T(a, a_o) = \begin{bmatrix}
1 - \sum_{k=0}^{a-o} p_{0,k} & 0 & \cdots & 0 & 0 \\
p_{0,1} & 1 - \sum_{k=1}^{a-o} p_{1,k} \left(1 - p^1_o(a | a_o)\right) & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
p_{0,N-1} & p_{1,N-1} \left(1 - p^1_o(a | a_o)\right) & \cdots & \left(1 - p^{N-1}_o(a | a_o)\right) & 0 \\
0 & p^1_o(a | a_o) & \cdots & p^{N-1}_o(a | a_o) & 1
\end{bmatrix}
\]

Where, for survival model 2,

\[
p^K_o(a | a_o) = \begin{cases}
p^K_o(a = a_o) \\
p^K_o(a_0 < a \leq a_0 + 4) \\
p^K_o(a_0 + 4 < a)
\end{cases}
\]

**Disease mortality**

The probability of dying from the disease in stage K at age \(a\), is denoted \(p^K_{m,0}(a)\) is given by the equation

\[
p^K_{m,0}(a) = \sum_{K=1}^{K-1} p^K_{m,0}(a | a_o, K_o)p^K_{m}(a-1 | a_o)p^K_{m,0}(a_o, K_o)
\]

**Estimating survival**

When the disease mortality is known, here denoted \(p^K_{m,0}(a)\), the sets of survival parameters (3 for each state, and possibly stratified by age) can be estimated by minimizing

\[
S = \sum_{\text{state } k} \left( \sum_{\text{AgeGroup } a} \frac{(p^K_{m,0}(a) - p^K_{m,0}(a))^2}{\sigma^2(a)} \right)
\]

**Approximating attributable cases**

The number of smoking-attributable cases (\(I_d\)) for a disease (d) is calculated by dividing the number of new cases of a disease among individuals who are either smokers or ex-smokers (n) by the total number of people in the population in a given year.

\[
I_d(y) = \frac{n(y)}{N(y)}
\]
**Potential Years of Life Lost**

The PYLL (Gardner and others 1990; Health and Social Care Information Centre 2015) for an individual ($PYLL(i)$) who dies in a given year will be calculated from the following equation (0.49).

$$PYLL(i) = \begin{cases} Age_{ref} - Age_{death} & \text{if } Age_{death} < Age_{ref} \\ 0 & \text{if } Age_{death} \geq Age_{ref} \end{cases}$$  \hspace{1cm} (0.49)

For each individual the difference between the reference age (life expectancy) and the age of death will be calculated. The total PYLL each year ($TotalPYLL(year)$) will be calculated each year in the microsimulation. This metric will consider individuals who have died in a given year ($N_{died(year)}$).

As the simulation projects into the future, and simulates a cohort of children with defined age groups and therefore year of birth, life expectancy values for each simulated individual will be based on their life-expectancy values at birth, obtained for each country from national statistics repositories.

$$TotalPYLL(year) = \frac{\sum_{i=1}^{N_{pop}(year)} PYLL(i)}{N_{population}(year)}$$ \hspace{1cm} (0.50)

**Premature Mortality Rate**

The premature mortality rate ($PM(year)$) based on the number of individuals who die prematurely in a given year is calculated based on equations (0.51).

$$premature(i) = \begin{cases} 1 & \text{if } age_{death} < 70 \\ 0 & \text{if } age_{death} \geq 70 \end{cases}$$ \hspace{1cm} (0.51)

$$PM(year) = \frac{\sum_{i=1}^{N_{pop}(year)} premature(i)}{N_{population}(year)}$$ \hspace{1cm} (0.52)

**Costs Module**

The costs module includes both direct and indirect cost calculations.

**Direct costs**

Direct costs are calculated based on a cost per case which is constant throughout the simulation.

$$\text{Direct cost (£) per individual (year)} = \frac{\text{Cost per case ($)} \times \text{Prevalence(year)}}{\text{Alive (year)}}$$ \hspace{1cm} (0.53)

The direct costs are displayed per million US dollars ($) and at a rate defined by the user.
Direct costs (M£ per rate (year)) = \frac{\text{Direct cost ($ per individual (year)) \times rate}}{10^6} \quad (0.54)

95 percent confidence intervals are calculated from the prevalence rates per individual (\( P \)) by the equation below.

\[
95\% CI = \text{Direct Costs (year)} \times 1.96 \sqrt{\frac{P(1-P)}{\text{Trials}}} \quad (0.55)
\]

**Premature Mortality Costs**

\[
PMC(i) = \begin{cases} 
\sum_{i=\text{age at birth}}^{\text{age death}} \text{Income}(i) & \text{if } \text{age death} < \text{LE at birth} \\
0 & \text{if } \text{age death} \geq \text{LE at birth}
\end{cases} \quad (0.56)
\]

The premature mortality costs (Gold and others 1996; Menzin and others 2012) for each individual \((PMC(i))\) are calculated by summing over the income costs from the age of death until the maximum age. The maximum age can be defined as the pension age or by some other value.

The model outputs average PMC per 100,000 as shown in equation (0.57).

\[
\text{PMC per 100,000} = \frac{PMC_{\text{Total}}}{N_{\text{population}}} \times 100,000 \quad (0.57)
\]

**Propagation of errors equation**

To include totals for each of the outputs, the sum of each disease output (e.g. incidence, prevalence) was summed. The total errors \((E_n)\) were calculated using the propagation of errors equation:

\[
E_n = \sqrt{E_1^2 + E_2^2 + \ldots + E_n^2} \quad (0.58)
\]

Where \(E_n\) is the error for each individual disease output which has been included in the sum.

**Software architecture**

**Aim of the Model**

This model uses a common method (microsimulation methods) to predict the impact of changing risk factors to measure chronic disease.

The model is an epidemiological/medical competing risk application that uses both stochastic and deterministic processing capable of projecting cohort mortality rates for individuals or a population, taking individuals’ risk factors and medical profiles into account. Through its interactive scenario specification, the model allows for the effects...
of ageing and the projection of future mortality rates, either with or without taking into account possible future trends in risk factors or medical conditions.

**Summary of the Architecture of the Existing Model**

The existing solution is written in C++ (compiler Embarcadero C++ Builder). It is a modular, object-oriented design and is compiled to run using the Windows operating system (see figure A2).

The application has a limited interactive graphics capability designed for the rapid assessment of outputs and comparative assessments of batched runs. Diagrams and graphs produced in this way can be exported from the application in suitable file formats. The model is equipped with a suite of editors, allowing flexible and traceable input of individual, cohort or population data.

The model’s inputs are in the form of tab-delimited text files. The application has a number of editors that can create, edit and store these files. The simulation, disease and scenario editors allow the user to specify all input data files, parameters and processing rules necessary for a run of the program. The application’s many data inputs are processed in a similar fashion – for a specified run-configuration the application dynamically creates and maintains lists of software objects, each object being constructed from a designated data file (the disease and scenario objects that follow are examples of this process). Files inputted in this way into their corresponding dynamical objects are automatically checked for their data integrity by the newly created software object’s own methods. Each run creates and stores a time-tagged configuration file specifying the complete set of input files, output files, parameter settings and rule set. Provided that the necessary input data are available, it is possible to rerun a simulation by reusing the configuration file.

Outputs are handled in a similar way but in reverse: run-time generated outputs are stored in dynamically created output objects; at the end of a run the objects write their data to tab-delimited text files. Outputs can be summary files, or medium- or low-level data files that can be further processed by standard software packages.
Main C++ classes used by the model

Individual members of the population and diseases are modelled by the C++ classes Tperson and Tdisease respectively. The risk factor trends and scenarios are modelled by the C++ class Tscenario. The principal operations of the program can be regarded as the interactions of Tperson, Tdisease and Tscenario objects. These classes have some of their fields and methods highlighted in the following section and subsections to give an indication of the processing chain implemented in the model. The software closely follows the real world life of individuals – they age in a personal risk factor environment, and possibly develop diseases from which they may recover or die (see figure 1, page 12 for a diagram of the model’s inputs and outputs).

Tperson C++ class

People are implemented as instances of the C++ class Tperson; an indication of its data fields and methods (provided in table A4) are grouped into the state record. The medHistory record maintains their current disease and risk factor status together with data necessary for the computation disease-related transitions. The Tperson object’s data fields are updated annually with the yearByYear(diseaseList, scenarioList, …) method. The method needs to be supplied with pointers to the list of disease object pointers being modelled and the list of risk factor scenario object pointers that determines how a person’s set of risk factors change over the year.
Tdisease C++ class

Both single and multistage diseases are implemented as instances of the C++ class Tdisease. An example of their structure is shown in figure A3.

Each stage of a disease has its own set of disease statistics such as incidence risks, remission risks and survival risks. Moreover, each disease stage will also contain economic data such as direct and indirect costs. In addition to the data fields there are also methods that are included in the Tdisease class (examples are provided in table A5). Disease data are stored in structured text files – one file for each disease or version of that disease. The key method of the Tdisease class is function GetRisk() which, for a specified person’s state, medical history and risk factor type the method returns the relevant transition probability. If the application is running in a stochastic transition mode this probability is compared to an application-generated random number to determine if the transition takes place; in deterministic mode the same transition probability is included in the relevant life-disease table that computes and lists the probabilities of being alive and in possible exclusive disease states, or dead.

### Table A4: C++ Tperson class

<table>
<thead>
<tr>
<th>TPERSON</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data field</td>
<td></td>
</tr>
<tr>
<td>State</td>
<td>State vector record</td>
</tr>
<tr>
<td>medHistory</td>
<td>Medical history record</td>
</tr>
<tr>
<td>Method</td>
<td></td>
</tr>
<tr>
<td>Tperson(state0,medHistory0)</td>
<td>Constructor for initial state and history</td>
</tr>
<tr>
<td>yearByYear(...)</td>
<td>Updates state and medHistory by one year</td>
</tr>
</tbody>
</table>

Tdisease C++ class

Both single and multistage diseases are implemented as instances of the C++ class Tdisease. An example of their structure is shown in figure A3.

Each stage of a disease has its own set of disease statistics such as incidence risks, remission risks and survival risks. Moreover, each disease stage will also contain economic data such as direct and indirect costs. In addition to the data fields there are also methods that are included in the Tdisease class (examples are provided in table A5). Disease data are stored in structured text files – one file for each disease or version of that disease. The key method of the Tdisease class is function GetRisk() which, for a specified person’s state, medical history and risk factor type the method returns the relevant transition probability. If the application is running in a stochastic transition mode this probability is compared to an application-generated random number to determine if the transition takes place; in deterministic mode the same transition probability is included in the relevant life-disease table that computes and lists the probabilities of being alive and in possible exclusive disease states, or dead.
Figure A3: Multistage disease architecture

Key:
- inc: incidence
- rem: remission
- pre: prevalence
- mor: mortality
- sur: survival
- bmi: body mass index
- smk: smoking

01 → stage0 to stage1
12 → stage1 to stage2
20 → stage1 to stage2
11 → stage1
Processing is user-specified to be either random (Monte Carlo) or deterministic. The random option can process any specified population or cohort; the deterministic option processes only cohorts. In this context: a population is a specified number of males and females whose age distributions and risk factor distributions are inputted as appropriate tab-delimited text files; a cohort is a text file of individuals specifying, for each individual, their initial state and medical history. The user options and necessary data files are specified in the application’s simulation editor.

The user must also specify the set of diseases and the set of risk factors that is being simulated. Again this is done via the appropriate application editor: The disease editor allows the construction and identification of a batch file of disease files; the simulation editor allows for the specification of the mix of risk factors and, where necessary, their distributions by age and gender. The simulation editor also provides the mechanism by which essential run parameters are specified – the start year, stop year, number of trials, and so on.

<table>
<thead>
<tr>
<th>DATA FIELD</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>terminal</td>
<td>Boolean, true if the disease is terminal</td>
</tr>
<tr>
<td>state</td>
<td>Disease state (normal, severe,...)</td>
</tr>
<tr>
<td>*DataAvailability</td>
<td>Boolean array of data availability by risk type</td>
</tr>
<tr>
<td>**IncidenceRisk</td>
<td>Incidence rates by age, gender</td>
</tr>
<tr>
<td>***SurvivalRisk</td>
<td>Survival rates by age, gender, state</td>
</tr>
<tr>
<td>**PrevalenceRisk</td>
<td>Prevalence rates by age, gender</td>
</tr>
<tr>
<td>***RemissionRisk</td>
<td>Remission rates by age and gender, state</td>
</tr>
<tr>
<td>***MortalityRisk</td>
<td>Mortality rates by age, gender, state</td>
</tr>
<tr>
<td>***RelRisk</td>
<td>Relative risks by risk factor type, age, gender</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>METHOD</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>TDisease(aFile)</td>
<td>Constructor using data from aFile</td>
</tr>
<tr>
<td>LoadFromFile(aFile)</td>
<td>Fills the data fields from aFile</td>
</tr>
<tr>
<td>WriteToFile(aFile)</td>
<td>Writes the data fields to aFile</td>
</tr>
<tr>
<td>GetRisk(state, medHistory, risktype)</td>
<td>Returns risk for specified risktype</td>
</tr>
</tbody>
</table>
Individuals are processed one at a time from the simulation’s start year until they either die or reach the simulation’s stop year. In each simulated year they can either contract any mix of the modelled diseases that they do not yet have; achieve remission from any disease or disease stage they might have; die from any terminal disease they might have; or die from other causes (other causes are modelled as a single, instantly fatal, terminal disease; its incidence probability is constructed via the disease editor from the modelled diseases’ mortality statistics and the appropriate national mortality statistics).

Each run of the model requires the specification of a risk factor scenario for each risk factor modelled. These scenarios can simply maintain risk factor distributions at their start year values or they can allow for the modelling of risk factor trends or medical advances resulting in the reduction of disease incidence or improvements in the survivability of specified diseases.

**Tscenario C++ class**

Scenarios are modelled as instances of the C++ Tscenario class and are constructed by the scenario editor that is accessed via the simulation editor.

Runs can be organized into batches with different runs having different risk factor scenarios. This allows for direct comparisons to be made – for example, what happens to life expectancy with or without improvements to the treatment of stroke.

Scenarios are implemented as instances of the C++ class, Tscenario; an indication of its data fields and methods are provided in table A6. The scenario objects are constructed from files that are created by the scenario editor.

Much of the input data (disease data, mortality data, demographic data, etc.) is typically changed on an annual basis. Such changes are easily accommodated and logged via the disease, distribution and simulation editors. New diseases that are described by the current set of risk factors can be added to (or subtracted from) the simulation via the disease editor.

The model has essentially only two external software dependencies: its own C++ development environment and its host processor’s operating system. The configuration was chosen for its maintainability.
Table A6: The C++ Tscenario class

<table>
<thead>
<tr>
<th>TSCENARIO</th>
<th>DATA FIELD</th>
<th>METHOD</th>
</tr>
</thead>
<tbody>
<tr>
<td>scenarioType</td>
<td>Type of scenario e.g. (smoking...)</td>
<td>Tscenario(aFile)</td>
</tr>
<tr>
<td>start year</td>
<td>Year at which scenario starts</td>
<td>LoadFromFile(aFile)</td>
</tr>
<tr>
<td>stop year</td>
<td>Year at which scenario stops</td>
<td></td>
</tr>
<tr>
<td>futureRiskFile</td>
<td>File specifying future risk distribution</td>
<td></td>
</tr>
<tr>
<td>targetAgeGroup</td>
<td>Target age group e.g. (18+)</td>
<td></td>
</tr>
<tr>
<td>targetGenderGroup</td>
<td>Target gender group e.g. (males,females)</td>
<td></td>
</tr>
</tbody>
</table>

---
APPENDIX B: CIGARETTE TAX
SCENARIOS OUTPUT, 2015–2017, UKRAINE (RESULTS FROM TAXSIM MODELLING)
**GOVERNMENT REVENUE TYPE**

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Actual 2015</th>
<th>Expected contribution to GDP 2016: Ad valorem (12%) Minimum specific (HRV 8.515) and simple specific HRV 6.365)</th>
<th>Expected contribution to GDP 2017: Ad valorem remains equal and 12% tax increase in Minimum specific excise (HRV 9.54), and simple specific (HRV 7.13)</th>
<th>Expected contribution to GDP 2017: Increase Ad valorem (15%), and 30% increase in the Minimum specific excise (HRV 11.08), and simple specific (HRV 8.28)</th>
<th>Expected contribution to GDP 2017: Increase 30% the Ad valorem, and 50% increase in the Minimum specific excise (HRV 12.77), and simple specific (HRV 9.55)</th>
<th>Expected contribution to GDP 2017: Increase Ad valorem and specific tax (40%), adopting a simplified tax structure with uniform specific excise (HRV 7.19)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>73.8</td>
<td>48.8</td>
<td>53.4</td>
<td>60.1</td>
<td>64.0</td>
<td>66.9</td>
</tr>
<tr>
<td></td>
<td>15.2</td>
<td>41.4</td>
<td>32.9</td>
<td>24.7</td>
<td>21.2</td>
<td>19.2</td>
</tr>
<tr>
<td></td>
<td>US$0.63</td>
<td>US$1,01</td>
<td>US$0.87</td>
<td>US$0.81</td>
<td>US$0.81</td>
<td>US$0.81</td>
</tr>
<tr>
<td></td>
<td>308.9</td>
<td>482.6</td>
<td>428.6</td>
<td>573.0</td>
<td>573.0</td>
<td>573.0</td>
</tr>
<tr>
<td></td>
<td>22.8</td>
<td>1.2%</td>
<td>1.3%</td>
<td>1.3%</td>
<td>1.3%</td>
<td>1.3%</td>
</tr>
<tr>
<td></td>
<td>US$0.94</td>
<td>US$1,01</td>
<td>US$0.87</td>
<td>US$1,17</td>
<td>US$1,17</td>
<td>US$1,17</td>
</tr>
<tr>
<td></td>
<td>6.0</td>
<td>0.3%</td>
<td>0.3%</td>
<td>0.3%</td>
<td>0.3%</td>
<td>0.3%</td>
</tr>
<tr>
<td></td>
<td>US$0.33</td>
<td>US$0.85</td>
<td>US$0.85</td>
<td>US$1.21</td>
<td>US$1.21</td>
<td>US$1.21</td>
</tr>
<tr>
<td></td>
<td>34.9</td>
<td>1.6%</td>
<td>1.6%</td>
<td>4.58</td>
<td>4.58</td>
<td>4.58</td>
</tr>
<tr>
<td></td>
<td>US$1.44</td>
<td>US$1.8</td>
<td>US$1.8</td>
<td>US$1.8</td>
<td>US$1.8</td>
<td>US$1.8</td>
</tr>
<tr>
<td></td>
<td>56.3</td>
<td>1.0%</td>
<td>1.0%</td>
<td>1.0%</td>
<td>1.0%</td>
<td>1.0%</td>
</tr>
<tr>
<td></td>
<td>US$0.97</td>
<td>US$1.03</td>
<td>US$1.03</td>
<td>US$1.03</td>
<td>US$1.03</td>
<td>US$1.03</td>
</tr>
<tr>
<td></td>
<td>6.0</td>
<td>0.3%</td>
<td>0.3%</td>
<td>0.3%</td>
<td>0.3%</td>
<td>0.3%</td>
</tr>
<tr>
<td></td>
<td>US$0.25</td>
<td>US$1.8</td>
<td>US$1.8</td>
<td>US$1.8</td>
<td>US$1.8</td>
<td>US$1.8</td>
</tr>
<tr>
<td></td>
<td>34.9</td>
<td>1.6%</td>
<td>1.6%</td>
<td>4.58</td>
<td>4.58</td>
<td>4.58</td>
</tr>
<tr>
<td></td>
<td>US$1.44</td>
<td>US$1.8</td>
<td>US$1.8</td>
<td>US$1.8</td>
<td>US$1.8</td>
<td>US$1.8</td>
</tr>
<tr>
<td></td>
<td>56.3</td>
<td>1.0%</td>
<td>1.0%</td>
<td>1.0%</td>
<td>1.0%</td>
<td>1.0%</td>
</tr>
<tr>
<td></td>
<td>US$0.97</td>
<td>US$1.03</td>
<td>US$1.03</td>
<td>US$1.03</td>
<td>US$1.03</td>
<td>US$1.03</td>
</tr>
<tr>
<td></td>
<td>6.0</td>
<td>0.3%</td>
<td>0.3%</td>
<td>0.3%</td>
<td>0.3%</td>
<td>0.3%</td>
</tr>
<tr>
<td></td>
<td>US$0.25</td>
<td>US$1.8</td>
<td>US$1.8</td>
<td>US$1.8</td>
<td>US$1.8</td>
<td>US$1.8</td>
</tr>
<tr>
<td></td>
<td>34.9</td>
<td>1.6%</td>
<td>1.6%</td>
<td>4.58</td>
<td>4.58</td>
<td>4.58</td>
</tr>
<tr>
<td></td>
<td>US$1.44</td>
<td>US$1.8</td>
<td>US$1.8</td>
<td>US$1.8</td>
<td>US$1.8</td>
<td>US$1.8</td>
</tr>
<tr>
<td></td>
<td>56.3</td>
<td>1.0%</td>
<td>1.0%</td>
<td>1.0%</td>
<td>1.0%</td>
<td>1.0%</td>
</tr>
<tr>
<td></td>
<td>US$0.97</td>
<td>US$1.03</td>
<td>US$1.03</td>
<td>US$1.03</td>
<td>US$1.03</td>
<td>US$1.03</td>
</tr>
</tbody>
</table>

**Notes:**
- World Bank Group forecast: annual average exchange rate = 2016 (1 US$/HRV 23.8); 2017 (1 US$/HRV 24.4)
- *Expected contribution to GDP* is calculated based on the government revenue type mentioned.
- The table includes data on cigarette consumption, average price, excise tax revenue, and additional tobacco excise.
REFERENCES


