Modeling the Long-Term Health and Cost Impacts of Reducing Smoking Prevalence Through Tobacco Taxation in Ukraine

WORLD BANK GROUP

Acknowledgements

Modeling the Long-Term Health and Cost Impacts of Reducing Smoking Prevalence through Tobacco Taxation in Ukraine

Introduction

Smoking is a leading cause of preventable premature deaths. Smoking's effects will continue to devastate lives in many countries, including Ukraine, if measures are not implemented to reduce its prevalence. Smoking is a major cause of many chronic diseases, such as cardiovascular disease, respiratory disease, and smoking-related cancers.

Over recent years, successful tobacco control policies in Ukraine have resulted in one of the fastest declines in smoking prevalence in the world (1). This is largely due to multifaceted tobacco control legislation, adopted from 2005 and subsequently upgraded.

Ukraine ratified the WHO Framework Convention on Tobacco Control (FCTC) in 2006. Currently, Ukrainian legislation basically corresponds to FCTC requirements. In 2005, Ukraine adopted a first tobacco-control law. Since then, several additional tobacco-control policies have been implemented in the country.

Smoke-free policies supported by media campaigns have covered many workplaces and public places since the middle of 2006. Under these policies, at least 50 percent of the area of restaurants and bars had to be isolated from the smoking area, so that tobacco smoke did not penetrate into smoke-free areas. This measure was supported by an intensive media campaign and public movement in favor of smoke-free policies. Many restaurants went completely smoke-free both before and after implementing this measure. As of December 2012, restaurants, workplaces, and other public places became 100 percent smoke-free. Designated smoking places, which figured in the legislation between 2006 and 2012, were abolished in the amended laws. As of late 2006, cigarette packs carried textual warning labels covering 30 percent of their surface, in place of a previous warning which covered 10 percent of the front surface and stated: "Ministry of Health warns: Smoking is bad for your health." Since October 4, 2012, large (50 percent of the pack surface area), graphic health-warning labels on tobacco packaging have been introduced.

A ban on outdoor tobacco advertising since January 2009 was followed by a more comprehensive tobacco advertising ban, which entered into force on September 16, 2012.

In addition to the reduced tobacco affordability observed during the global economic recession, the average tax incidence was increased between August 2008 and July 2010 from 0.5 UAH (Ukrainian hryvnia, the national currency of Ukraine) to 3 UAH per cigarette pack. Further changes in tobacco tax rates were less substantial and were above inflation only in some years. However, while the policies described were definitely beneficial, much progress remains to be made. As of late 2015, the prevalence of current smoking among men in Ukraine was 45 percent (2, 3), although prevalence is much lower in women, at 11 percent (2, 3).

It is not a given that smoking trends will continue to decrease in Ukraine, unless effective tobacco control measures are sustained and strengthened. Especially when the economy grows, commodities/luxuries such as smoking will become more affordable.

Tobacco industry tactics can also become an important factor in determining the level of cigarette consumption. One of the mechanisms of this influence derives from the industry's right to determine the maximum retail price of cigarettes and thus to manipulate the net-of-tax portion of the price. In 2016, a new tax policy stipulated that the minimal specific tobacco tax increase by 40 percent. Thus, the retail price was expected to increase and the consumption of cigarettes to decrease. However, the actual level of cigarette consumption increased. This happened because tobacco companies, aiming to keep their customers, initiated "price wars." This example illustrates that more factors are at play than are usually taken into account in weighing policy choices. Price and tax factors are extremely important and need to be considered when forecasting trends.

The present report provides evidence from a modeling exercise undertaken to predict the health and related cost impacts that may stem from the implementation of a tobacco excise tax increase in Ukraine. 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 taxation on the future burden of a range of non-communicable diseases (NCDs). Specifically, the disease outcomes quantified were coronary heart disease (CHD), stroke, chronic obstructive pulmonary disease (COPD), and lung cancer. The microsimulation model has been deemed by the OECD the most relevant method for NCD modeling based on risk-factor data (4). 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 Ukraine. That work was carried out by the World Bank, using the Tobacco Tax Simulation Model (TaXSiM) developed by the World Health Organization (WHO) (5, 6).

Summary of results

Table 1 presents a summary of total disease cases (epidemiological) and costs (economic) by parameter, year, and scenario as rates per Ukraine population.

The model estimated that by 2035 the specified tax increase would result in the avoidance of 126,730 new cases of smoking-related disease; 29,172 premature deaths; and 267,098 potential years of life lost, relative to no change in tax. These reductions in disease and death will avoid 1.5 billion UAH in healthcare costs and 16.5 billion UAH in premature mortality costs, respectively.

Table 1: **Summary Table of the Outputs as Rates per Ukraine Population, by Year**

Epidemiological outputs	Year	Sc0 (Baseline)	Sc1
Cumulative incident cases	2025	5480948[±4237]	5427558[±4237]
	2035	11366868[±5753]	11255173[±5753]
Cumulative incident cases avoided	2025	NA	56224[±6341]
	2035	NA	126730[±9123]
Incident cases per year	2025	589035[±1545]	582341[±1545]
	2035	646600[±1545]	640799[±1727]
Attributable incident cases	2025	218221[±1121]	208475[±1121]
	2035	222603[±1041]	211984[±1041]
Cumulative premature deaths avoided	2025	NA	6372
	2035	NA	29172
Cumulative potential years of life lost relative to baseline	2025	NA	48923
	2035	NA	267098
Economic outputs			
Direct costs avoided (millions UAH)	2025	NA	542.23
	2035	NA	1545.81
Cumulative premature mortality costs avoided (millions UAH)	2025	NA	3568.4
	2035	NA	16536.4

Summary of Methods

Methods

The model simulates a virtual population of Ukraine, based on known population statistics.

Initial smoking prevalence by age and sex was extracted from the 2015 Annual Household Survey conducted by the National Statistics Service of Ukraine.

Scenarios took account of price impacts on uptake of smoking and cessation.

Individuals within 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 were extracted from the Global Burden of Disease database.

Relative risks of contracting diseases in smokers compared to non-smokers were extracted from peer-reviewed literature.

A five-module microsimulation model was used to predict the future health and economic impacts of smoking prevalence by 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 2015 estimates.

A specified percentage of smokers who are affected by the tax increase move to the "never-smoker" category in 2017, in order to account for reductions in uptake due to price increases.

If an individual's smoking status is changed by the intervention, their smoking status will then remain fixed for the entire simulation.

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

Smokers react quickly to the tax: we modeled an immediate effect and then a linear trend, in line with TaXSiM.

Limitations

No data on survival for the specified NCDs were available.

Data on the percentages of ex-smokers in Ukraine are limited.

The model does not take account of future changes in policy or technology.

No change in secondhand smoke exposure is modeled.

Baseline is static over time.

The simulation only includes four smoking-related diseases, so results are likely underestimates of the true effects.

No data on non-healthcare costs, e.g. lost productivity due to disease, were available.

No data were available to explore differences by social groups.

For the scenarios, smokers were moved to never smokers in order to account for change in uptake. This will overestimate the impact of the tax increases.

No in-depth uncertainty analysis was conducted.

Full methods

Data Collection

Smoking Prevalence Data
Baseline smoking prevalence data was extracted from the 2015 Annual
Household Survey, conducted by the National Statistic Service in Ukraine (7).
Additional data on percentages of occasional smokers and ex-smokers were
extrapolated from the omnibus surveys conducted by Kiev International
Institute of Sociology (2, 3).

Data manipulation and assumptions

1. Daily versus current smokers: The Annual Household Survey provides
prevalence data on daily smoking only, rather than current smoking. Current
smoking data are preferable, since the WHO target is focused on a total
reduction in smoking prevalence, rather than number of cigarettes smoked.
Modeling proceeded with a focus on prevalence of current smoking. Pooled
estimates from other smaller surveys with more detailed collection of smoking
status data were collated and the 2015 Household survey adjusted to include
estimated proportions of occasional smokers.

2. Ex-smokers: No data by age and sex were available for percentages of
ex-smokers within the Annual Household Survey. However, some data
were available from the recent omnibus surveys (2013-2015). Therefore, in
order to take account of ex-smokers (who have a greater disease risk than
never smokers), the distributions of ex-smokers and never smokers from the
omnibus surveys were used to proportion out the non-smoker data into ex-
smoker and never smoker from the Annual Household Survey. This enabled
us to initialize the model in the start year with a more accurate estimation of
ex-smokers than would be done using proxy ex-smoker data.

3. Sample sizes: Often, sample sizes by age group were not presented,
therefore the total sample size was proportioned across the five-year age
groups and the variance increased based on data from the UN population
prospects database (8).

4. Age groups: Data for some years were in wide age groups of more than
20 years (e.g., 30-59 years), therefore prevalence was assumed to be the same
across these groups. Once raw/more detailed data/data by five-year age
groups become available, then the data can be updated.

Disease Data

For this study, the following smoking-related NCDs were modeled: 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 Global Burden of Disease study database (9). Lung cancer data were grouped with trachea and bronchus data in the database, which slightly overestimates cases relative to Globocan (10). No survival data were available for these diseases in Ukraine, therefore survival data were calculated using DISMOD equations from the World Health Organization (11).

Relative risks (RR) for smokers compared to non-smokers were extracted from prospective cohort studies which observed the development of CHD (12-17), COPD (17-21), lung cancer (17, 21-23), and stroke (24-26). 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. However, if RRs for neighboring age groups from various studies differed much, some smoothing was undertaken. Appendix 3 describes the method of creating RRs in more detail.

For ex-smokers, RRs were assumed to decrease over time since cessation. The ex-smoker RR was computed using a decay function method developed by Hoogenveen and colleagues (27). This function uses the current smoker RR for each disease as the starting point and then models the decline in relative risk of disease for an ex-smoker over time, as detailed in Appendix 1 of the supplementary appendix.

Health-Economic Data

Data on direct health care costs by disease were extracted from the literature (28), but no data on indirect, non-healthcare costs by disease were available. Data on direct health care costs were included in the model and the direct healthcare cost impacts output from the model.

It was possible to calculate premature mortality costs by including average annual income.[1] This accumulates the lost earnings due to death before age 65 to provide a different measure of lost productivity in terms of losses of GDP due to death. However, this does not take account of losses in productivity due to morbidity. We carried out a sensitivity analysis on the costs, running the model with a discount rate of 5 percent, as is used in Russia (http://www.ispor.org/peguidelines/countrydet.asp?c=18&t=4). No discount rate was available for Ukraine.

[1] http://data.worldbank.org/indicator/NY.GDP.PCAP.CD?year_high_desc=true - average of 5 years taken from the World Bank and OECD national accounts data: $3320 per year

Table 2: **References for Disease Data**

	Incidence	Mortality	Survival	Direct healthcare costs
CHD	GBD 2016 (9)	GBD 2016 (9)	Converted from incidence and mortality	I Denisova, P Kuznetsova 2014 (28)
Stroke	GBD 2016 (9)	GBD 2016 (9)	Converted from incidence and mortality	I Denisova, P Kuznetsova 2014 (28)
COPD	GBD 2016 (9)	GBD 2016 (9)	Converted from incidence and mortality	I Denisova, P Kuznetsova 2014 (28)
Lung Cancer	GBD 2016 (9)	GBD 2016 (9)	Converted from incidence and mortality	I Denisova, P Kuznetsova 2014 (28)

Population Data
In order to simulate the population of Ukraine, the population by age and sex, births by mother's age, and total fertility rate statistics were taken from the UN population prospects database (8).

Total mortality rates were taken from the WHO global health estimates database (29).

These parameters enable the model to simulate the Ukrainian population as close to reality as is possible.

The Microsimulation Model

The UK Health Forum (UKHF) microsimulation model was originally developed for the English government's Foresight enquiry (30, 31) and has been further developed over the past decade to incorporate a number of additional interacting risk factors, including smoking. (Methods are described in greater detail in (32, 33) and in our supplementary appendix 1.) The model simulates a virtual population that reproduces the characteristics and behavior of a large sample of individuals (20-100 million). These characteristics (age, sex, smoking 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.

Figure 1 illustrates the modular nature of the model.

Module 1 uses cross-sectional data on the prevalence of the risk factor - cigarette smoking in this case. For the current study, 2015 smoking prevalence data for Ukraine was extrapolated forward to 2035. It was assumed that the proportions of the population within each smoking category as calculated in 2015 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.

The model produces a wide range of different outputs, including incidence, cumulative incidence, prevalence, premature mortality, direct healthcare costs avoided, and disability-adjusted life years.

To our knowledge, no other studies have used a microsimulation model to quantify the future costs and health impacts of tobacco taxation policy scenarios in Ukraine.
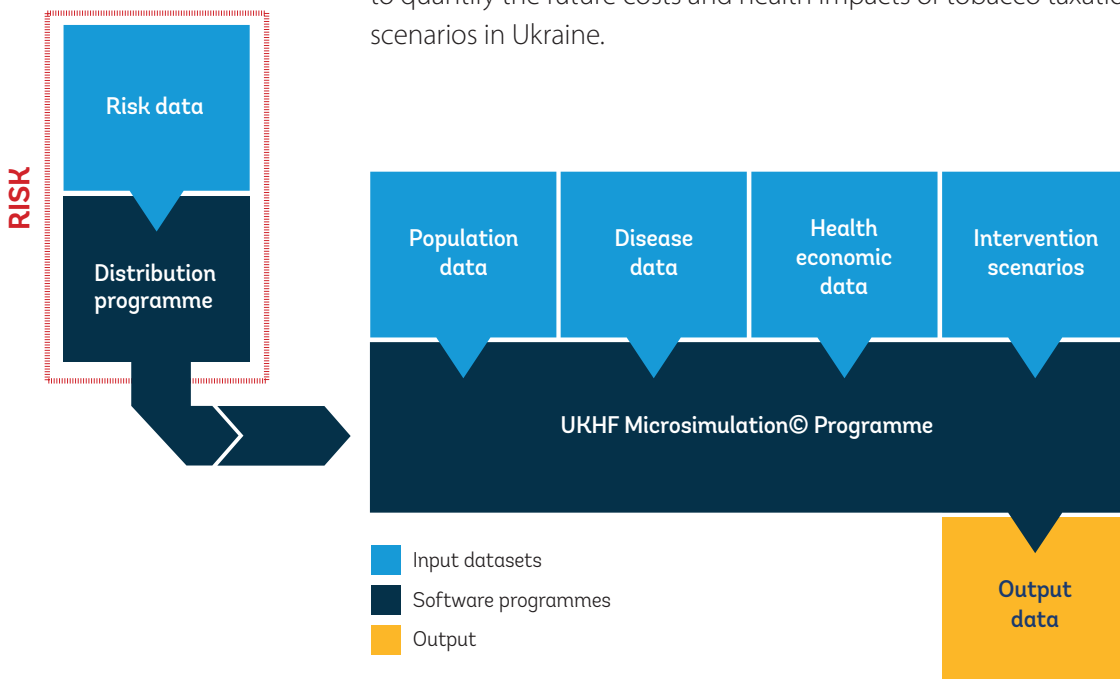
Figure 1: **Illustration of the Microsimulation Model**

Development of Scenarios

An initial modeling study was carried out using the World Health Organization (WHO) TaXSiM model.[2] 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 (Table 3).

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. Details of the TaXSiM model scenarios can be found in Table 3.

[2] WHO tobacco tax simulation model (TaXSiM) http://who.int/tobacco/economics/taxsim/en/

Table 3: **TaXSiM Model Scenarios and Outputs**

	Actual 2015	Expected Contri-bution to GDP	Baseline Situation (2016): Ad valorem (12%) minimum specific (8.515 UAH) and simple specific (6.365 UAH)	Expected Contri-bution to GDP	SCENARIO 1 (2017): Ad valorem tax is the same as in 2016 (12%), and 40% increase in both the minimum specific excise (11.92 UAH), and simple specific (8.91 UAH)**	Expected Contri-bution to GDP
Total cigarettes taxed (billion pieces)	74.0		77.0		70.1	
Average cigarette price (UAH per pack)	15.3		20.8		25.7	
Average cigarette price (US$ per pack).*	$0.63		$0.81		$0.92	
Average excise tax (UAH per 1000 pieces)	308.9		431.4		600.0	
Total excise tax revenue (billion UAH)	22.9	1.0%	33.2	1.4%	41.8	1.6%
Total excise tax revenue (US$ billion).*	$0.94		$ 1.30		$1.50	
Total government revenue (excise, VAT and levies, billion UAH)	34.9	1.6%	49.9	2.2%	60.1	2.3%
Total government revenue (excise, VAT and levies, US$ billion).*	$1.44		$ 1.95		$2.16	
Total expenditure on cigarettes (billion UAH)	56.4		79.9		90.0	
Percentage change in total cigarette consumption (%)			4.1		-9.0	

* World Bank Group actual (2015 -2016) and forecast (2017): Annual average exchange rate = 2016 (1US$/25.6 UAH); 2017 (1US$/27.80 UAH)
** per pack of 20 cigarettes

Scenario Assumptions

1. Several studies suggest that around 50 percent of the effect of price increases on overall cigarette consumption results from participation changes (34, 35). Therefore, 50 percent of the estimated reduction in cigarette consumption was used as an estimate of the reduction in the total prevalence of smoking. While taxation which results in increased real prices of tobacco might reduce the intensity of smoking, research suggests that people who cut down actually inhale more, as measured by serum cotinine levels (36). 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.

2. Our analysis of the omnibus surveys 2013-2015 showed that, in males, 55 percent of the change in smoking prevalence was due to a reduction in uptake. Specifically, the percentage of smokers decreased, the percentage of ex-smokers did not change, while the percentage of never smokers increased. Therefore, these changes were probably due to males' "not starting smoking." Among females, 100 percent of the change in consumption was due to "not starting smoking" (2).

3. While these average changes were not the same for each group, and people usually initiate smoking while they are under 30 years old (2), the model did not take these age differences into account, and the relative decline in percentages of current smokers was applied to all age groups.
Therefore, it was assumed that taxation would result in changes in uptake.

4. A baseline "static" trend was included. This assumed that smoking prevalence remains constant at 2015 rates. The tax increase scenario was compared to this baseline.

5. The tax increase scenario represents the tax change adopted in January 2017.

6. The change in smoking prevalence occurs in the second year of the simulation only (2017). This is in line with TaXSiM. As noted above, tobacco companies' "price wars" were observed to cause an increase in cigarette consumption, rather than the decrease expected according to taxation policies. Taking this into account, it was assumed that, in 2016, there would be no change in smoking prevalence from the 2015 level (37). The baseline used in the model was 2015 smoking prevalence held constant.

7. The scenarios are based on Monte Carlo simulations (Individuals were sampled from the population and simulated through).

8. The specified percentage of smokers who are affected by the tax increase move to the never-smoker category in 2017.

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

10. We assumed an immediate reduction in smoking prevalence due to the tax increases in 2017. We learned via personal communication with Prof. 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) (38). Townsend and Atkinson take the opposite view (39): That smokers tend to react quickly to a price change. We used a model similar to theirs, with an immediate effect and then a linear trend, and in line with the TaXSiM model outputs.

There were two scenarios:

1. A baseline 'static' trend. This assumed that smoking prevalence stays constant at 2015 rates.

2. A tax increase scenario. An earlier iteration of TaXSiM calculated that an increase in Ad valorem tax of 15 percent, a 30 percent increase in the minimum specific excise, 11.08 UAH, and a simple specific of 8.28 UAH would result in a reduction of 10.2 percent in cigarettes smoked. Using the assumptions above, this translated into a reduction in uptake of 5.61 percent in males, and 10.2 percent in females. Therefore, in 2017, this specified percentage of smokers was moved to the never-smoker group in order to take account of uptake and maintain 100 percent of smokers in the model (the population cannot exceed 100 percent). This slightly underestimates the effect of the scenario, as described in the discussion. This change occurred in 2017 only. Appendix 2 provides the full TaxSiM analysis from the earlier iteration.

Scenario 1 is summarized in Table 4:

Table 4: **Summary of Scenarios**

% reduction in cigarettes consumption as per Table 2	Estimated expected reduction in smoking (males)		Estimated expected reduction in smoking (females)	
(%)	Uptake (%)	Number of cigarettes smoked (%)	Uptake (%)	Number of cigarettes smoked (%)
10.2	5.61	4.59	10.2	0

Results

A number of different outputs are produced from the model, and these are defined below:

Smoking Prevalence (%)

Table 5 shows smoking prevalence for males, females, and both males and females combined for the baseline and Scenario 1.

Table 5: **Smoking Prevalence by Year, Sex and Scenario (%)**

Year	Scenario 0 (baseline)			Scenario 1		
	M	F	TOTAL	M	F	TOTAL
2016	40.8	5.7	21.5	40.8	5.7	21.5
2020	40.3	5.4	21.1	38.2	4.9	19.8
2025	39.4	5.3	20.5	37.5	4.8	19.4
2030	38.9	5.1	20.1	37.2	4.8	19.2
2035	38.2	5.0	19.8	36.7	4.7	19.0
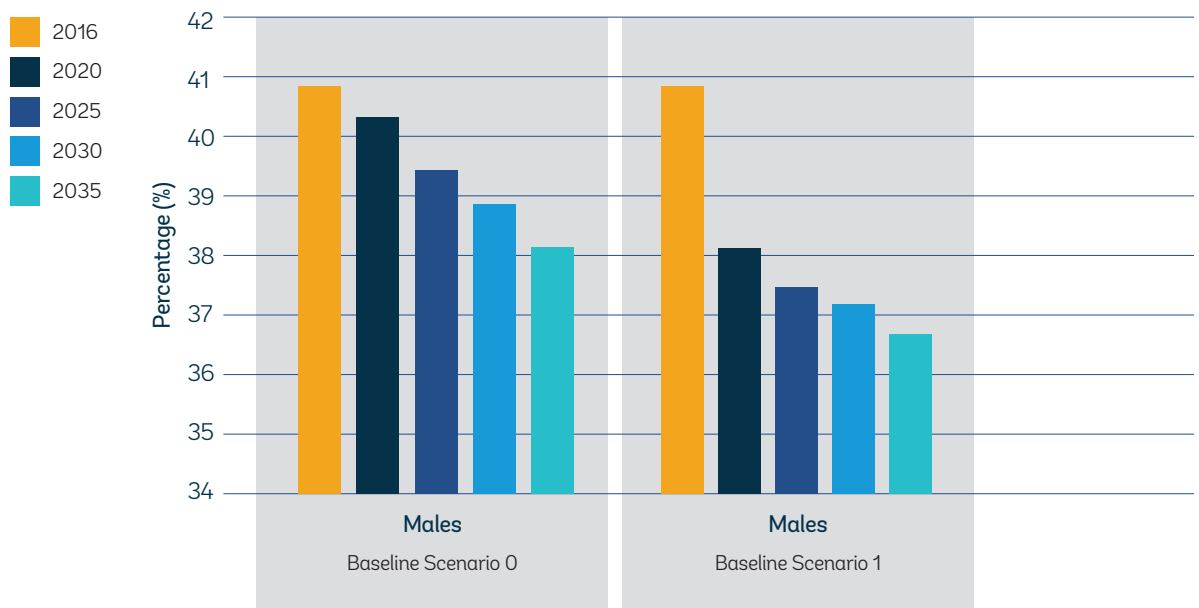
Figure 2: **Male Smoking Prevalence by Year for Each Scenario.**
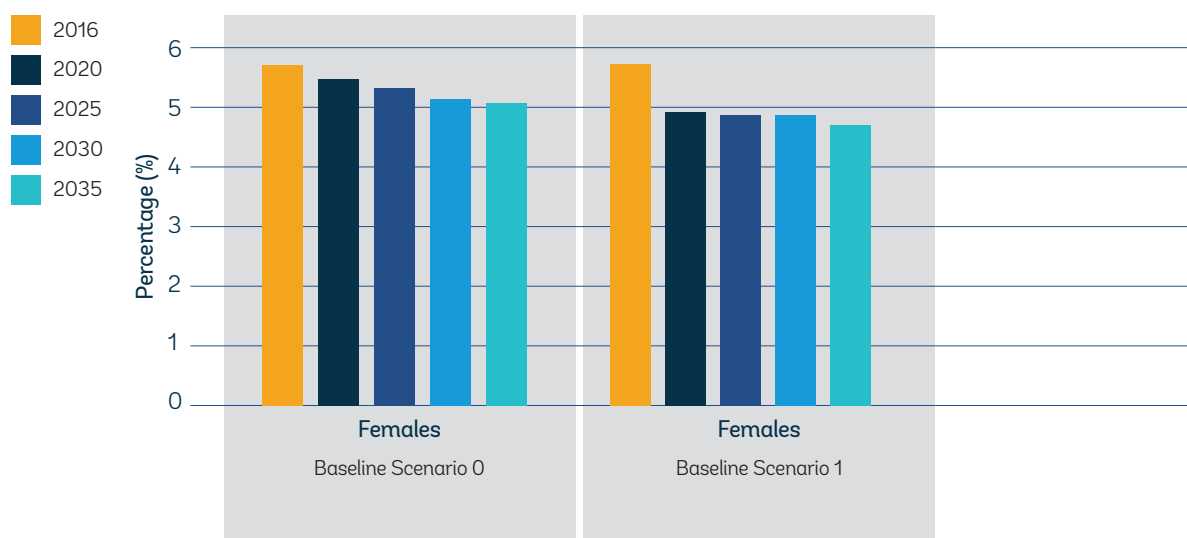
Figure 3: **Female Smoking Prevalence by Year for Each Scenario.**

Epidemiological Indicators

Results from the microsimulation are presented as rates per 100,000, then scaled to the Ukraine population for that year, as estimated by the UN population prospects (8).

1. Cumulative incidence rate per year per Ukraine population
The total number of new cases of a disease 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 incident cases represents a sum of all of the incident cases from the start of the simulation.

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

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

4. Attributable incidence rate per Ukraine population per year
The number of new cases of a disease attributable to being a smoker or ex-smoker in the Ukraine population.

15

5. Premature mortality rates per Ukraine population

Premature mortality refers to the total number of deaths in a given year below the life expectancy of that individual in the Ukraine population. Results are presented per year in the total population and cumulative over a given period of the simulation.

6. Potential years of life lost per Ukraine population

For each individual, the difference between the reference age (life expectancy at birth) and the age of death is calculated. The average annual PYLL was calculated each year in the microsimulation. This metric considers individuals who have died in a given year and is output as a rate per 100,000, which is then scaled to a rate per Ukraine population.

Economic outputs

7. Direct cost 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 Ukraine.

8. Premature mortality costs

This relates to lost earnings due to premature deaths. The premature mortality costs for each individual in the year of death are calculated by summing over the income costs from the age of death until the individual's life expectancy (LE) at birth.

Summary Table

Table 6 presents a summary table of total disease cases (epidemiological) and costs (economic) by parameter, year, and scenario as rates per Ukraine population.

Scenario 0 (Sc0) refers to the baseline scenario where smoking prevalence was assumed constant based on 2015 smoking prevalence.

Scenario 1 (Sc1) refers to the one-off tax scenario as summarized in table 3.

Cumulative Incident Cases

Table 7 presents the cumulative incident cases for each disease by year, and Table 8 presents the cumulative incident cases avoided.

Table 6: **Summary Table of the Outputs as Rates per Ukraine Population, by Year**

Epidemiological outputs	Year	Sc0 (Baseline)	Sc1
Cumulative incident cases	2025	5480948[±4237]	5427558[±4237]
	2035	11366868[±5753]	11255173[±5753]
Cumulative incident cases avoided	2025	NA	56224[±6341]
	2035	NA	126730[±9123]
Incident cases per year	2025	589035[±1545]	582341[±1545]
	2035	646600[±1545]	640799[±1727]
Attributable incident cases	2025	218221[±1121]	208475[±1121]
	2035	222603[±1041]	211984[±1041]
Cumulative premature deaths avoided	2025	NA	6372
	2035	NA	29172
Cumulative potential years of life lost relative to baseline	2025	NA	48923
	2035	NA	267098
Economic outputs			
Direct costs avoided (millions UAH)	2025	NA	542.23
	2035	NA	1545.81
Cumulative premature mortality costs avoided (millions UAH)	2025	NA	3568.4
	2035	NA	16536.4

Table 7: **Cumulative Incident Cases for Each Disease by Year for the Total Ukraine Population**

		CHD	COPD	Lung Cancer	Stroke	Total
Year 2025	Sc 0	3712722[+-3390]	665256[+-1695]	194068[+-847]	908901[+-1695]	5480948[±4237]
	Sc 1	3679248[±3390]	655510[±1695]	188136[±847]	904664[±1695]	5427558[±4237]
Year 2035	Sc 0	7707697[+-4719]	1346232[+-1966]	392110[+-1180]	1920828[+-2360]	11366868[±5753]
	Sc 1	7638872[±4719]	1323421[±1966]	379132[±1180]	1913749[±2360]	11255173[±5753]
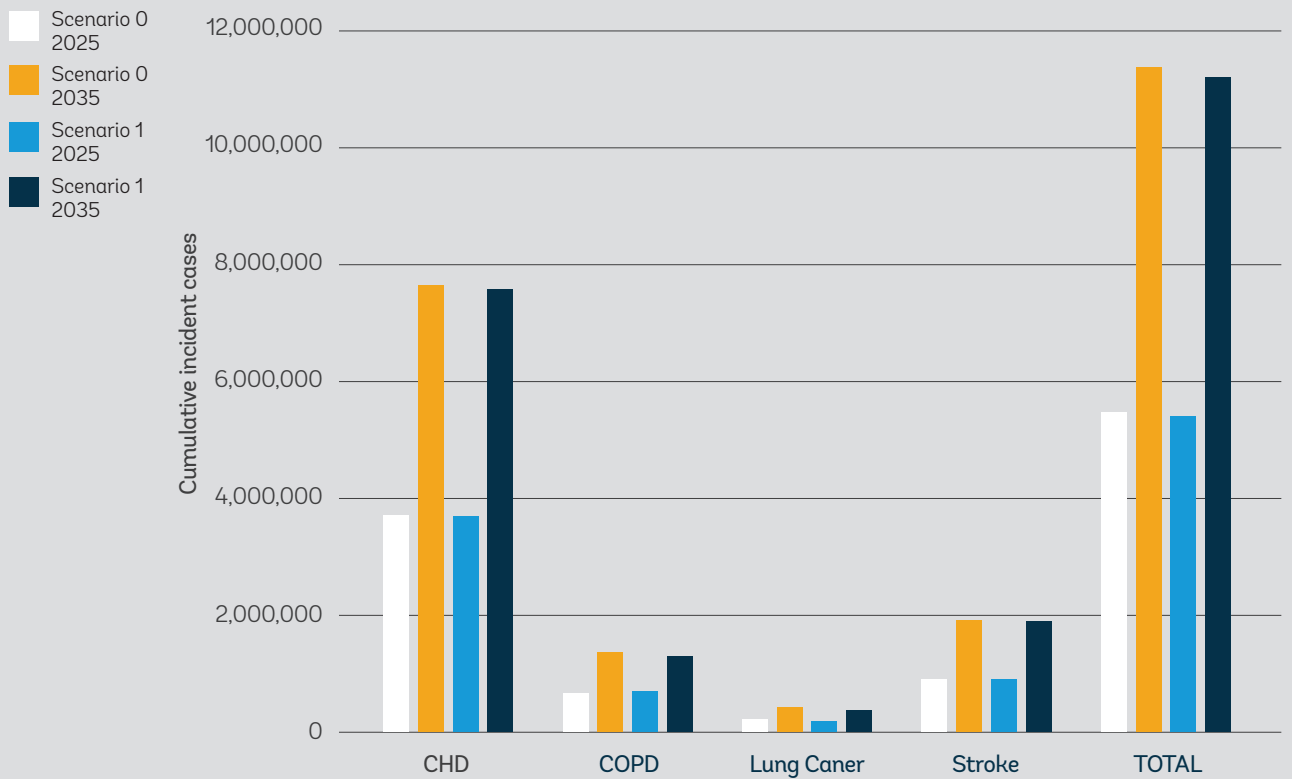
Figure 4: **Cumulative Incident Cases per Ukraine Population by 2025 and 2035.**

Table 8: **Cumulative Incident Cases Avoided Relative to Scenario 0 for the Ukraine Population by 2025 and 2035**

	CHD	COPD	Lung Cancer	Stroke	Total
Year 2025	35252[+-4908]	10263[+-2677]	6247[+-1338]	4462[+-2677]	56224[±6341]
Year 2035	78092[+-7586]	25881[+-3123]	14725[+-1784]	8032[+-3569]	126730[±9123]
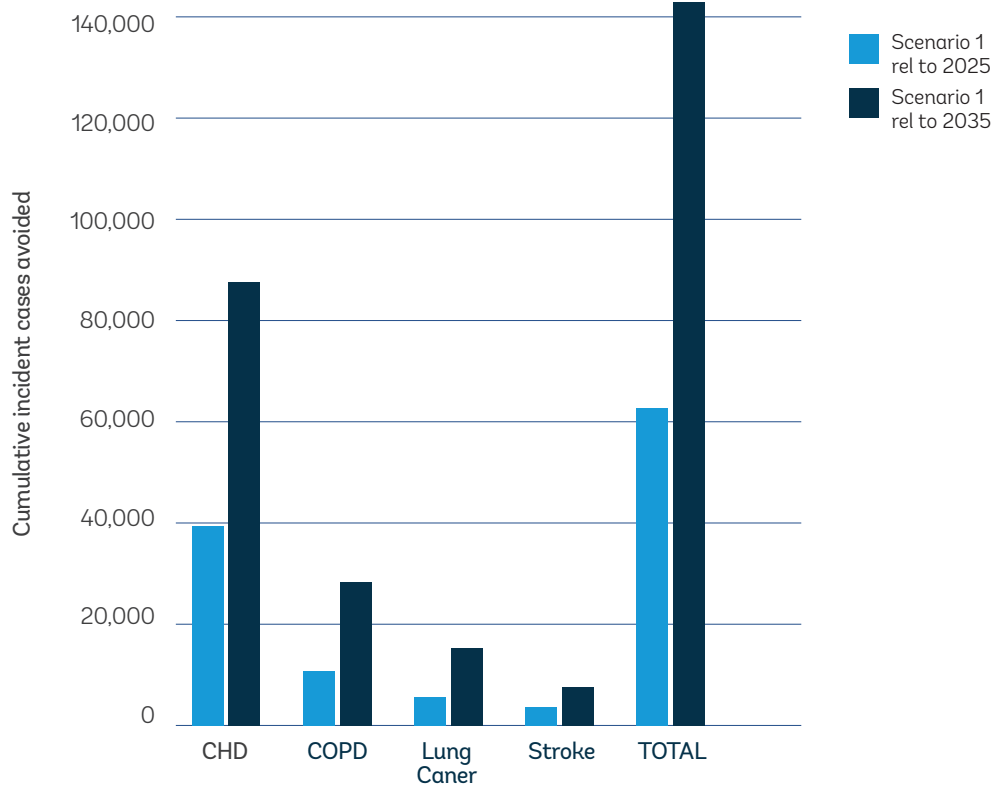
Figure 5: **Cumulative Incident Cases Avoided.**

Incident and Attributable Incident Cases per Year

Table 9 presents the incidence rates per Ukraine population, and Table 10
shows the incidence rate attributable to smoking per Ukraine population by
year for each disease. Figure 7 presents the incident cases by scenario for 2025
by disease. The blue bars show incident cases and red bars show incident
cases attributable to smoking in the specified year per Ukraine population. For
scenario 1, the cases attributable to smoking contribute a smaller portion to
the overall new cases when compared with baseline (scenario 0). This is to be
expected, since the scenario is impacting smokers, so we would expect the
avoided cases attributable to smoking to increase over time.

Table 9: **Incident Cases in the Total Population per Year**

		CHD	COPD	Lung Cancer	Stroke	Total
Year 2025	Sc 0	398492[+-1338]	70059[+-446]	20080[+-446]	100404[+-446]	589035[±1545]
	Sc 1	394476[+-1338]	68274[+-446]	19634[+-446]	99957[+-446]	582341[±1545]
Year 2035	Sc 0	439546[+-1338]	73629[+-446]	21419[+-446]	112006[+-446]	646600[±1545]
	Sc 1	435976[+-1338]	71844[+-446]	20527[+-446]	112452[+-892]	640799[±1727]
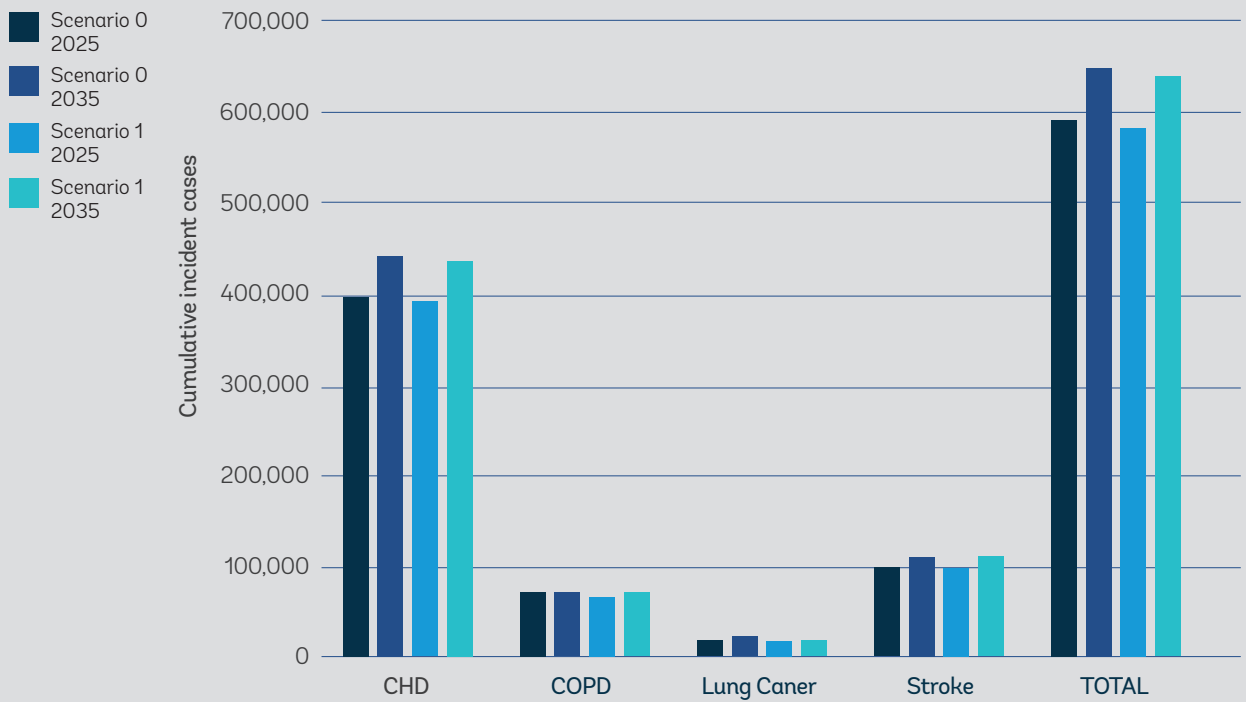
Figure 6: **Incident Cases in Ukraine in 2025 and 2035.**

Table 10: **Attributable Incident Cases per Year**

		CHD	COPD	Lung Cancer	Stroke	Total
Year 2025	Sc 0	145339[+-847]	35170[±424]	15678[±424]	22034[±424]	218221[±1121]
	Sc 1	138983[+-847]	33051[±424]	15254[±424]	21187[±424]	208475[±1121]
Year 2035	Sc 0	148664[+-787]	36183[±393]	15338[±393]	22418[±393]	222603[±1041]
	Sc 1	142371[+-787]	33823[±393]	14552[±393]	21238[±393]	211984[±1041]
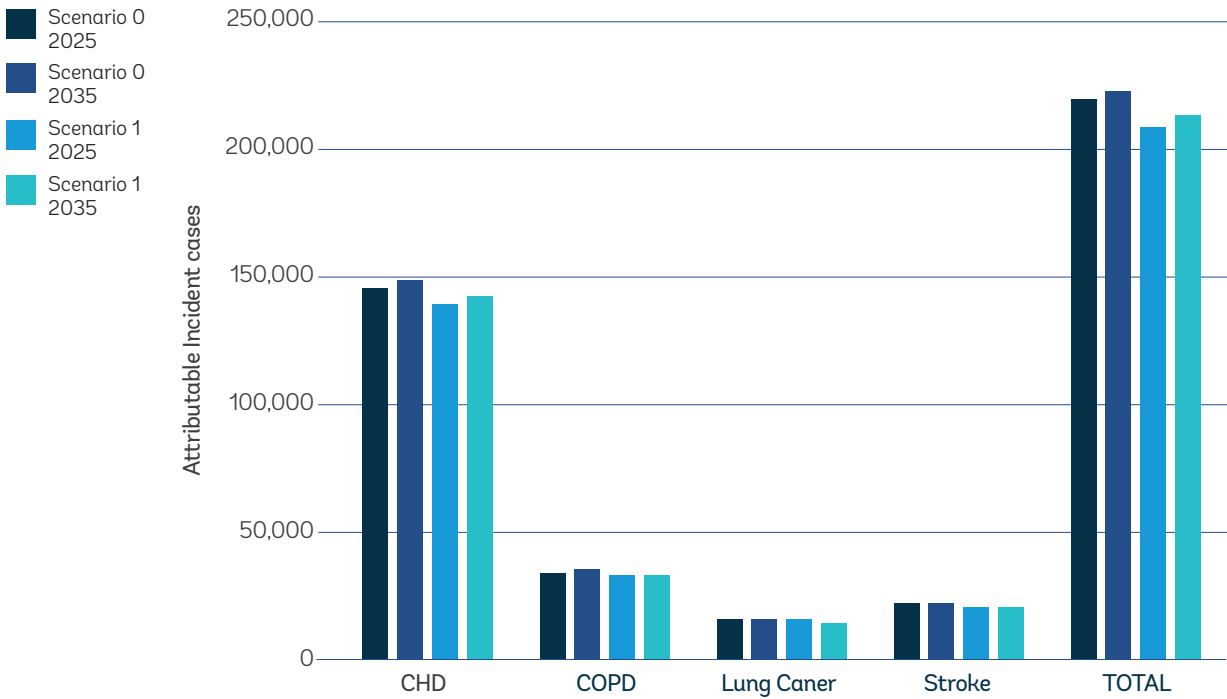
Figure 7: **Attributable Incident Cases per Ukraine Population
in the Years 2025 and 2035.**
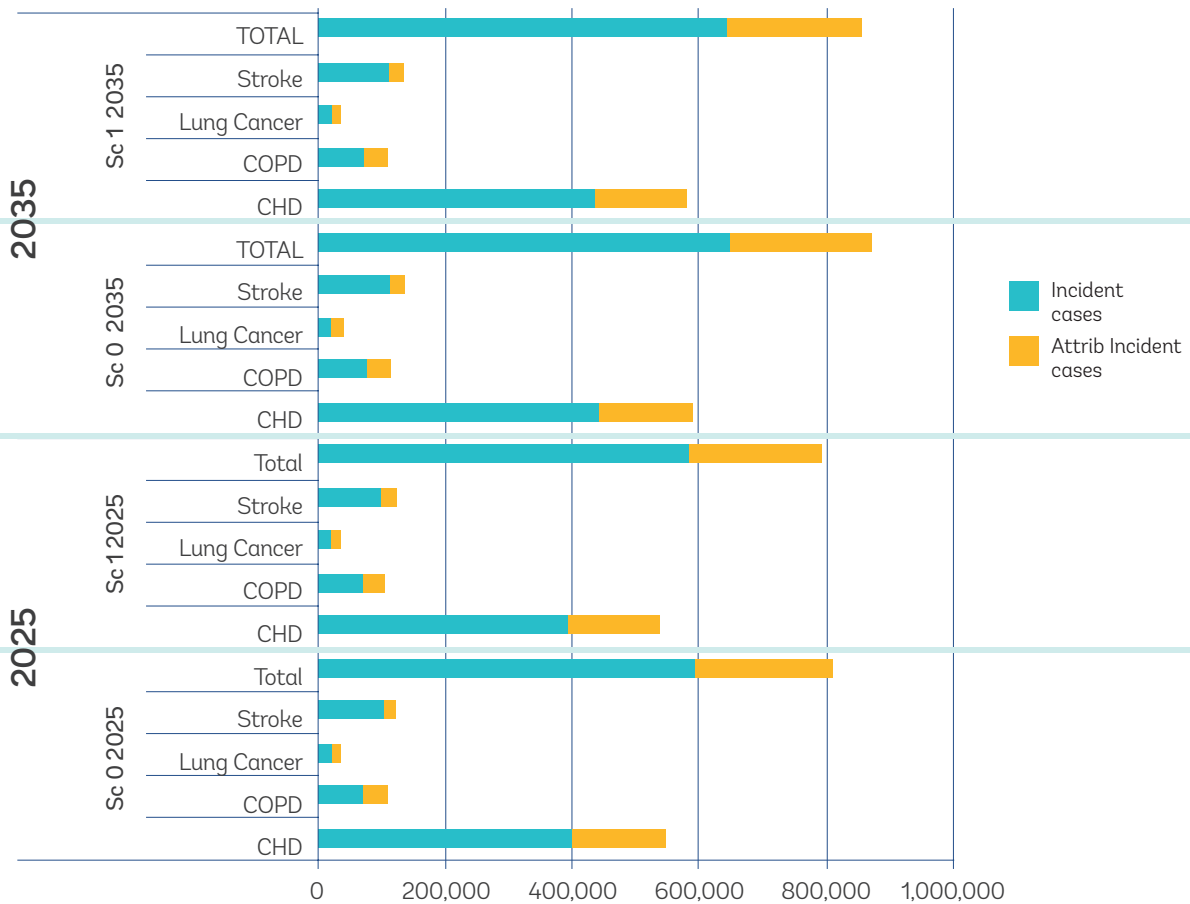
Figure 8: **Incident and Attributable Incident Cases in 2025 for
Baseline and Scenario 1 by Disease.**

Premature Deaths

Table 11 presents the premature deaths, premature deaths avoided, and cumulative premature deaths avoided in the total Ukraine population relative to scenario 0. The results show that, by 2025, there will be 6,372 premature deaths averted given the scenario 1 tax increase. By 2035, this increases to 29,172 premature deaths averted for scenario 1. Figure 8 presents the cumulative premature deaths avoided by scenario for 2025 and 2035.

Table 11: **Premature Mortality in the Total Ukraine Population**

		Premature deaths	**Premature deaths avoided**	**Cumulative premature deaths avoided**
Year 2025	Sc 0	307204	NA	NA
	Sc 1	305086	2119	6372
Year 2035	Sc 0	328004	NA	NA
	Sc 1	326824	1180	29172
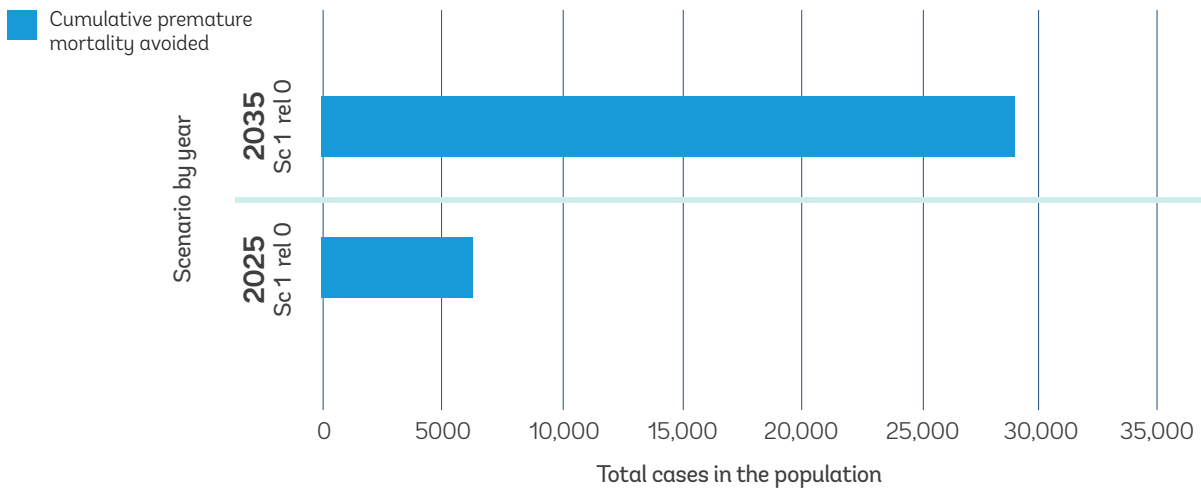
Figure 9: **Cumulative Premature Deaths Avoided in the Ukraine Population by 2025 and 2035.**

Potential Years of Life Lost

Table 12 presents the cumulative potential years of life lost (PYLL) for each scenario by year in the total Ukraine population. By 2025, scenario 1 is predicted to avoid 267,098 PYLL relative to baseline.

Table 12: **Cumulative PYLL by scenario and year, and PYLL avoided due to scenario 1 relative to scenario 0**

Year	Sc 0	Sc 1	PYLL avoided Sc1 rel 0
2025	47173316	47124394	48923
2035	92950950	92683852	267098

Direct Costs Avoided

Table 13 presents the cumulative direct healthcare costs avoided for scenario 1 relative to scenario 0. Relative to scenario 0, scenario 1 results in the avoidance of the following direct healthcare costs by disease: CHD (UAH1.1bn/US$130 million[3]), followed by COPD (UAH 0.16bn/US$25 million).

Table 13: **Direct Cumulative Healthcare Costs Avoided (UAH millions)**

		CHD	COPD	Lung Cancer	Stroke	Total
Year 2025	Sc 1 rel 0	408.37 [+-0.06]	26.16 [+-0.06]	54.05 [+-0.06]	53.65 [+-0.06]	542.22 [+-0.12]
Year 2035	Sc 1 rel 0	1133.1 [+-0.08]	160.54 [+-0.08]	143.35 [+-0.08]	108.82 [+-0.08]	1545.82 [+-0.16]

Figure 10: **Direct Cumulative Healthcare Costs Avoided (UAH millions)**
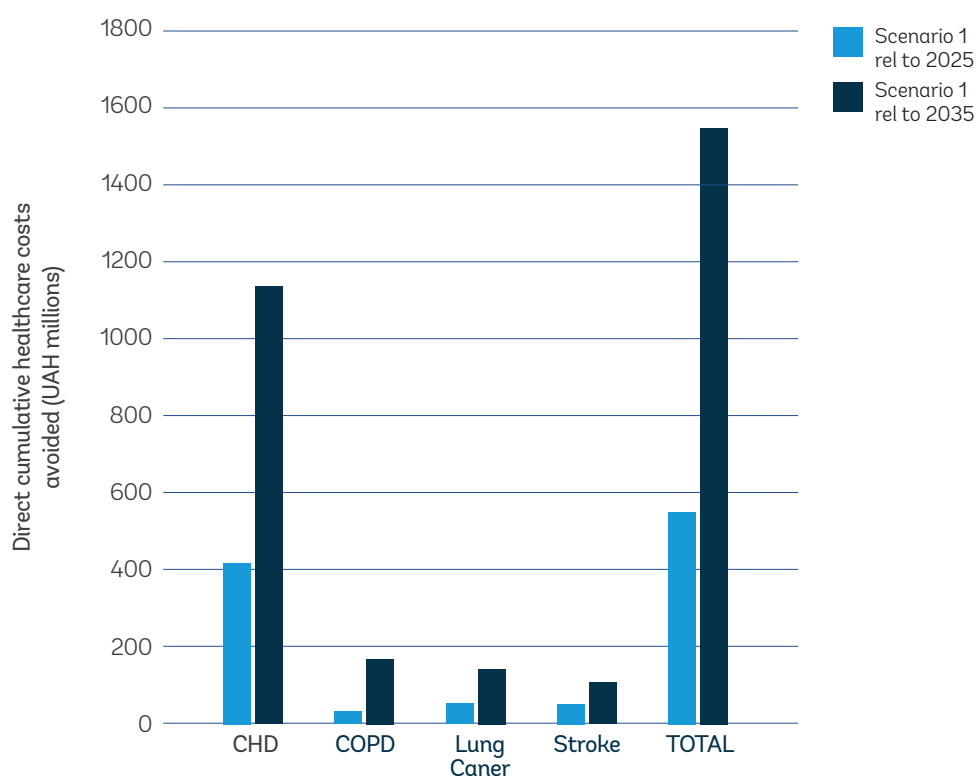
[3] The exchange rate of 1US$/23.8 UAH is used here.

23

As expected, discounting at 5 percent has a large impact on the cumulative direct costs avoided by 2025 and 2035. For example, scenario 1 is predicted to result in 399 million UAH avoided by 2025, compared to 542 million UAH avoided by 2025 without discounting. The results are presented in Table 14.

Table 14: **Direct Cumulative Healthcare Costs Avoided (UAH millions) (with 5 percent discounting)**

		CHD	COPD	Lung Cancer	Stroke	Total
Year 2025	Sc 1 rel 0	301.18	16.66	40.76	40.13	398.74
Year 2035	Sc 1 rel 0	660.51	82.23	85.24	68.36	896.34

Premature Mortality Costs Avoided

Table 14 presents the premature mortality costs avoided, relative to baseline. In 2035 alone, UAH 1.97 billion (US$82.7 million) premature mortality costs could be avoided. Cumulatively, by 2035, UAH 16.5 billion (US$695 million) premature mortality costs could be avoided relative to baseline.

Table 15: **Premature Mortality Costs Avoided (UAH millions)**

	PM costs avoided	Cumulative costs avoided
Year 2025	206	3568.4
Year 2035	1968.3	16536.4

Discussion

This study explored the impact of a one-time tobacco tax increase in Ukraine on the future burden of four smoking-related diseases through 2035. The results showed that small changes in smoking prevalence in one year can have large impacts in terms of disease incidence and premature mortality cases avoided into the future. Our results show that implementation of a one-off tax has an impact on smoking-related health burden, but highlights the need for continuous tobacco control measures, if smoking prevalence is to continue to decrease and have sizeable impacts on related disease occurrence.

As well as benefits in terms of morbidity, particularly CHD cases avoided, we observe large savings in terms of premature mortality and potential years of life lost. This is important, since Ukraine is experiencing a decreasing population over time, and a lower life expectancy compared to the EU average of 78 for males and around 83 for females (40). Tobacco taxation is one important step to improving life expectancy in Ukraine, especially amongst men, whose smoking prevalence is high.

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 (41). 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.

While the microsimulation method is advantageous in NCD modeling, one key disadvantage is that the model is data intensive. Fortuitously, during the period of the study, the Global Burden of Disease team published an online database that included many of the data inputs that were required (9). While country-specific data are preferable, and the GBD is based on modeled estimates and recommended as a cross-country comparative tool, few other data were available for Ukraine. Inter alia, there were neither survival data nor relative risks specific for Ukraine available. Once better data become available, the model can easily be updated.

No data on indirect costs such as productivity losses by disease were available. Large savings to the health system were observed with just a small change in smoking prevalence. However, wider societal costs such as losses in productivity are likely to be higher than those reflected here, making a stronger case for the implementation of regular tax hikes for tobacco control (42). If indirect cost data by disease become available, then the model can once again easily be updated in the future.

One notable limitation of our scenario methodology is that smokers were moved to the never-smokers category to account for changes in uptake due to the tax. While this is not realistic, it was the only solution by which to model change in uptake within the total population (and ensure we maintain 100 percent of people in the population). This approach could result in an underestimation of the health impact of a tobacco tax increase. This effect could arise, because some of those smokers who become never smokers may already have a smoking-related disease.

We know that social groups react differently to tax increases (43). Due to small sample sizes, it was not possible to model the long-term health impacts on different social groups within the microsimulation. However, we can infer from research conducted in Ukraine (43, 44) that the largest impact of taxation will be observed in the poorest social groups. This is important, since it means that tobacco taxation could contribute to reducing social inequalities in health. 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 additional scenarios in the future relative to a "no change" scenario.

At present, the model does not take account of multimorbidity 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 by 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 time scale. However, we did carry out a small sensitivity analysis of the costs – running the model with and without a 5 percent discount rate. Further work should develop more sophisticated interventions, for example,

individuals of different ages could be affected differently by the intervention. It was beyond the scope of this study to include this development within the microsimulation. However, a prototype, user-friendly tool has been developed for Ukraine that enables the user to select different age cohorts (as opposed to a population distribution of individuals) and run simulations to quantify health and cost impacts by population groups.

Further work should also explore the impact of other potential policies in Ukraine, such as a tobacco duty escalator, as well as a combination of tobacco-control measures including smoking cessation services.

This study complements that which was carried out using the TaXSiM model and shows the health and related economic benefits of increasing tobacco tax in Ukraine. Even small reductions in smoking prevalence in one year will have long-term impacts on disease incidence and subsequent health costs.

Bibliography

1. WHO. WHO report on the global tobacco epidemic 2011: *Warning about the dangers of tobacco.* Geneva: WHO; 2011.

2. Andreeva T. Results of omnibus surveys with tobacco-related questions conducted in Ukraine in 2013, 2014, 2015. 2016.

3. The results of KIIS survey on tobacco smoking in Ukraine as of 2015 compared to 2013-2014 Kiev: Kiev International Institute of Sociology; 2016. Available at: http://www.kiis.com.ua/?lang=eng&cat=reports&id=587&page=1.

4. Oderkirk J, Sassi F, Cecchini M, Astolfi R, OECD Health Division. Toward a New *Comprehensive International Health and Health Care Policy Decision Support Tool.* OECD Directorate for Employment, Labour and Social Affairs; 2012.

5. Feenberg D, Coutts E. An introduction to the TAXSIM model. *Journal of Policy Analysis and Management.* 1993;12(1):189-94.

6. Butrica BA, Burkhauser RV. Estimating federal income tax burdens for Panel Study of Income Dynamics (PSID) families using the National Bureau of Economic Research TAXSIM model. 1997.

7. State Statistical Service of Ukraine. Population's self-perceived health status and availability of selected types of medical aid in 2015 (in Ukrainian). Kiev: State Statistical Service of Ukraine; 2016. Available at: http://ukrstat.gov.ua/druk/katalog/kat_u/2015/sb/zb_snsz_2015.zip.

8. United Nations. *World population prospects 2015*. Available at: http://esa.un.org/unpd/wpp/.

9. Global Burden of Disease. Global Health Data Exchange. In: Institute for Health Metrics and Evaluation, editor. Available at: http://ghdx.healthdata.org/gbd-results-tool2016.

10. World Health Organization. Globocan 2012: Estimated Cancer Incidence, Mortality and Prevalence Worldwide in 2012. Available at: http://globocan.iarc.fr/Default.aspx.

11. World Health Organization. Health statistics and information systems - Software tools - DISMOD II 2014 [23/02/15]. Available at: http://www.who.int/healthinfo/global_burden_disease/tools_software/en/.

12. Song YM, Cho HJ. Risk of stroke and myocardial infarction after reduction or cessation of cigarette smoking: A cohort study in Korean men. *Stroke.* 2008;39(9):2432-8.

13. Baba S, Iso H, Mannami T, Sasaki S, Okada K, Konishi M, et al. Cigarette smoking and risk of coronary heart disease incidence among middle-aged Japanese men and women: The JPHC Study Cohort I. *European Journal of Cardiovascular Prevention & Rehabilitation.* 2006;13(2):207-13.

14. Tolstrup JS, Hvidtfeldt UA, Flachs EM, Spiegelman D, Heitmann BL, Balter K, et al. Smoking and risk of coronary heart disease in younger, middle-aged, and older adults. *American Journal of Public Health.* 2014;104(1):96-102.

15. Burns DM. Epidemiology of smoking-induced cardiovascular disease. *Progress in Cardiovascular Diseases.* 2003;46(1):11-29.

16. Cronin EM, Kearney PM, Kearney PP, Sullivan P, Perry IJ. Impact of a national smoking ban on hospital admission for acute coronary syndromes: A longitudinal study. *Clinical Cardiology.* 2012;35(4):205-9.

17. U.S. Department of Health and Human Services. *The health consequences of smoking—50 years of progress: a report of the Surgeon General.* Washington, DC: U.S. Department of Health and Human Services; 2014.

18. Prescott E, Bjerg AM, Andersen PK, Lange P, Vestbo J. Gender difference in smoking effects on lung function and risk of hospitalization for COPD: Results from a Danish longitudinal population study. *The European Respiratory Journal.* 1997;10(4):822-7.

19. Johannessen A, Omenaas E, Bakke P, Gulsvik A. Incidence of GOLD-defined chronic obstructive pulmonary disease in a general adult population. *International Journal of Tuberculosis and Lung Disease.* 2005;9(8):926-32.

20. Terzikhan N, Verhamme KMC, Hofman A, Stricker BH, Brusselle GG, Lahousse L. Prevalence and incidence of COPD in smokers and non-smokers: The Rotterdam Study. *European Journal of Epidemiology.* 2016;31(8):785-92.

21. Thun MJ, Carter BD, Feskanich D, Freedman ND, Prentice R, Lopez AD, et al. 50-Year Trends in Smoking-Related Mortality in the United States. *New England Journal of Medicine.* 2013;368(4):351-64.

22. Freedman ND, Leitzmann MF, Hollenbeck AR, Schatzkin A, Abnet CC. Cigarette smoking and subsequent risk of lung cancer in men and women: Analysis of a prospective cohort study. *Lancet Oncol.* 2008;9(7):649-56.

23. Bae JM, Lee MS, Shin MH, Kim DH, Li ZM, Ahn YO. Cigarette smoking and risk of lung cancer in Korean men: The Seoul male cancer cohort study. *J Korean Med Sci*. 2007;22(3):508-12.

24. Mannami T, Iso H, Baba S, Sasaki S, Okada K, Konishi M, et al. Cigarette smoking and risk of stroke and its subtypes among middle-aged Japanese men and women: The JPHC Study Cohort I. *Stroke*. 2004;35(6):1248-53.

25. Shinton R, Beevers G. Meta-analysis of relation between cigarette smoking and stroke. *BMJ*. 1989;298(6676):789-94.

26. Wannamethee SG, Shaper AG, Whincup PH, Walker M. Smoking cessation and the risk of stroke in middle-aged men. *Journal of the American Medical Association*. 1995;274(2):155-60.

27. Hoogenveen RT, van Baal PH, Boshuizen HC, Feenstra TL. Dynamic effects of smoking cessation on disease incidence, mortality and quality of life: The role of time since cessation. *Cost Eff Resour Alloc*. 2008;6:1.

28. Denisova I, Kuznetsova P. The effects of tobacco taxes on health : *An analysis of the effects by income quintile and gender in Kazakhstan, the Russian Federation, and Ukraine.* The World Bank; 2014 Oct.

29. World Health Organisation. World Health Statistics 2015. Global Health Observatory (GHO) data 2015. Available at: http://www.who.int/gho/publications/world_health_statistics/2015/en/.

30. McPherson K, Marsh T, Brown M. Foresight tackling obesities: Future choices – modelling future trends in obesity and the impact on health. Foresight Tackling Obesities Future Choices. 2007.

31. Wang YC, McPherson K, Marsh T, Gortmaker SL, Brown M. Health and economic burden of the projected obesity trends in the USA and the UK. *Lancet*. 2011;378(9793):815-25.

32. Forum CRUUH. *Aiming High: Why the UK should aim to be tobacco-free.* 2016.

33. Forum UH. Appendix B4. Detailed Methodology Technical Document. http://econdaproject.eu/2015.

34. Farrelly MC, Bray JW, Zarkin GA, Wendling BW. The joint demand for cigarettes and marijuana: Evidence from the National Household Surveys on Drug Abuse. *J Health Econ*. 2001;20(1):51-68.

35. Response to increases in cigarette prices by race/ethnicity, income, and age groups--United States, 1976-1993. *Morbidity and Mortality Weekly Report.* 1998;47(29):605-9.

36. Fidler JA, Stapleton JA, West R. Variation in saliva cotinine as a function of self-reported attempts to reduce cigarette consumption. *Psychopharmacology (Berl).* 2011;217(4):587-93.

37. Krasovsky K. Public health and revenue impact of cigarette "price wars" in Ukraine. ECTOH-20172017.

38. Becker G, Murphy K. A Theory of Rational Addiction. *Journal of Political Economy.* 1988;96(4):675-700.

39. Atkinson AB, Skegg JL. Anti-Smoking Publicity and the Demand for Tobacco in the U.K.*†. The Manchester School. 1973;41(3):265-82.

40. Eurostat: Statistics Explained. File: Life expectancy at birth, EU-28, 2002-14 2016 [cited 2016 20.12.2016]. Available at: http://ec.europa.eu/eurostat/statistics-explained/index.php/File:Life_expectancy_at_birth,_EU-28,_2002%E2%80%9314_%28%C2%B9%29_%28years%29_YB16.png.

41. Centers for Disease Control and Prevention. *Health Effects of Cigarette Smoking 2016.* Atlanta: Centers for Disease Control and Prevention; 2016. Available at: https://www.cdc.gov/tobacco/data_statistics/fact_sheets/health_effects/effects_cig_smoking/.

42. Action on smoking and health. The economics of tobacco. ASH factsheet. 2015.

43.   Krasovsky K. Sharp changes in tobacco products affordability and the dynamics of smoking prevalence in various social and income groups in Ukraine in 2008-2012. *Tob Induc Dis.* 2013;11(1):21.

44. Krasovsky K, Andreeva T, Krisanov D, Mashliakivsky M, Rud G. *The Economics of tobacco control in Ukraine from the public health perspective.* Kiev 2002. 128 p.

Appendix 1. Technical Appendix

1. Microsimulation Framework

Our simulation 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 analyzed risk factors. The detailed description of the two modules is presented below.

1.1 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, …, 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 a multinomial logistic regression model with prevalence of RF category k as the outcome, and time t as a single explanatory variable. For $k < N$ we have

$$\ln\left(\frac{p_k(t)}{p_1(t)} \right) = \beta_0^k + \beta_1^k t \tag{1.1}$$

The prevalence of the first category is obtained by using the normalization constraint $\sum_{k=1}^{N} p_k(t) = 1$. Solving equation (1.1) for $p_k(t)$, we obtain

$$p_k(t) = \frac{\exp\left(\beta_0^k + \beta_1^k t\right)}{1 + \sum_{k'=1}^{N} \exp\left(\beta_0^{k'} + \beta_1^{k'} t\right)} , \tag{1.2}$$

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

1.1.1 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 distributed[1] random variable;

together they are a set of N groups (number of years) of K probabilities $\{\{t_i, \mu_{ki}, \sigma_{ki} \mid k \in [0,K\text{-}1]\} \mid i \in [0,N\text{-}1]\}$. For each year the set of K probabilities form a distribution – their sum is equal to unity.

The regression consists of fitting a set of logistic functions $\{p_k(\mathbf{a}, \mathbf{b}, t) \mid k \in [0,K\text{-}1]\}$ 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 and 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 $p_k(\mathbf{a},\mathbf{b};t)$ are chosen to be ratios of sums of exponentials (This is equivalent to modeling the log probability ratios, p_k/p_0, as linear functions of time).

$$S(\mathbf{a}, \mathbf{b}) = \tfrac{1}{2} \sum_{k=0}^{k=K-1} \sum_{i=0}^{i=N-1} \frac{\left(p_k(\mathbf{a},\mathbf{b};t_i) - \mu_{ki}\right)^2}{\sigma_{ki}^2} \tag{1.3}$$

$$p_k(\mathbf{a},\mathbf{b},t) \equiv \frac{e^{A_k}}{1 + e^{A_1} + .. + e^{A_{K-1}}}$$
$$\mathbf{a} \equiv (a_0, a_1, .., a_{K-1}), \quad \mathbf{b} \equiv (b_0, b_1, .., b_{K-1}) \tag{1.4}$$
$$A_0 \equiv 0, \quad A_k \equiv a_k + b_k t$$

The parameters A_0, 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 – the remaining function being determined from the requirement that the complete set of K form a distribution and sum to unity.

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

The minimum of the function S is determined from the equations

$$\frac{\partial S}{\partial a_j} = \frac{\partial S}{\partial b_j} = 0 \qquad \text{for j=1,2,....,k-1} \tag{1.5}$$

noting the reations

$$\frac{\partial p_k}{\partial A_j} = \frac{\partial}{\partial A_j}\left(\frac{e^{A_k}}{1+e^{A_1}+..+e^{A_{K-1}}}\right) = p_k \delta_{kj} - p_k p_j$$

$$\frac{\partial}{\partial a_j} = \frac{\partial}{\partial A_j} \tag{1.6}$$

$$\frac{\partial}{\partial b_j} = t\frac{\partial}{\partial A_j}$$

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

1.1.2 Bayesian interpretation

The 2K-2 regression parameters {**a,b**} are regarded as random variables whose posterior distribution is proportional to the function exp(-S**(a,b)**). The maximum likelihood estimate of this probability distribution function, the minimum of the function S, is obtained at the values $\hat{\mathbf{a}}, \hat{\mathbf{b}}$. 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 $\left(\mathbf{a}-\hat{\mathbf{a}}\right),\left(\mathbf{b}-\hat{\mathbf{b}}\right)$ about the maximum likelihood estimate $\hat{\mathbf{S}} \equiv S\left(\hat{\mathbf{a}},\hat{\mathbf{b}}\right)$. Hence

$$S\left(\mathbf{a},\mathbf{b}\right) = \tfrac{1}{2}\sum_{k=0}^{k=K-1}\sum_{i=0}^{i=N-1}\frac{\left(p_k\left(\mathbf{a},\mathbf{b};t_i\right)-\mu_{ki}\right)^2}{\sigma_{ki}^2}$$

$$\equiv S\left(\hat{a},\hat{b}\right)+\tfrac{1}{2}\left(a-\hat{a},b-\hat{b}\right)P^{-1}\left(a-\hat{a},b-\hat{b}\right)+...$$

$$\approx S\left(\hat{a},\hat{b}\right)+\tfrac{1}{2}\sum_{i,j}(a_i-\hat{a}_i)\frac{\partial^2\hat{S}}{\partial\hat{a}_i\partial\hat{a}_j}\left(a_j-\hat{a}_j\right)+\tfrac{1}{2}\sum_{i,j}(a_i-\hat{a}_i)\frac{\partial^2\hat{S}}{\partial\hat{a}_i\partial\hat{b}_j}\left(b_j-\hat{b}_j\right)+ \tag{1.7}$$

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

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.

1.1.3 Estimation of the confidence intervals

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

$$
\begin{aligned}
p_k(\mathbf{a},\mathbf{b},t) &= p_k\left(\hat{\mathbf{a}} + \mathbf{a} - \hat{\mathbf{a}}, \hat{\mathbf{b}} + \mathbf{b} - \hat{\mathbf{b}}, t\right) \\
&= \hat{p}_k(t) + \left(\nabla_{\hat{a}}, \nabla_{\hat{b}}\right)\hat{p}_k(t)\begin{pmatrix} \mathbf{a} - \hat{\mathbf{a}} \\ \mathbf{b} - \hat{\mathbf{b}} \end{pmatrix} + \dots
\end{aligned}
\tag{1.8}
$$

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

$$
\sigma_k^2(t) \equiv \left\langle \left(p_k(\mathbf{a},\mathbf{b},t) - \hat{p}_k(t)\right)^2 \right\rangle = \left(\nabla_{\hat{a}}\hat{p}_k(t), \nabla_{\hat{b}}\hat{p}_k(t)\right)\left\langle \begin{pmatrix} \mathbf{a} - \hat{\mathbf{a}} \\ \mathbf{b} - \hat{\mathbf{b}} \end{pmatrix}\begin{pmatrix} \mathbf{a} - \hat{\mathbf{a}} \\ \mathbf{b} - \hat{\mathbf{b}} \end{pmatrix}^T \right\rangle \times
$$

$$
\left(\nabla_{\hat{a}}\hat{p}_k(t), \nabla_{\hat{b}}\hat{p}_k(t)\right)^T = \left(\nabla_{\hat{a}}\hat{p}_k(t), \nabla_{\hat{b}}\hat{p}_k(t)\right) P \left(\nabla_{\hat{a}}\hat{p}_k(t), \nabla_{\hat{b}}\hat{p}_k(t)\right)^T
\tag{1.9}
$$

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

$$
\sigma_k^2(t) = \left(\dfrac{\partial \hat{p}_k(t)}{\partial \hat{a}_1} \quad \dfrac{\partial \hat{p}_k(t)}{\partial \hat{a}_2} \quad \dfrac{\partial \hat{p}_k(t)}{\partial \hat{b}_1} \quad \dfrac{\partial \hat{p}_k(t)}{\partial \hat{b}_2}\right)
\begin{bmatrix}
P_{aa11} & P_{aa12} & P_{ab11} & P_{ab12} \\
P_{aa21} & P_{aa22} & P_{ab21} & P_{ab22} \\
P_{ba11} & P_{ba12} & P_{bb11} & P_{bb12} \\
P_{ba21} & P_{ba22} & P_{bb21} & P_{bb22}
\end{bmatrix}
\begin{pmatrix}
\dfrac{\partial \hat{p}_k(t)}{\partial \hat{a}_1} \\
\dfrac{\partial \hat{p}_k(t)}{\partial \hat{a}_2} \\
\dfrac{\partial \hat{p}_k(t)}{\partial \hat{b}_1} \\
\dfrac{\partial \hat{p}_k(t)}{\partial \hat{b}_2}
\end{pmatrix}
$$

$$\tag{1.10}$$

where $P_{cdij} \equiv \left\langle \left(c_i - \hat{c}_i\right)\left(d_j - \hat{d}_j\right)\right\rangle$. The 95% confidence interval for $p_k(t)$ is centred given as $\left[\hat{p}_k(t) - 1.96\sigma_k(t), \mathrm{p}_k(t) + 1.96\sigma_k(t)\right]$.

1.2 Module Two: Microsimulation

1.2.1 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 (QALYs), direct and indirect lifetime health-care costs, and, as a consequence of interventions, incremental cost effectiveness ratios (ICERs) over a specified time scale. Outputs can be discounted for any specific discount rate.
The section provides an overview of the initialization of the microsimulation model and will be expanded upon in the next sections.

1.2.2 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).

1.2.2.1 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 1. It shows the 2015 population distribution in Ukraine used in the initialization of the model.

males (20.8m)   Ukraine   females (24.1m)

(90+) 0.3%   (90+) 0.8%

(80-89) 2.2%   (80-89) 4.8%

(70-79) 5.8%   (70-79) 9.7%

(60-69) 10.5%   (60-69) 13.0%

(50-59) 14.2%   (50-59) 14.6%

(40-49) 14.2%   (40-49) 13.1%

(30-39) 17.1%   (30-39) 14.6%

(20-29) 14.9%   (20-29) 12.3%

(10-19) 9.9%   (10-19) 8.1%

(0-9) 9.9%   (0-9) 8.8%
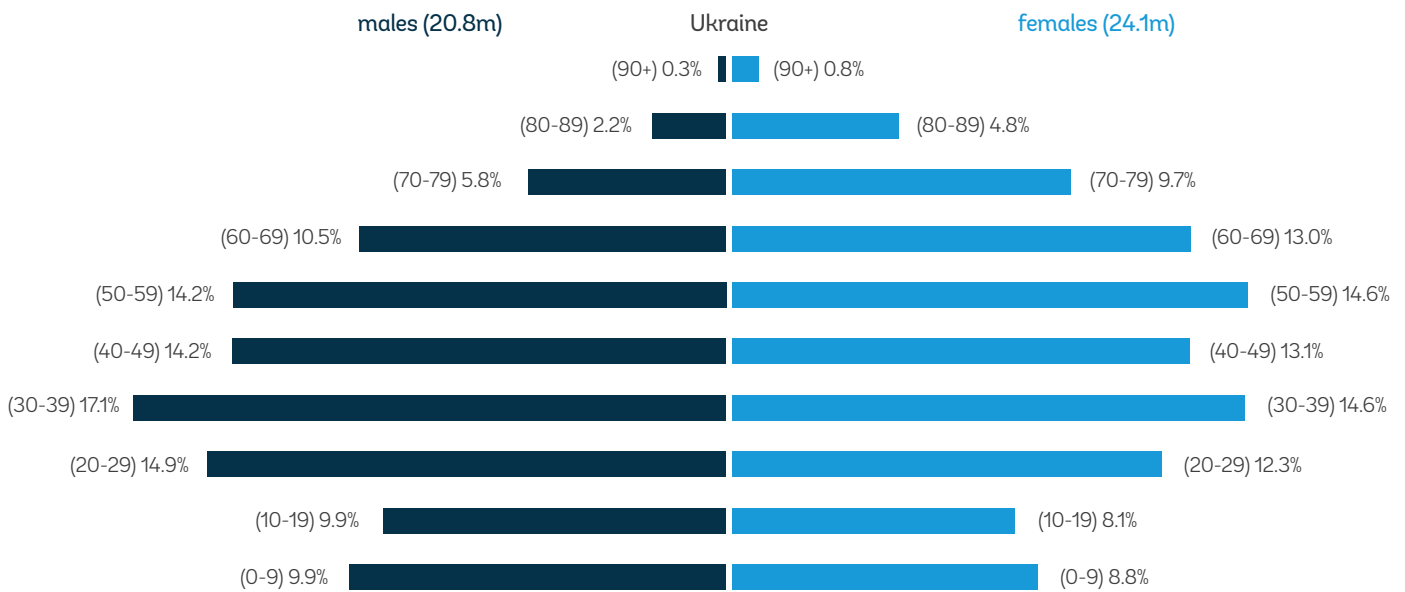
Figure 1: **Population Pyramid in 2015 in Ukraine**

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 1).

Table 1: **Summary of the Parameters Representing the Distribution Component**

Distribution name	symbol	note
MalesByAgeByYear	$p_m(a)$	Input in $year_0$ – probability of a male having age a
FemalesByAgeByYear	$p_f(a)$	Input in $year_0$ – probability of a female having age a
BirthsByAgeofMother	$p_b(a)$	Input in $year_0$ – conditional probability of a birth at age a| the mother gives birth.
NumberOfBirths	$p_\lambda(n)$	$\lambda \equiv$ TFR, Poisson distribution, probability of giving birth to n children

1.2.2.2 Birth model

Any female in the childbearing years {AgeAtChild.lo, 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.[2]

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 probability that a mother who produces N children produces n of them at age a is given as the Binomially distributed variable,

$$p_b\left(\text{n at a} \mid N\right) = \frac{N!}{n!(N-n)!}\left(p_b\left(a\right)\right)^n\left(1 - p_{bm}\left(a\right)\right)^{N-n} \tag{1.11}$$

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

$$p_b\left(n\,at\,a\right) = e^{-\lambda}\sum_{N=n}^{\infty}\frac{\lambda^N}{N!}\,p_b\left(\text{n at a} \mid N\right) = e^{-\lambda}\sum_{N=n}^{\infty}\frac{\lambda^N}{n!(N-n)!}\left(p_b\left(a\right)\right)^n\left(1 - p_b\left(a\right)\right)^{N-n} \tag{1.12}$$

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

$$p_b\left(n\,at\,a\right) = e^{-\lambda p_b(a)}\frac{\left(\lambda p_b\left(a\right)\right)^n}{n!} = p_{\lambda p_b(a)}\left(n\right) \tag{1.13}$$

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

The gender of the children[3] is determined by the probability p_{male}=1-p_{female}
In the baseline model this is taken to be the probability $N_m/(N_m+N_f)$.

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.

1.2.2.3 Deaths from modeled 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 modeled diseases will change as modeled 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.

[2] This could be made to be time dependent; in the baseline model it is constant.

[3] The probability of child gender can be made time dependent.

1.2.3 The risk factor model

The distribution of risk factors (RF) in the population is estimated using regression analysis stratified by both sex S = {male, 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 $D = \{p_k(t)|k = 1, \ldots N; \, t > 0\}$, 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).

1.2.3.1 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_0, p_1, p_2}. 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).

The complete set of longitudinal smoking trajectories and the probabilities of their happening is generated for the simulation years by allowing all possible transitions between smoking categories:

{never smoked} — {never smoked, smoker}
{ex-smoker} — {ex-smoker, smoker}
{smoker} — {ex-smoker, smoker}

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

$$\begin{bmatrix} p_0^{'} \\ p_1^{'} \\ p_2^{'} \end{bmatrix} = \begin{bmatrix} 1-p & 0 & 0 \\ 0 & 1-p & 1-p \\ p & p & p \end{bmatrix} \begin{bmatrix} p_0 \\ p_1 \\ p_2 \end{bmatrix} \tag{1.14}$$

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 will consist in summing over the probability 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 x 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 evolution of the three-dimensional states are the smoking trajectories necessary for the computation of disease-table disease and death probabilities.

1.2.3.2 Smoking

The microsimulation framework 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: i) Chronic obstructive pulmonary disease (COPD), ii) Coronary heart disease (or Myocardial Infarction if CHD data are not available), iii) stroke, and iv) 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 associated with ex-smokers ($RR_{ex-smoker}$) are related to the relative risk of smokers (RR_{smoker}). The ex-smoker relative risks are assumed to decrease over time with the number of years since smoking cessation (Tcessation). These relative risks are computed in the model using equations 1.19 and 1.20 (1).

$$RR_{ex-smoker}(A, S, T_{cessation}) = 1 + (RR_{smoker}(A, S) - 1)\exp(-\gamma(A)T_{cessation}) \qquad (1.15)$$

$$\gamma(A) = \gamma_0 \exp(-\eta A) \qquad (1.16)$$

where γ is the regression coefficient of time dependency. The constants γ0 and η are intercept and regression coefficient of age dependency, respectively, which are related to the specified disease Table 2.

Table 2: **Parameter Estimates for γ0 and η Related to Each Disease (1)**

Distribution name	γ0	η
AMI	0.24228	0.05822
Stroke	0.31947	0.01648
COPD	0.20333	0.03087
Lung cancer	0.15637	0.02065

However, a minimum exists when the cessation time is equal to η-1. The minimum value was calculated by the method detailed below (equations (1.17), (1.18) and (1.19)). Where time t is equal to the age A of an individual.

$$\rho_{Exsmk}(t) = 1 + (\rho_{smk} - 1) f(t) \tag{1.17}$$

$$f(t) = \exp\left(-\gamma_0 (t - t_0) \exp(-\eta t)\right)$$
$$\Rightarrow \tag{1.18}$$
$$f'(t) = -\gamma_0 f(t) e^{-\eta t} \left(-\eta(t - t_0) + 1\right)$$

The function f(t) has the following properties:

$$\begin{aligned} f(t_0) &= 1 \\ f'(t_0) &= -\gamma_0 e^{-\eta t_0} \\ f(t) &\text{ has a minimum at } t = t_0 + \eta^{-1} \\ f(\infty) &= A \end{aligned} \tag{1.19}$$

In order to keep the RR$_{ex\text{-}smoker}$ from increasing, the cessation time was set equal to η-1 when the cessation time was greater than η-1 (see equation (1.20)).

$$RR_{ex\text{-}smoker}(A, S, T_{cessation}) = \begin{cases} 1 + (RR_{smoker}(A,S) - 1)\exp(-\gamma(A)T_{cessation}) & T_{cessation} < \eta^{-1} \\ 1 + (RR_{smoker}(A,S) - 1)\exp(-\gamma(A)\eta^{-1}) & T_{cessation} \geq \eta^{-1} \end{cases}$$
$$(1.20)$$

$$\gamma(A) = \gamma_0 \exp(-\eta A) \tag{1.21}$$
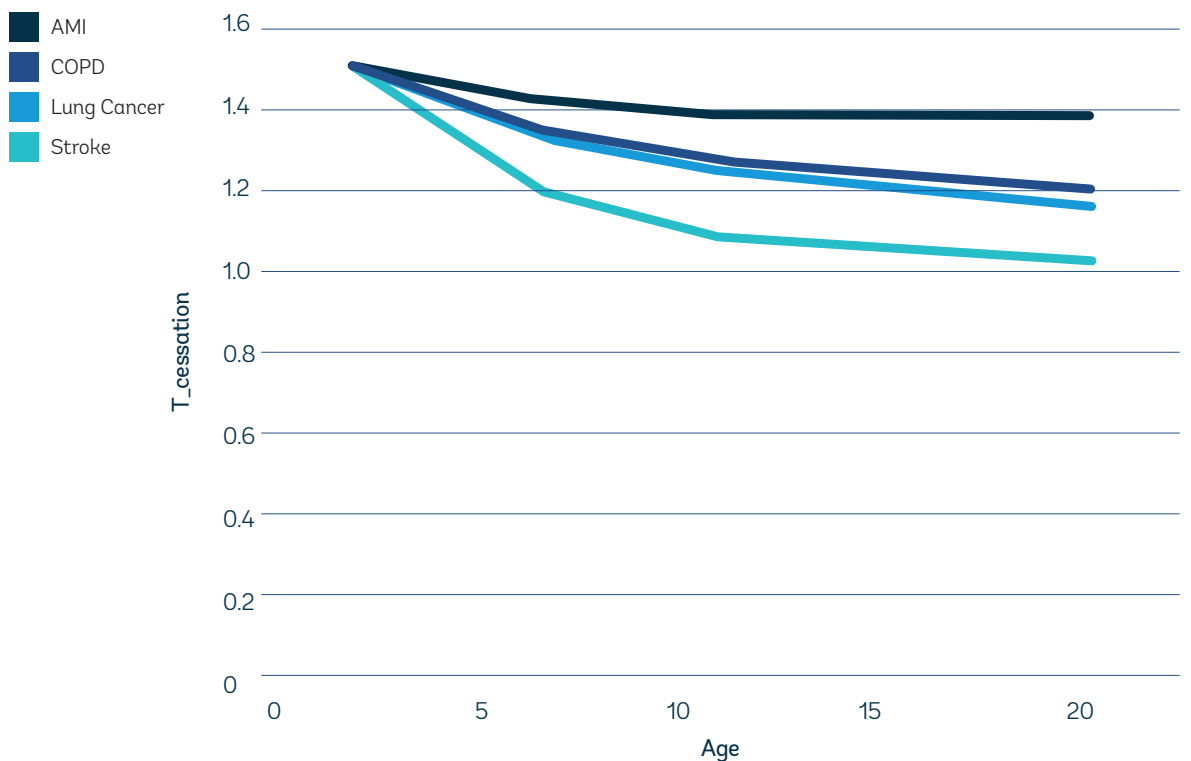
1.2.4 Relative risks

The reported incidence risks for any disease do not make reference to any underlying risk factor. The microsimulation requires this dependence to be made manifest.

The risk factor dependence of disease incidence has to be inferred from the distribution of the risk factor in the population (here denoted as π); it is a disaggregation process:

Suppose that α is a risk factor state of some risk factor A, and denote by $p_A(d|\alpha,a,s)$ the incidence probability for the disease d given the risk state, α, the person's age, a, and gender, s. The relative risk ρ_A is defined by equation (1.22).

$$p_A\left(d\,|\,\alpha,a,s\right) = \rho_{A|d}\left(\alpha\,|\,a,s\right)p_A\left(d\,|\,\alpha_0,a,s\right)$$
$$\rho_{A|d}\left(\alpha_0\,|\,a,s\right) \equiv 1$$

(1.22)

Where α_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,

$$p\left(d\,|\,a,s\right) = \sum_{\alpha}p_A\left(d\,|\,\alpha,a,s\right)\pi_A\left(\alpha\,|\,a,s\right)$$
$$= p_A\left(d\,|\,\alpha_0,a,s\right)\sum_{\alpha}\rho_{A|d}\left(\alpha\,|\,a,s\right)\pi_A\left(\alpha\,|\,a,s\right)$$

(1.23)

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

$$p\left(d\,|\,\alpha,a,s\right) = \rho_{A|d}\left(\alpha\,|\,a,s\right)\frac{p\left(d\,|\,a,s\right)}{\sum_{\beta}\rho_{A|d}\left(\beta\,|\,a,s\right)\pi_A\left(\alpha\,|\,a,s\right)}$$

(1.24)

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

1.2.5 Modeling diseases

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

The microsimulation uses risk-dependent incidence statistics, and these are inferred from the relative risk statistics and the distribution of the risk factor within the population. 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 modeled as age- and gender-dependent exponential distributions.

1.2.6 Methods for approximating missing disease statistics

A large amount data are required for modeling these diseases. Where possible, these datasets have been collected from published sources or analyzed from either cross-sectional or longitudinal datasets. Another limitation is that often these data need to be in a specific format. For example, the model updates the individual's disease status every year, so the RR's used in the model need to be annual RR's.

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

1.2.6.1 Terminal and non-terminal single-state disease incidence from prevalence

For terminal diseases, to estimate incidence knowing prevalence and mortality, one can proceed by finding those incidence probabilities that minimize the distance between the known $\bar{p}_{pre}^{K}$ and computed prevalence p_{pre}^{K}

$$S = \sum_{a \in \text{AgeGroup}} \frac{\left(\bar{p}_{pre}^{K}(a) - p_{pre}^{K}(a) \right)^2}{\bar{\sigma}^2(a)}$$

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

1.2.6.2 Mortality statistics

In any year, in some population, in a sample of N people who have the disease, a subset N_{ω} will die from the disease.

Mortality statistics record the cross-sectional probabilities of death as a result of the disease – possibly stratifying by age

$$p_{\omega} = \frac{N_{\omega}}{N} \tag{1.25}$$

Within some such subset N_ω 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 $\{p_{\omega 0}, p_{\omega 1}, p_{\omega 2}, p_{\omega 3}, p_{\omega 4}\}$

The probability of being alive after N years is simply that you don't die in each year

$$p_{survive}(a_0 + N) = (1 - p_{\omega 0})(1 - p_{\omega 1})(1 - p_{\omega 2})..(1 - p_{\omega N-1}) \tag{1.26}$$

1.2.6.3 The survival models
There are three in use (they are easily extended if the data merit):

Survival model 0: a single probability of dying $\{p_{\omega 0}\}$

$p_{\omega 0}$ is valid for all years

Survival model 1: two different probabilities of dying $\{p_{\omega 0}, p_{\omega 1}\}$

$p_{\omega 0}$ is valid for the first year; $p_{\omega 1}$ thereafter.

Survival model 2: three different probabilities of dying $\{p_{\omega 0}, p_{\omega 1}, p_{\omega 5}\}$

$p_{\omega 0}$ is valid for the first year; $p_{\omega 1}$ for the second to the fifth year; $p_{\omega 5}$ thereafter

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

1.2.6.4 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

$$
\begin{aligned}
\hat{p}_{mortality}(a) = {} & p_{inc}(a-1)\,p_{\omega 0} + \\
& + p_{inc}(a-2)(1-p_{\omega 0})\,p_{\omega 1} + \\
& + p_{inc}(a-3)(1-p_{\omega 0})(1-p_{\omega 1})\,p_{\omega 1} + \\
& + p_{inc}(a-4)(1-p_{\omega 0})(1-p_{\omega 1})^2\,p_{\omega 1} + \\
& + p_{inc}(a-5)(1-p_{\omega 0})(1-p_{\omega 1})^3\,p_{\omega 1} + \\
& + p_{inc}(a-6)(1-p_{\omega 0})(1-p_{\omega 1})^4\,p_{\omega 5} + \\
& + p_{inc}(a-7)(1-p_{\omega 0})(1-p_{\omega 1})^4(1-p_{\omega 5})\,p_{\omega 5} + \\
& + ...
\end{aligned}
\tag{1.27}
$$

The three probabilities $\{p_{\omega 0}, p_{\omega 1}, p_{\omega 5}\}$ are estimated by minimizing

$$S = \sum_{a \in AgeGroup} \frac{\left(\bar{p}_{mortality}(a) - \hat{p}_{mortality}(a)\right)^2}{\bar{\sigma}^2} \qquad (1.28)$$

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

$$p_{inc}(a) = (1 - p_i(0))(1 - p_i(1))..(1 - p_i(a-1))p_i(a) \qquad (1.29)$$

1.2.6.4.1 Survival Statistics CRUK 2010/11

The following table (Table 3) is taken from the CRUK website (www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/lung-cancer/survival) as an example. It gives 1-, 5- and 10-year survival percentages for lung cancer:

Table 3: **Survival Percentage for Lung Cancer**

Cancer	1 year	5 year	10 year	$1 - p_{\omega 0}$	$1 - p_{\omega 1}$	$1 - p_{\omega 5}$
Lung	32	10	5	0.32	0.75	0.71

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

$$p_{survival}(a_0 + 1) = (1 - p_{\omega 0})$$
$$p_{survival}(a_0 + 5) = (1 - p_{\omega 0})(1 - p_{\omega 1})^4 \qquad (1.30)$$
$$p_{survival}(a_0 + 10) = (1 - p_{\omega 0})(1 - p_{\omega 1})^4 (1 - p_{\omega 5})^5$$

1.2.6.5 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_0 years from some time t0 is given as

$$p_{survival}(t) = 1 - R^{-1} \int_0^t du\, e^{-Ru} = e^{-Rt} \qquad (1.31)$$

For a time period of 1 year

$$p_{survival}(1) = e^{-R}$$
$$\Rightarrow \qquad (1.32)$$
$$R = -\ln\left(p_{survival}(1)\right) = -\ln(1 - p_\omega)$$

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

$$p_{\text{survival}}(t=4) = 1 - R^{-1} \int_0^4 du\, e^{-Ru} = e^{-4R} = (1-p_\omega)^4 \qquad (1.33)$$

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

1.2.6.6 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.

1.2.6.6.1 Survival model 0

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

The model uses 1 parameter {R}

$$R = -\ln\left(p_{survival}(1)\right) \qquad (1.34)$$

1.2.6.6.2 Survival model 1

The model uses two parameters {p_1, R}

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

$$p_1 = 1 - p_{survival}(1)$$
$$R = -\frac{1}{4}\ln\left(\frac{p_{survival}(5)}{p_{survival}(1)}\right) \qquad (1.35)$$

1.2.6.6.3 Survival model 2

The model uses three parameters {p_1, R, $R_{>5}$}

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

$$p_1 = 1 - p_{survival}(1)$$
$$R = -\frac{1}{4}\ln\left(\frac{p_{survival}(5)}{p_{survival}(1)}\right) \qquad (1.36)$$
$$R_{>5} = -\frac{1}{5}\ln\left(\frac{p_{survival}(10)}{p_{survival}(5)}\right)$$

1.2.6.7 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 4 following states:

state	Description
0	alive without disease *d*
1	alive with disease *d*
2	dead from disease *d*
3	dead from another disease

p_{ik} is the probability of disease d incidence, aged *k*

$p_{\omega k}$ is the probability of dying from the disease *d*, aged *k*

$p_{\overline{\omega} k}$ is the probability of dying other than from disease *d*, aged *k*

The state transition matrix is constructed as follows

$$
\begin{bmatrix} p_0(k+1) \\ p_1(k+1) \\ p_2(k+1) \\ p_3(k+1) \end{bmatrix} = \begin{bmatrix} (1-p_{\overline{\omega}k})(1-p_{ik}) & (1-p_{\overline{\omega}k}-p_{\omega k})p_{\alpha k} & 0 & 0 \\ (1-p_{\overline{\omega}k})p_{ik} & (1-p_{\overline{\omega}k}-p_{\omega k})(1-p_{\alpha k}) & 0 & 0 \\ 0 & p_{\omega k} & 1 & 0 \\ p_{\overline{\omega}k} & p_{\overline{\omega}k} & 0 & 1 \end{bmatrix} \begin{bmatrix} p_0(k) \\ p_1(k) \\ p_2(k) \\ p_3(k) \end{bmatrix} \tag{1.37}
$$

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

The disease mortality equation is that for state-2,

$$
p_2(k+1) = p_{\omega k}p_1(k) + p_2(k) \tag{1.38}
$$

The probability of dying from the disease in the age interval [k, k+1] is $p_{\omega k}p_1(k)$ - this is otherwise the (cross-sectional) disease mortality, $p_{mor}(k)$. $p_1(k)$ is otherwise known as the disease prevalence, $p_{pre}(k)$. Hence the relation

$$
p_{\omega k} = \frac{p_{mor}(k)}{p_{pre}(k)} \tag{1.39}
$$

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

$$p_{\omega k} = 1 - e^{-R_k} \quad \Rightarrow \quad R_k = -\ln\left(1 - p_{\omega k}\right) \tag{1.40}$$

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 any longer of any use, the microsimulation uses survival statistics inferred from the prevalence and mortality statistics (equation (1.39)).

An alternative derivation equation (1.39) 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 $p_{\omega k} n_k$ and, equivalently, as $p_{mor}(k) N_k$. Observing that the disease prevalence is n_k/N_k leads to the equation

$$p_{\Omega k} n_k = p_{mor}\left(k\right) N_k$$
$$p_{pre}\left(k\right) = \frac{n_k}{N_k}$$
$$\Rightarrow \tag{1.41}$$
$$p_{\Omega k} = \frac{p_{mor}\left(k\right)}{p_{pre}\left(k\right)}$$

1.2.6.8 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 some earlier year.

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

$$p_{mor}\left(a\right) = \sum_{a_0 < a} p_{\omega}\left(a | a_0\right) p_{inc}\left(a_0\right) \tag{1.42}$$

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 (1.42) 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.

1.2.6.8.1 Setup

For each sex, consider an N-stage, terminal disease for which both the inter-stage transition probabilities, $p_{i,j}(a)$ (the probability to go from stage i to stage j), and the state-dependent mortality probabilities $p_{mor}^{K}(a)$ (K denotes the stage number and a the age) are known. The following algorithm allows for an 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.

1.2.6.8.2 Definitions

$p_K(a \mid a_0, K_0)$ 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_0 in state K_0

$p_\omega(a \mid a_0, K_0)$ the probability of being dead (from the disease) at age a, given that the disease was contracted at a_0 in state K_0

$p_{inc}(a_0, K_0)$ the probability of first getting the disease in a_0 in state K_0, given no disease at age 0

$p_\omega^K(a \mid a_0)$ the probability of dying from the disease in stage K at age a, given that the disease was contracted at age a_0 and that the person was alive at age a-1

These probabilities are linked by the state-update equation,

1.2.6.8.3 Disease incidence

The probability that a person, who at age 0 does not have the disease, first gets the disease at age a_0 in state K_0 given as

$$p_{inc}(a_0, K_0) = \prod_{a=0}^{a=a_0-1}\left(1 - \sum_{k=1}^{K} p_{0k}(a)\right)p_{0K_0} \tag{1.43}$$

1.2.6.8.4 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_0 in stage K, then the initial state vector is determined from the initial conditions $p_i(a_0) = \delta_{iK_0}$ $(K_0 > 0)$, $p_\omega(a_0) = 0$. At subsequent ages, the state probabilities are given by the recursion equation

$$\begin{pmatrix} p_0\left(a+1\,|\,a_0,K_0\right) \\ p_1\left(a+1\,|\,a_0,K_0\right) \\ .. \\ p_{N-1}\left(a+1\,|\,a_0,K_0\right) \\ p_\omega\left(a+1\,|\,a_0,K_0\right) \end{pmatrix} = T\left(a,a_0\right)\begin{pmatrix} p_0\left(a\,|\,a_0,K_0\right) \\ p_1\left(a\,|\,a_0,K_0\right) \\ ... \\ p_{N-1}\left(a\,|\,a_0,K_0\right) \\ p_\omega\left(a\right)|\,a_0,K_0 \end{pmatrix}$$

$$T\left(a,a_0\right) \equiv \begin{pmatrix} \left(1-\sum_{k>0} p_{k0}\right) & 0 & ... & 0 & 0 \\ p_{0,1} & \left(1-\sum_{k>1} p_{1,k}\right)\left(1-p_\omega^1\left(a\,|\,a_0\right)\right) & ... & 0 & 0 \\ ... & ... & ... & ... & ... \\ p_{0,N-1} & p_{1,N-1}\left(1-p_\omega^1\left(a\,|\,a_0\right)\right) & ... & \left(1-p_\omega^{N-1}\left(a\,|\,a_0\right)\right) & 0 \\ 0 & p_\omega^1\left(a\,|\,a_0\right) & ... & p_\omega^{N-1}\left(a\,|\,a_0\right) & 1 \end{pmatrix}$$

$$(1.44)$$

Where, for survival model 2,

$$p_\omega^K\left(a\,|\,a_0\right) = \begin{cases} p_{\omega 0}^K & \left(a=a_0\right) \\ p_{\omega 1}^K & \left(a_0 < a \le a_0+4\right) \\ p_{\omega 5}^K & \left(a_0+4 < a\right) \end{cases} \tag{1.45}$$

1.2.6.8.5 Disease mortality

The probability of dying from the disease in stage K at age a, denoted $p_{mor}^K\left(a\right)$, is given by the equation

$$p_{mor}^K(a) = \sum_{K_0=1}^{K_0=N-1}\sum_{a_0=0}^{a_0=a-1} p_\omega^K\left(a\,|\,a_0,K_0\right)p_K\left(a-1\,|\,a_0\right)p_{inc}\left(a_0,K_0\right) \tag{1.46}$$

1.2.6.8.6 Estimating survival

When the disease mortality is known, here denoted $\bar{p}_{mor}^K\left(a\right)$, the sets of survival parameters (3 for each state, and possibly stratified by age) can be estimated by minimizing

$$S = \sum_{\text{stage K}}\left(\sum_{a\in\text{AgeGroup}} \frac{\left(\bar{p}_{mor}^K\left(a\right)-p_{mor}^K\left(a\right)\right)^2}{\bar{\sigma}^2\left(a\right)}\right) \tag{1.47}$$

$$(1.48)$$

1.2.7 Approximating attributable cases

The smoking attribute cases (I_A) 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.

$$I_{A_d}(y) = \frac{n(y)}{N(y)} \tag{1.1}$$

1.2.8 Potential Years of Life Lost (PYLL) (2, 3)

The PYLL for an individual (PYLL(i)) who dies in a given year will be calculated from the following equation (1.2).

$$PYLL(i) = \begin{cases} Age_{ref} - Age_{death} & \text{if } Age_{death} < Age_{ref} \\ 0 & \text{if } Age_{death} \geq Age_{ref} \end{cases} \tag{1.2}$$

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.

$$TotalPYLL(year) = \frac{\sum_{i=1}^{N_{died}(year)} PYLL(i)}{N_{population}(year)} \tag{1.3}$$

1.2.9 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

$$premature(i) = \begin{cases} 1 & \text{if } age_{death} < 70 \\ 0 & \text{if } age_{death} \geq 70 \end{cases} \tag{1.4}$$

$$PM(year) = \frac{\sum_{i=1}^{N_{died}(year)} premature(i)}{N_{population}(year)} \tag{1.5}$$

1.2.10 Costs module

The cost module includes both direct and indirect cost calculations.

1.2.10.1 Direct costs

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

$$\text{Direct cost (£) per individual (year)} = \frac{\text{Cost per case (\$) * Prevalence(year)}}{\text{Alive (year)}} \qquad (1.6)$$

The direct costs are displayed per million $ and at a rate defined by the user.

$$\text{Direct costs (M£) per rate (year)} = \frac{\text{Direct cost (\$) per individual (year) * rate}}{10^6} \qquad (1.7)$$

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

$$95\%CI = \text{Direct Costs (year)} * 1.96 \sqrt{\frac{P(1-P)}{Trials}} \qquad (1.8)$$

1.2.11 Premature mortality costs (PMC)(4, 5)

$$PMC(i) = \begin{cases} \displaystyle\sum_{i=age_{death}}^{i=\text{LE at birth-1}} Income(i) & \text{if } age_{death} < \text{LE at birth} \\ 0 & \text{if } age_{death} \geq \text{LE at birth} \end{cases} \qquad (1.9)$$

The premature mortality costs 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 the equation below (1.10).

$$\text{PMC per } 100,000 = \frac{PMC_{Total}}{N_{population}} * 100,000 \qquad (1.10)$$

1.2.12 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 (ET) were calculated using the propagation of errors equation:

$$E_T = \sqrt{E_1^2 + E_2^2 + \quad + E_n^2} \qquad (1.11)$$

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

2 Software Architecture

2.1 Aim of the Model

This model utilizes a common bespoke method to predict the impact of changing risk factors to measure chronic disease. UKHF currently have risk-factor models for obesity, tobacco, albuminuria, and eGFR and related diseases that include myocardial infarction/coronary heart disease, stroke, Type 2 diabetes, chronic obstructive pulmonary disease, chronic kidney disease, and lung cancer.

The model is an epidemiological/medical competing-risk application that uses both stochastic and deterministic processing capable of projecting cohort mortality rates of an individual or a population, taking account of the individuals' risk factors and medical profiles. 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.

2.2 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 under the Windows operating system.

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 sketched below are examples of this process). Files input 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, medium or low-level data files which can be further processed by standard
software packages.
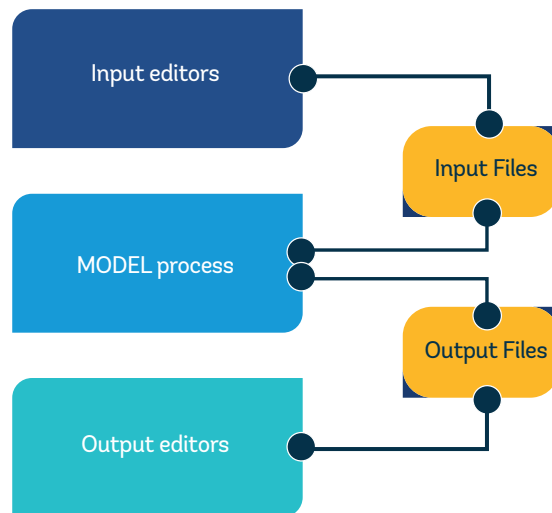
Figure 2.2: **The Model Structure.**

2.3 *Main C++ Classes Used by the Model*

Individual members of the population and diseases are modeled by the C++
classes Tperson and Tdisease respectively. The risk-factor trends and scenarios
are modeled 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; the idea is 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 catch diseases from which they may recover or die.

Figure 3 shows a schematic of the model illustrating the overall structure of the
model and each class.
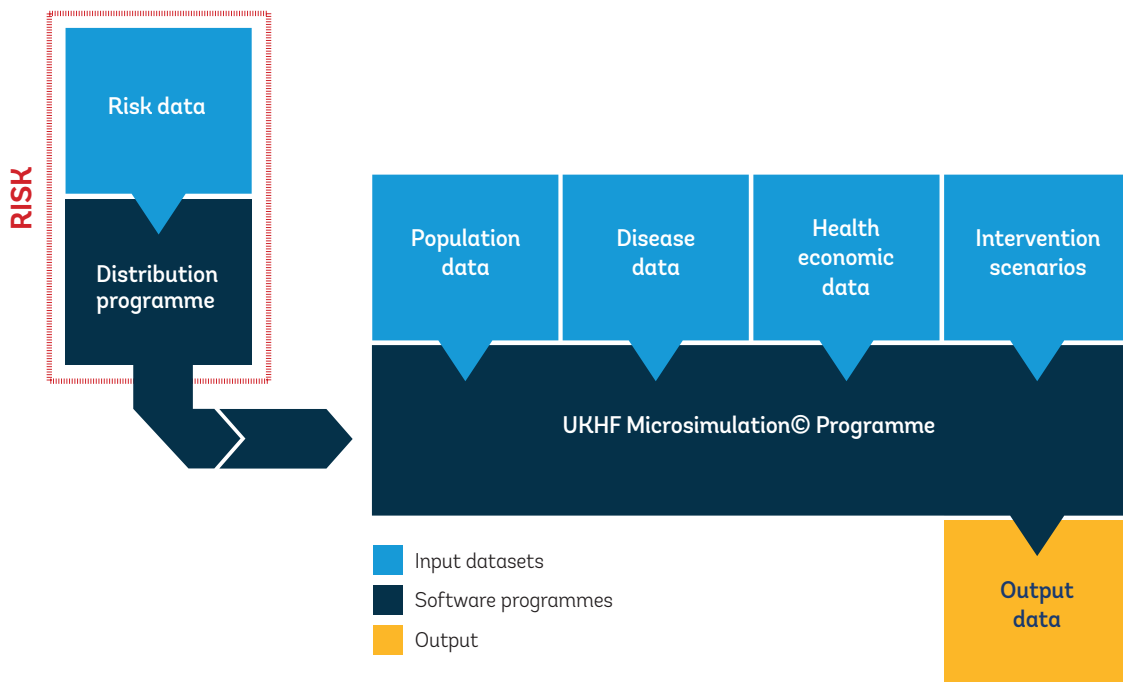
Figure 3: **Schematic of the Model.**

2.3.1 Tperson C++ class

People are implemented as instances of the C++ class, Tperson; an indication of its data fields and methods provided in Table 4 are grouped into the state record. The medHistory record maintains their current disease and risk factor status together with data necessary for the computation of 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 modeled and the list of risk-factor scenario object pointers, which determines how the person's set of risk factors change over the year.

Table 4: **The C++ Tperson Class**

Tperson	Description	
Data field		
	State	State vector record
	medHistory	Medical history record
	…	…
Method		
	Tperson(state0,medHistory0)	Constructor for initial state and history
	yearByYear(…)	Updates state and medHistory by one year
	…	…

2.3.2 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 4. 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, some examples of which are
provided in Table 5. 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() . For a specified person 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
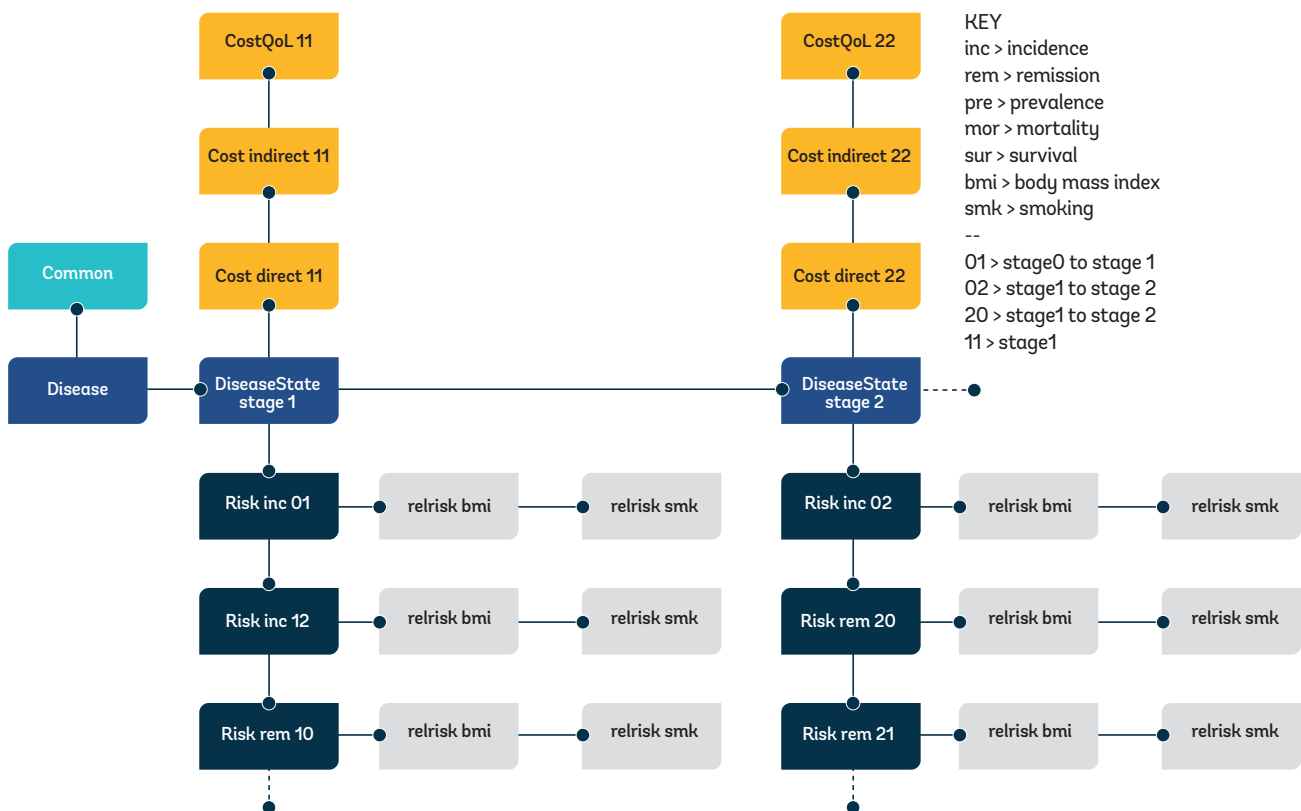and in possible exclusive disease states or dead.

Figure 4: **Multistage Disease Architecture**

Table 5: **The C++ Tdisease Class**

TDisease	Description	
Data field	name	Disease name
	terminal	Boolean, true if the disease is terminal
	state	Disease state {normal, severe,…}
	*DataAvailability	Boolean array of Data availability by risk type
	**IncidenceRisk	Incidence rates by age, gender
	***SurvivalRisk	Survival rates by age, gender, state
	**PrevalenceRisk	Prevalence rates by age, gender
	***RemissionRisk	Remission rates by age and gender,state
	***MortalityRisk	Mortality rates by age, gender, state
	***RelRisk	Relative risks by risk factor type, age, gender
	…	…
Method		
	TDisease(aFile)	Constructor using data from aFile
	LoadFromFile(aFile)	Fills the data fields from aFile
	WriteToFile(aFile)	Writes the data fields to aFile
	GetRisk(state, medHistory,risktype)	Returns risk for specified risktype
	…	…

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 input 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 are 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.

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 modeled diseases which 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 modeled as a single, instantly fatal, terminal disease; its incidence probability is constructed via the disease editor from the modeled 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 modeled. These scenarios can simply maintain risk-factor distributions at their start year values, or they can allow for the modeling of risk factor trends or medical advances resulting in the reduction of disease incidence or improvements in the survivability of specified diseases.

2.3.3 Tscenario C++ class

Scenarios are modeled as instances of the C++ Tscenario class and are constructed by the scenario editor, which 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 is provided in Table 6.

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 input editors – 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 ease of its maintainability.

Table 6: **The C++ Tscenario Class**

Tscenario		
Data field		
	scenarioType	Type of scenario eg. {,smoking,… }
	start year	Year at which scenario starts
	stop year	Year at which scenario stops
	futureRiskFile	File specifying future risk distribution
	targetAgeGroup	Target age group eg. {18+}
	targetGenderGroup	Target gender group eg. {males,females}
	…	…
Method		
	Tscenario(aFile)	Constructor using data from aFile
	LoadFromFile(aFile)	Fills data fields from aFile
	…	…

References

1. Hoogenveen RT, van Baal PH, Boshuizen HC, Feenstra TL. Dynamic effects of smoking cessation on disease incidence, mortality and quality of life: The role of time since cessation. *Cost Eff Resour Alloc.* 2008.

2. Gardner JW, Sanborn JS. Years of Potential Life Lost (YPLL)-What Does it Measure? Epidemiology. 1990;1(4):322-9.

3. Health and Social Care Information Centre. Indicator Specification: CCG OIS 1.1 Potential Years of Life lost (PYLL) from causes considered amenable to healthcare 2015. Available at: https://indicators.hscic.gov.uk/download/Clinical%20Commissioning%20Group%20Indicators/Specification/CCG_1.1_I00767_S.pdf.

4. Gold M, Siegel JE, Russell LB, Weinstein MC. Cost-effectiveness in Health and Medicine. Press OU, editor1996.

5. Menzin J, Marton JP, Menzin JA, Willke RJ, Woodward RM, Federico V. Lost productivity due to premature mortality in developed and emerging countries: An application to smoking cessation. *BMC medical research methodology.* 2012;12(1):1.

Appendix 2. Results of the TaXSiM model: Ukraine Summary Cigarette Tax - Scenarios Output – 2015 – 2017

Government Revenue Type	Actual 2015	Expected Contribution to GDP	Baseline Situation (2016): Ad valorem (12%) minimum specific (8.515 UAH) and simple specific (6.365 UAH)	Expected Contribution to GDP	SCENARIO 1 (2017): Ad valorem remains equal and 12% tax increase in minimum specific excise (9.54 UAH), and simple specific (7.13 UAH)
Total cigarettes taxed (billion pieces)	73.8		66.9		64.0
Average cigarette price (UAH per pack)	15.2		19.2		21.2
Average cigarette price (US$ per pack) *	$0.63		$0.81		$0.87
Average excise tax (UAH per 1000 pieces)	308.9		430.7		482.6
Total excise tax revenue (billion UAH)	22.8	1.0%	28.8	1.3%	30.9
Total excise tax revenue (US$ billion) *	$0.94		$1.21		$ 1.27
Additional tobacco excise (billion UAH) /percentage of GDP			6.0	0.3%	2.1
Additional tobacco excise (U$ million) *			$ 254		$85
Total government revenue (excise, VAT and levies, billion UAH)	34.9	1.6%	42.1	1.8%	45.0
Total government revenue (excise, VAT and levies, US$ billion) *	$1.4		$1.8		$1.8
Total expenditure on cigarettes (billion UAH)	56.3		64.3		67.9
Percentage change in total cigarette consumption (%)			-9.3		-4.3

* World Bank Group forecast:
Annual average exchange rate = 2016 (1US$/23.8 UAH); 2017 (1US$/24.4 UAH)

Expected Contribution to GDP	SCENARIO 2 (2017): Increase Ad valorem tax (15%), and 30% Increase in the minimum specific excise (11.08 UAH), and simple specific (8.28 UAH)	Expected Contribution to GDP	SCENARIO 3 (2017): Increase 30% the Ad valorem, and 50% Increase in the minimum specific excise (12.77 UAH), and simple specific (9.55 UAH)	Expected Contribution to GDP	SCENARIO 4 (2017): Increase Ad valorem and specific tax (40%), adopting a simplified tax structure with uniform specific excise tax (11.92 UAH)	Expected Contribution to GDP
	60.1		53.4		48.8	
	24.7		32.9		41.4	
	$1.01		$ 1.35		$1.69	
	573.0		825.8		1106.1	
1.2%	34.4	1.3%	44.1	1.7%	54.0	2.1%
	$1.41		$ 1.81		$ 2.21	
0.1%	5.6	0.2%	15.3	0.6%	25.1	1.0%
	$230		$ 626		$1,030	
1.8%	49.9	1.9%	61.9	2.4%	73.9	2.9%
	$2.0		$ 2.5		$ 3.0	
	74.2		88.0		100.9	
	-10.2		-20.2		-27.1	

Appendix 3. Adjustment of Epidemiological Input Data for the Microsimulation Model of the Health Impacts of Tobacco Taxation in Ukraine

The goal of this annex is to summarize research findings relevant to the microsimulation model, including data on the incidence of tobacco-use-related diseases, relative risks of their development in smokers and former smokers compared with never smokers, and related risk of premature death.

Below, diseases included in the model are considered and, for each of them, extracts are shown from studies related to incidence, relative risks, and mortality. Based on the published evidence and statistics available for Ukraine, estimates of incidence and relative risks were elaborated and suggested as inputs for the microsimulation model.

As estimates from the Global Burden of Disease database became available during our research (Global Burden of Disease, 2016), we utilized incidence inputs from this database. Estimates for relative risks were used, as explained below.

3.1 Cardiovascular Diseases, Particularly Coronary Heart Disease (CHD)

3.1.1 Incidence
Because there are no proper estimates of CHD incidence in Ukraine, and most studies report only relative risks and no absolute risks, data were adapted from the seminal study on ischemic heart disease – the Framingham study (Castelli, 1984; Lerner & Kannel, 1986).

The Framingham study reports average morbidity and mortality for men and women. It was calculated that the number of CHD cases is about five times greater than the number of CHD deaths. SDR for CHD in Ukraine was taken for the years 2012-2014 and multiplied by the coefficient to estimate the approximate level of average incidence.

Table 3.1: **Calculation of Estimated CHD Morbidity for Ukraine (per 100 000)**

	Framingham		Morbidity/ mortality	Ukraine	
	morbidity	mortality	ratio	morbidity	mortality
Males	28.7	6.2	4.6	2777	600
Females	14.5	2.8	5.2	1864	360

Subsequently, from the graphs included in papers on the Framingham study, the biennial levels of new cases of CHD per 1000 were extracted, smoothed to five-year intervals, and multiplied by 50 to switch to rates per year per 100000. Then, average unweighted incidence for people aged 35-84 was calculated, the coefficients were determined to adjust the incidence to the expected value based on mortality levels (1.6 for men and 1.9 for women), and the expected incidence was calculated. It was assumed that people older than 84 years have the same incidence as people aged 80-84. In Table 3.2, columns which correspond to age groups 0-34 are not shown, as their incidence of CHD is assumed to be 0.

Table 3.2: **Calculation of Estimated Incidence of CHD in Ukraine By Gender And Age, per 100 000 Population per Year**

	Age groups											Coefficients
	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	>84	
Framingham biennial per 1000												
male	9	9	21	21	40	40	48	48	52	52		
female	1	1	7	7	20	20	26	26	46	46		
Smoothed												
male	9	13	17	27	34	43	45	49	51	52		
female	1	3	5	11	16	22	24	33	39	46		
Per year per 1000000												
male	450	650	850	1367	1683	2133	2267	2467	2533	2600		1700
female	50	150	250	567	783	1100	1200	1633	1967	2300		1000
Multiplied for Ukraine												
male	735	1062	1389	2233	2750	3485	3703	4030	4139	4248	4248	1.6
female	93	280	466	1056	1460	2051	2237	3045	3666	4288	4288	1.9

3.1.2 Relative risk

Peer-reviewed literature shows that younger smokers have a much greater relative risk of contracting CHD than older ones. Relative to never smokers, CHD risk among current smokers was highest in the youngest and lowest in the oldest participants. For example, among women aged 40 to 49 years, the hazard ratio was 8.5 (95% confidence interval [CI] = 5.0, 14), while it was 3.1 (95% CI = 2.0, 4.9) among those aged 70 or older. The largest absolute risk differences between current smokers and never smokers were observed among the oldest participants. Finally, the majority of CHD cases among smokers were attributable to smoking. For example, attributable proportions of CHD by age group were 88% (40-49 years), 81% (50-59 years), 71% for (60-69 years), and 68% (70 years) among women who smoked (Tolstrup et al., 2014).

Graphs from this systematic review which display RR by age groups are shown below.
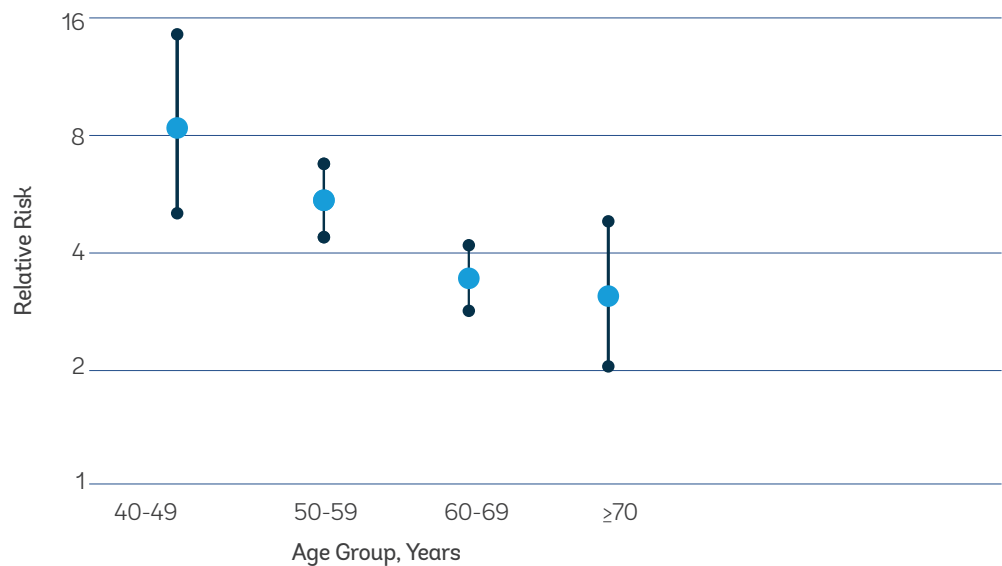
Figure 3.1: **RRs of Developing CHD among Female Smokers Compared with Non-Smokers by Age Groups, Results from a Systematic Review (Tolstrup et al., 2014).**
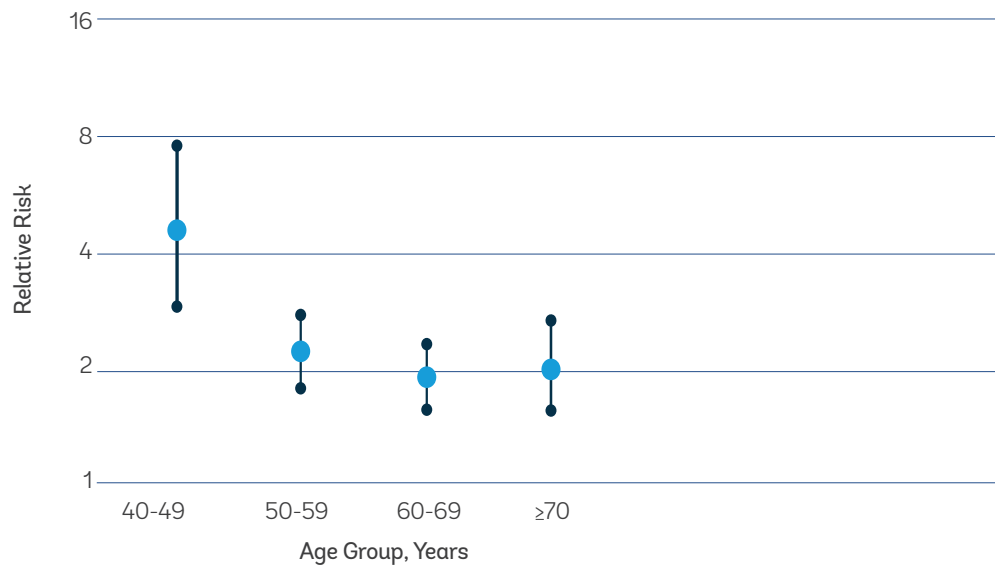
Figure 3.2: **RRs of Developing CHD among Male Smokers Compared
with Non-Smokers by Age Groups, Results from a Systematic Review
(Tolstrup et al., 2014).**

A population-based prospective cohort study of 19 782 men and 21 500
women aged 40-59 years between 1990-1992 and 2001 was conducted
to examine the relationship between smoking status and the risk of CHD.
A total of 260 incidences of CHD were confirmed among men, including
174 myocardial infarctions (MI). The numbers among women were 66 and
43, respectively. The multivariate relative risk [95% confidence interval (CI)]
for current smokers versus never-smokers in men, after adjustment for
cardiovascular risk factors and several lifestyle factors, was 2.85 (1.98, 4.12)
for total CHD and 3.64 (2.27, 5.83) for MI. These respective risks in women
were 3.07 (1.48, 6.40) and 2.90 (1.18, 7.18). Among men, a dose-dependent
relationship was observed between the number of cigarettes and the risk of
MI. The population-attributable risk percent (95% CI) of CHD was 46% (34, 55)
in men and 9% (0, 18) in women. Smoking cessation, however, led to a rapid
decline in the CHD risk within 2 years (Baba et al., 2006).

Smoking fewer cigarettes/day for a longer duration was more deleterious
than smoking more cigarettes/day for a shorter duration (P < 0.01). For 50
pack-years (365,000 cigarettes), estimated RRs of CVD were 2.1 for accrual at 20
cigarettes/day and 1.6 for accrual at 50 cigarettes/day (Lubin et al., 2016).

3.1.2.1 Women versus men
A systematic review and meta-analysis of prospective cohort studies with data
for 3 912 809 individuals and 67 075 coronary heart disease events from 86
prospective trials concluded as follows: In 75 cohorts (2.4 million participants)
that adjusted for cardiovascular risk factors other than coronary heart disease,

the pooled adjusted female-to-male RRR of smoking compared with not smoking for coronary heart disease was 1.25 (95% CI 1.12-1.39, p<0.0001). This outcome was unchanged after adjustment for potential publication bias, and there was no evidence of important between-study heterogeneity (p=0.21). The RRR increased by 2% for every additional year of study follow-up (p=0.03). In pooled data from 53 studies, there was no evidence of a sex difference in the RR between participants who had previously smoked compared with those who never had (RRR 0.96, 95% CI 0.86-1.08, p=0.53) (Huxley & Woodward, 2011).

3.1.2.2 Effects of quitting smoking

In a cohort of 475 734 Korean men aged 30 to 58 years, compared with non-reducing heavy smokers (>= 20 cigarettes/d), those who quit smoking showed significantly lower risks of MI with hazard ratios (95% confidence intervals [CI]) of 0.43 (0.34 to 0.53) (Song & Cho, 2008).

3.1.2.3 Suggested RR for smokers compared with non-smokers

Based on the above results from (Tolstrup et al., 2014) and (Song & Cho, 2008), the updated risk ratios might be as follows.

Table 3.3: **Suggested Input Risk Ratios of Developing CHD among Smokers Compared to Never Smokers in Ukraine, by Gender and Age**

Gender	Age groups					
	35-40	45-50	50-55	55-60	60-65	>65
men	5.0	4.0	3.0	3.0	2.0	2.0
women	8.5	8.5	6.6	4.8	3.4	3.1

3.1.3 Mortality

In the Greek cohort study (Notara et al., 2015), which observed 10-year Acute Coronary Syndrome (ACS) prognosis among 2172 cardiovascular patients, patients with >60 pack-years of smoking had 57.8 % higher ACS mortality and 24.6 % higher risk for any ACS event. A nested model, adjusted only for age and sex, revealed that, for every 30 pack-years of smoking increase, the associated ACS risk increased by 13 % (95 % CI 1.03, 1.30, p = 0.001).

Smoking is a strong independent risk factor for cardiovascular events and mortality even at older age, advancing cardiovascular mortality by more than five years, and demonstrating that smoking cessation in these age groups is still beneficial in reducing the excess risk. Random effects meta-analysis of the association of smoking status with cardiovascular mortality (based on the data of 503 905 participants aged 60 and older, of whom 37 952 died from

cardiovascular disease) yielded a summary hazard ratio of 2.07 (95% CI 1.82 to 2.36) for current smokers and 1.37 (1.25 to 1.49) for former smokers compared with never smokers. Corresponding summary estimates for risk advancement periods were 5.50 years (4.25 to 6.75) for current smokers and 2.16 years (1.38 to 2.39) for former smokers. The excess risk in smokers increased with cigarette consumption in a dose-response manner and decreased continuously with time since smoking cessation in former smokers (Mons et al., 2015).
In Sweden and Estonia, a 13-year follow-up regarding all-cause and cardiovascular mortality revealed that smoking and, to a lesser extent, plasma levels of interleukin-6 were significant predictors of CVD and non-CVD mortality in men, but none of the other conventional risk factors reached statistical significance (Jensen-Urstad, Viigimaa, Sammul, Lenhoff, & Johansson, 2014).

In a large prospective cohort of women (Sandhu et al., 2012) without coronary heart disease at baseline (among 101 018 women participating in the Nurses' Health Study), a strong dose-response relationship between cigarette smoking and SCD risk was observed, and smoking cessation significantly reduced and eventually eliminated excess SCD risk. Compared with never smokers, current smokers had a 2.44-fold (95% CI, 1.80-3.31) increased risk of SCD after controlling for coronary risk factors. In multivariable analyses, the quantity of cigarettes smoked daily (P value for trend, <0.0001) and smoking duration (P value for trend, <0.0001) were linearly associated with SCD risk among current smokers. Small-to-moderate amounts of cigarette consumption (1-14 per day) were associated with a significant 1.84-fold (95% CI, 1.16-2.92) increase in SCD risk and every 5 years of continued smoking was associated with an 8% increase in SCD risk (hazard ratio, 1.08; 95% CI, 1.05-1.12; P<0.0001). The SCD risk linearly decreased over time after quitting and was equivalent to that of a never-smoker after 20 years of cessation (P value for trend, <0.0001).

### 3.1.3.1 Effects of quitting smoking
A systematic review was conducted to determine the magnitude of risk reduction achieved by smoking cessation in patients with CHD. The researchers estimated a 36% reduction in crude relative risk (RR) of mortality for patients with CHD who quit, compared with those who continued smoking (RR, 0.64; 95% confidence interval [CI], 0.58-0.71) (Critchley & Capewell, 2003).

3.2 COPD

3.2.1 Incidence among the population

Incident cases of COPD in a population-based prospective 9-year study in Sao Paulo, Brazil, ranged from 1.4% to 4.0%, depending on the diagnostic criterion used (Moreira et al., 2015).

In the Rotterdam Study (Terzikhan et al., 2016), the overall IR was higher in men (13.3/1000 PY, 95 % CI 12.4–14.3) than in women (6.1/1000 PY, 95 % CI 5.6–6.6); age-specific IR ranged between 8.7 and 17.6/1000 PY in males and 3.0–7.9/1000 PY in females. The incidence of COPD increased from the age of 45 in both sexes to the age of 80 in men and 75 in women (Figure 3.3).
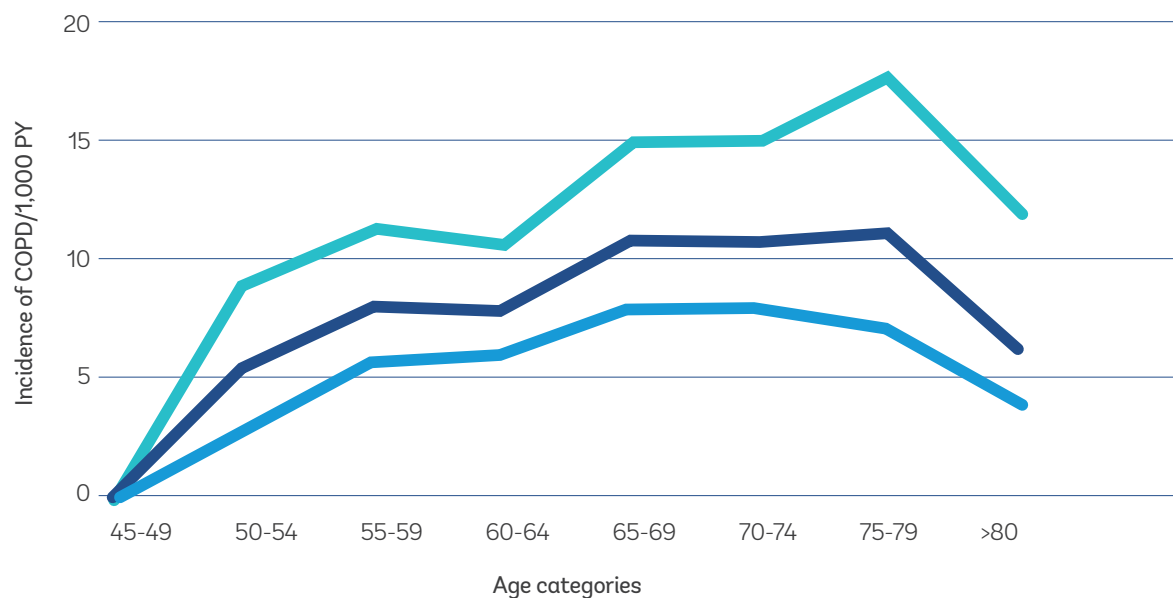
Figure 3.3: **Age-Specific Incidence of COPD by Sex and Age, Drawn from (Terzikhan et al., 2016)**

Although some studies report either relative risks of COPD among people in older age groups or mention that incidence increases after age 45 (Terzikhan et al., 2016), this does not mean that COPD is only occurring among people over 45. A systematic review on global burden of COPD (Halbert et al., 2006) reports pooled COPD prevalence at the level of 3.1% (1.8–5.0) among people younger than 40. Additionally, the WHO global report on mortality attributable to tobacco (World Health Organization, 2012) reports tobacco-related mortality starting from 30 years of age and estimates that 39% of COPD deaths among people aged 30-44 are attributable to tobacco.

The only study which reported the incidence of COPD by age and sex was conducted in Japan (Kojima et al., 2007). Its findings were used to estimate incidence by age groups in Ukraine.

Table 3.4: **Extract on the Incidence of COPD by Age and Sex in Japan (Kojima et al., 2007)**

Age (years)	Males			Females		
	n	Number of incidence cases for COPD	Incidence rate (per 100 person-years)	n	Number of incidence cases for COPD	Incidence rate (per 100 person-years)
Total	11,160	387	0.81	5,946	79	0.31
25-29	94	2	0.62	36	0	0.00
30-34	625	7	0.31	181	1	0.16
35-39	1,609	24	0.35	712	4	0.13
40-44	1,973	45	0.47	1,161	10	0.18
45-49	2,153	65	0.61	1,279	12	0.19
50-54	1,879	90	1.05	1,157	21	0.42
55-59	1,729	74	1.25	1,002	12	0.35
60-64	745	39	1.67	264	9	1.02
65-69	252	24	2.75	109	7	1.69
70-74	101	17	4.95	45	3	2.05

Another study conducted in Japan (Fukuchi et al., 2004) reported the prevalence of COPD among adults: 10.9% altogether, 16.4% among men and 5.0% among women.

These data on incidence and prevalence were considered in order to obtain extrapolated estimates for Ukraine. However, the only disease occurrence indicator available for Ukraine is the prevalence of COPD from the WHO Euro Health for all database, which reports a level of 3.7-3.9% in 2005-2015. However, studies aimed at COPD measurement conducted, for instance, in Norway (Johannessen, Omenaas, Bakke, & Gulsvik, 2005) found that about half of COPD cases remain undiagnosed. Another study (Nielsen, 2009) projected COPD prevalence to be 15-25% of the adult population. Yet the prevalence of COPD in Norway reported in HFADB is 0.2%.

Additionally, there is a recognized discrepancy in COPD prevalence across different countries and various studies. This is believed to be determined by the methods and definitions used to measure disease (Halbert, Isonaka, George, & Iqbal, 2003). Prevalence in most countries where proper measures were conducted was found to be between 4% and 10%.

Thus, there are no serious grounds to expect that the incidence of COPD in Ukraine should be lower than in Japan. Data on COPD incidence from the Japan study (Kojima et al., 2007) are suggested for use.

Table 3.4: **Suggested Incidence Rates for COPD**

	Age groups									
	25-29	**30-34**	**35-39**	**40-44**	**45-49**	**50-54**	**55-59**	**60-64**	**65-69**	**70-74**
Men	0.62	0.31	0.35	0.47	0.61	1.05	1.25	1.67	2.75	4.95
Women	0.00	0.16	0.13	0.18	0.19	0.42	0.35	1.02	1.69	2.05

3.2.2 Risk of COPD in smokers

The incidence rate (IR) (Terzikhan et al., 2016) was higher in current and former smokers than in never smokers (19.7/1000 PY, 95 % CI 18.1–21.4 in current smokers, 8.3/1000 PY, 95 % CI 7.6–9.1 in former smokers and 4.1/1000 PY, 95 % CI 3.6–4.7, in never smokers). The IR of COPD in smoking men was 15.0/1000 PY (95 % CI 13.9–16.2), compared to 8.6/1000 PY (95 % CI 7.8–9.5) in smoking women. The age-specific IR of COPD in ever smokers ranged between 7.3 and 15.3/1000 PY. The IR was 6.0/1000 PY (95 % CI 4.6–7.8) in never-smoking men and 3.7/1,000 PY (95 % CI 3.1–4.3) in never-smoking women. The age-specific incidence of COPD in never smokers increased by age, but to a lesser extent than the incidence of COPD in ever smokers (Figure 3. 5).
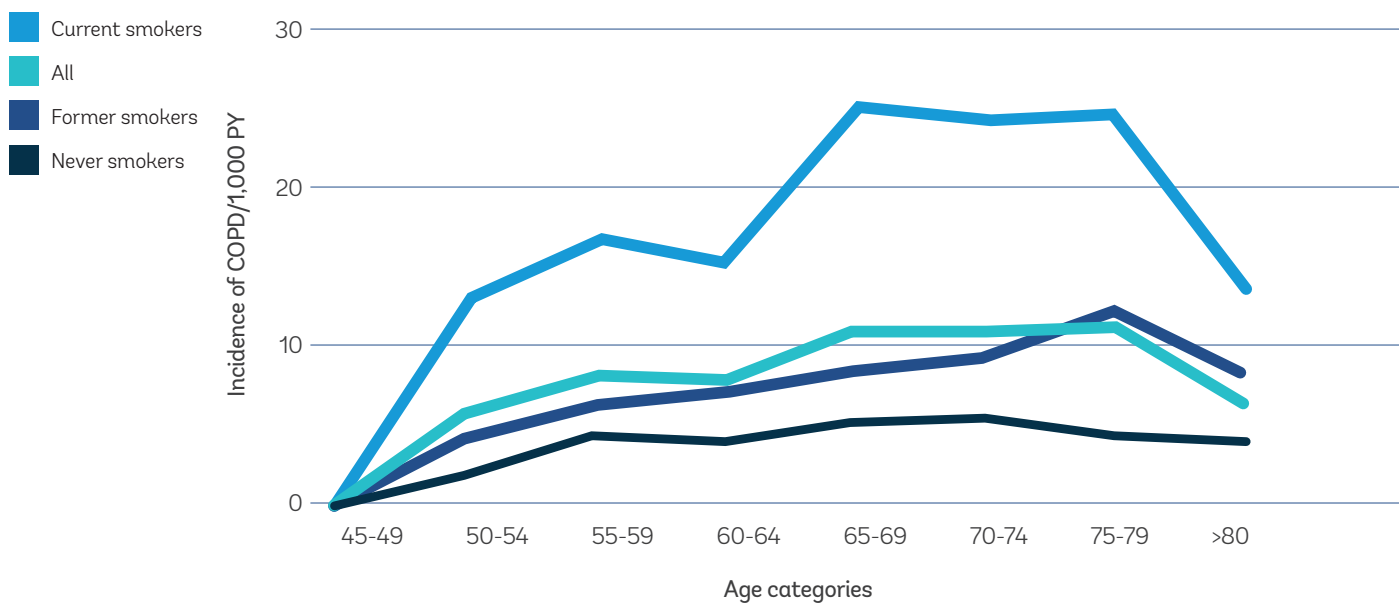
Figure 3.5: **Age-Specific Incidence of COPD by Smoking Status, Extracted from (Terzikhan et al., 2016)**

The abovementioned study from Norway found that adjusted odds ratios (OR) for current smokers and ex-smokers were 9.6 (95% CI 3.6-25.2) and 5.0 (95% CI 1.8-13.8), compared to never smokers (Johannessen et al., 2005).

The comparisons of risks from these two studies were extracted for the model.

Table 3.5: **Suggested Estimates for COPD RR in Smokers and Former Smokers, Compared with Never Smokers**

	Age groups							
	35-40	**45-50**	**50-55**	**55-60**	**60-65**	**65-69**	**70-74**	**75+**
Smokers	9.6	8.4	6.6	5.0	4.6	4.5	4.9	5.0
Ex smokers	5.0	4.7	4.2	3.8	4.0	4.5	4.8	5.0

The suggested RRs are equal for men and women. With regard to various tobacco-related diseases, some researchers have reported that the risk of contracting these diseases is greater in women than in men, while others found that the risk is identical. In Danish cohorts (Prescott, Bjerg, Andersen, Lange, & Vestbo, 1997), it was seen that risk associated with pack-years was higher in females than in males. As women smokers in Ukraine on average smoke fewer cigarettes, equal RRs for men and women can be considered grounded.

3.3 Hypertension

3.3.1.1 Effects of smoking

While some authors find an association between tobacco smoking and high blood pressure (Tesfaye, Byass, & Wall, 2009; Tesfaye, Byass, Wall, Berhane, & Bonita, 2008), it is necessary to distinguish between short-term, acute hypertensive effects and the long-term risk of developing chronic hypertension.

Cigarette smoking acutely exerts a hypertensive effect, mainly through the stimulation of the sympathetic nervous system. As regards the impact of chronic smoking on blood pressure, available data do not provide evidence of a direct causal relationship between these two cardiovascular risk factors (Poulter, 2002), a concept supported by the evidence that lower blood pressure values have not been observed after chronic smoking cessation (Virdis, Giannarelli, Neves, Taddei, & Ghiadoni, 2010). Though the prevalence of hypertension was higher in former smokers than in never smokers (13.5 versus 8.8%, P < 0.001) and the risk of hypertension was higher [odds ratio (OR) 1.31 (1.13-1.52), P< 0.001] in former smokers than in never smokers (Halimi et al., 2002), these findings were from a cross-sectional study, and no grounds for cause-and-effect association are found in this regard (Poulter, 2002).

3.3.1.1 Effects of quitting smoking

The effect of smoking cessation on the risk of developing hypertension (HPT) and on BP values was studied in a longitudinal study, with a follow-up period of 8 years, which included the participants of the Olivetti Heart Study. These were 430 untreated normotensive non-diabetic men with normal renal function (D'Elia et al., 2014). After 8 years of follow-up, BP changes (delta) were significantly lower in ex-smokers than in smokers (delta SBP/DBP: 12.6 +/- 13.4/7.9 +/- 8.1 vs. 16.0 +/- 14.9/10.3 +/- 10.1 mm Hg; P < 0.05; M +/- SD), also after adjustment for potential confounders. Moreover, at the last examination, the overall HPT prevalence was 33%, with lower values in ex-smokers than in smokers (25 vs. 38%, P = 0.01). After accounting for age, BP and BMI at baseline, and changes in smoking habits over the 8-year period, ex-smokers still had significantly lower risk of HPT than smokers (odds ratio 0.30, 95% confidence interval 0.15-0.58; P < 0.01).

Taking into account contradictory data on the impact of tobacco smoking on developing hypertension, we decided to exclude hypertension from the list of diseases modeled in the microsimulation of tobacco-related health impacts.

3.4 Lung Cancer

3.4.1 Relative risk
3.4.1.1 Effects of smoking

In the Seoul Male Cancer Cohort Study (SMCC), which included 14 272 men, cigarette smoking was associated with 4.18-fold risk of lung cancer in Korean men (Bae et al., 2007).

3.4.1.1 Women versus men

Analysis was conducted on the data of 279 214 men and 184 623 women from eight states in the USA, aged 50-71 years at study baseline, participating in the NIH-AARP Diet and Health study. Findings revealed that incidence rates were 20.3 (95% CI 16.3-24.3) per 100 000 person-years in men who had never smoked (99 cancers) and 25.3 (21.3-29.3) in women who had never smoked (152 cancers); for this group, the adjusted hazard ratio for lung cancer was 1.3 (1.0-1.8) for women compared with men.

Smoking was associated with increased risk of lung cancer in men and women. The incidence rate of current smokers who smoked more than two packs per day was 1259.2 (1035.0-1483.3) in men and 1308.9 (924.2-1693.6) in women. In current smokers, in a model adjusted for typical smoking dose, the HR was 0.9 (0.8-0.9) for women compared with men.

For former smokers, in a model adjusted for years of cessation and typical smoking dose, the HR was 0.9 (0.9-1.0) for women compared with men.

Incidence rates of adenocarcinoma, small-cell carcinoma, and undifferentiated tumors were similar in men and women; incidence rates of squamous tumors in men were about twice those in women. These findings suggest that women are not more susceptible than men to the carcinogenic effects of cigarette smoking in the lung. In smokers, incidence rates tended to be higher in men than women with comparable smoking histories, but differences were modest; smoking was strongly associated with lung cancer risk in both men and women (Freedman, Leitzmann, Hollenbeck, Schatzkin & Abnet, 2008).

3.4.2 Mortality

3.4.2.1 Effects of smoking

In the Japan Collaborative Cohort (JACC) Study, with 45 010 males and 55 724 females aged 40-79 years, 52.2% and 14.8% of lung cancer deaths were attributable to current and former cigarette smoking, respectively. In females, the corresponding figures were 11.8% and 2.8%. Among current male smokers, the relative risk was strongly correlated with the intensity and duration of cigarette smoking. In contrast, the PAR was associated with an intermediate level of smoking except for the years of smoking: the largest PARs were observed in those with 20-29 cigarettes per day, 40-59 pack-years and 20-22 years old at smoking inception. Absolute risks were estimated to increase with age and duration of smoking and not to decrease even after cessation (Ando et al., 2003).

3.4.1.1 Effects of quitting smoking

Pooled data from three large-scale cohort studies in Japan were used to evaluate the impact of smoking cessation on the decrease in risk of lung cancer death in male ex-smokers by age at quitting. For simplicity, subjects were limited to male never smokers and former or current smokers who started smoking at ages 18-22 years. 110 002 men aged 40-79 years at baseline were included. During the mean follow-up of 8.5 years, 968 men died from lung cancer. The mortality rate ratio compared to current smokers decreased with increasing attained age in men who stopped smoking before age 70 years. Among men who quit in their fifties, the cohort-adjusted mortality rate ratios (95% confidence interval) were 0.57 (0.40-0.82), 0.44 (0.29-0.66) and 0.36 (0.13-1.00) at attained ages 60-69, 70-79 and 80-89 years, respectively. The corresponding figures for those who quit in their sixties were 0.81 (0.44-1.48), 0.60 (0.43-0.82) and 0.43 (0.21-0.86). Overall, the mortality rate ratio for current smokers, relative to nonsmokers, was 4.71 (95% confidence interval 3.76-5.89) and those for ex-smokers who had quit smoking 0-4, 5-9, 10-14, 15-19, 20-24 and >= 25 years before were 3.99 (2.97-5.35), 2.55 (1.80-3.62), 1.87 (1.23-2.85), 1.21 (0.66-2.22), 0.76 (0.33-1.75) and 0.67 (0.34-1.32), respectively. Although earlier cessation of smoking generally resulted in a lower rate of lung cancer mortality in each group of attained age, the absolute mortality rate decreased appreciably after stopping smoking even in men who quit at ages 60-69 years (Wakai et al., 2007).

3.5 Peripheral arterial disease (PAD)

In a meta-analysis of the association between cigarette smoking and PAD, the pooled OR for current smokers was 2.71 (95% CI 2.28 to 3.21); for ex-smokers, the pooled OR was 1.67 (95% CI 1.54 to 1.81). The magnitude of the association is greater than that reported for coronary heart disease. The risk is lower among ex-smokers but, nonetheless, significantly increased compared with never smokers (Lu, Mackay, & Pell, 2014).

3.5.1 Any stroke

3.5.1.1 Effects of smoking

In a meta-analysis on the possible risks of stroke from cigarette smoking (Shinton & Beevers, 1989), the overall relative risk of stroke associated with cigarette smoking was 1.5 (95% confidence interval 1.4 to 1.6). Considerable differences were seen in relative risks among the subtypes: Cerebral infarction 1.9, cerebral hemorrhage 0.7, and subarachnoid hemorrhage 2.9. An effect of age on the relative risk was also noted; less than 55 years 2.9, 55-74 years 1.8, and greater than or equal to 75 years 1.1. A dose response between the number of cigarettes smoked and relative risk was noted, and there was a small increased risk in women compared with men. Ex-smokers under the age of 75 seemed to retain an appreciably increased risk of stroke (1.5); for all ages, the relative risk in ex-smokers was 1.2.

In a prospective study (Wannamethee, Shaper, Whincup, & Walker, 1995) of cardiovascular disease and its risk factors, 7735 men aged 40 through 59 years were drawn at random from the age-sex registers of one general practice in each of 24 British towns from 1978 through 1980 (the British Regional Heart Study). During the 12.75 years of follow-up, there were 167 major stroke events (43 fatal and 124 non-fatal) in the 7264 men with no recall of previous ischemic heart disease or stroke. After full adjustment for other risk factors, current smokers had a nearly fourfold relative risk (RR) of stroke compared with never smokers (RR, 3.7; 95% confidence interval [CI], 2.0 to 6.9).

Ex-smokers showed lower risk than current smokers, but showed excess risk compared with never smokers (RR, 1.7; 95% CI, 0.9 to 3.3; P = .11); those who switched to pipe or cigar smoking showed a significantly increased risk (RR, 3.3; 95% CI, 1.6 to 7.1), similar to that of current light smokers. Primary pipe or cigar smokers also showed increased risk (RR, 2.2; 95% CI, 0.6 to 8.0), but the number of subjects involved was small. The benefit of giving up smoking completely was seen within five years of quitting, with no further consistent decline in risk thereafter, but this was dependent on the amount of tobacco smoked. Light smokers (< 20 cigarettes/d) reverted to the risk level of those who had never smoked. Heavy smokers retained a more than twofold risk compared with never smokers (RR, 2.2; 95% CI, 1.1 to 4.3). The age-adjusted RR of stroke in those who quit smoking during the first five years of follow-up (recent quitters)

was reduced compared with continuing smokers (RR, 1.8; 95% CI, 0.7 to 4.6 vs. RR, 4.3; 95% CI, 2.1 to 8.8). The benefit of quitting smoking was observed in both normotensive and hypertensive men, but the absolute benefit was greater in hypertensive subjects. Thus, smoking cessation is associated with a considerable and rapid benefit in decreasing the risk of stroke, particularly in light smokers (< 20 cigarettes/d); a complete loss of risk is not seen in heavy smokers. Switching to pipe or cigar smoking confers little benefit, emphasizing the need for complete cessation of smoking. The absolute benefit of quitting smoking on risk of stroke is most marked in hypertensive subjects.

In the Japan Public Health Center-based Prospective Study on Cancer and Cardiovascular Disease (JPHC Study), relative risks (95% CIs) for current smokers compared with never-smokers, after adjustment for cardiovascular risk factors and public health center, were 1.27 (1.05 to 1.54) for total stroke, 0.72 (0.49 to 1.07) for intraparenchymal hemorrhage, 3.60 (1.62 to 8.01) for subarachnoid hemorrhage, and 1.66 (1.25 to 2.20) for ischemic stroke. The respective multivariate relative risks among women were 1.98 (1.42 to 2.77), 1.53 (0.86 to 4.25), 2.70 (1.45 to 5.02), and 1.57 (0.86 to 2.87). There was a dose-response relation between the number of cigarettes smoked and risks of ischemic stroke for men. A similar positive association was observed between smoking and risks of lacunar infarction and large-artery occlusive infarction, but not embolic infarction (Mannami et al., 2004).

3.5.1.1 Women versus men

In a systematic review and meta-analysis which aimed to estimate the effect of smoking on stroke in women compared with men (Peters, Huxley, & Woodward, 2013), with data from 81 prospective cohort studies that included 3 980 359 individuals and 42 401 strokes, the pooled multiple-adjusted RRR indicated a similar risk of stroke associated with smoking in women compared with men (RRR, 1.06 [95% confidence interval, 0.99-1.13]). In a regional analysis, there was evidence of a more harmful effect of smoking in women than in men in Western populations (RRR, 1.10 [1.02-1.18]), but not in Asian populations (RRR, 0.97 [0.87-1.09]). Compared with never-smokers, the beneficial effects of quitting smoking on stroke risk among former smokers were similar between the sexes (RRR, 1.10 [0.99-1.22]).

3.5.2 Ischemic stroke

Smoking is associated with an increased risk of ischemic stroke or CV death in the Atherosclerosis Risk in Communities (ARIC) Study, which comprised mostly middle-aged to young-old subjects (65-74 years), but not in the Cardiovascular Health Study (CHS), which comprised mostly middle-old or oldest-old (>= 75 years) adults with atrial fibrillation. Compared with never smokers, current smokers had a higher incidence of the composite endpoint in ARIC [HR: 1.65 (1.21-2.26)], but not in CHS [HR: 1.05 (0.69-1.61)] (Kwon et al., 2016).

3.5.2.1 Effects of quitting or reducing smoking

In a cohort of 475 734 Korean men aged 30 to 58 years, compared with non-reducing heavy smokers (>= 20 cigarettes/d), those who quit smoking showed significantly lower risks of ischemic stroke with hazard ratios (95% confidence intervals [CI]) of 0.66 (0.55 to 0.79). Compared with non-reducing heavy smokers, the risks of all strokes combined and MI among reducers tended to decrease, although the decrements were not statistically significant (Song & Cho, 2008).

3.5.3 Hemorrhagic stroke

3.5.3.1 Relative risk

3.5.3.1.1 Effects of smoking

In a study of incident cerebral microbleeds (CMBs), which are asymptomatic precursors of intracerebral hemorrhage, conducted among 2635 individuals aged 66 to 93 years from the population-based Age, Gene/Environment Susceptibility (AGES)-Reykjavik Study, relative risk for current smoking was 1.47 [95% CI, 1.11-1.94] (Ding et al., 2015).

3.5.1.1.1 Effects of quitting or reducing smoking

In a cohort of 475 734 Korean men aged 30 to 58 years, compared with non-reducing heavy smokers (>= 20 cigarettes/d), those who quit smoking showed significantly lower risks of subarachnoid hemorrhage with hazard ratios (95% confidence intervals [CI]) of 0.58 (0.38 to 0.90). For hemorrhagic stroke, quitters showed lower risk compared with heavy smokers, but the difference was not statistically significant (hazard ratio 0.82, 95% CI: 0.64 to 1.06). The risks of subarachnoid hemorrhage in those who reduced from moderate to light smoking tended to be lower than in non-reducing moderate (10 to 19 cigarettes/d) smokers (Song & Cho, 2008).

3.5.1.2 Suggested relative risks

Based on the above literature on cumulative risk of all strokes, the below RRs are suggested for the model.

Table 3.6: **Estimates of All Strokes Relative Risk for Smokers and Ex-Smokers Compared to Never Smokers for the Microsimulation Model by Age Group, Both Genders**

	Age groups					
	35-40	**40-50**	**50-55**	**55-60**	**60-65**	**>65**
Smokers	1.7	1.7	1.5	1.5	1.2	1.2
Ex smokers	4.3	3.7	2.9	1.8	1.8	1.2

3.6 All-Cause Mortality

3.6.1.1 Effects of smoking
In a large community-based prospective cohort study comprising 6209 Beijing adults (aged >= 40 years) studied for approximately eight years (1991-1999), the multivariable-adjusted HRs for all-cause mortality were 2.7 (95% confidence interval (CI):1.56-4.69) in young adult smokers (40-50 years) and 1.31 (95% CI: 1.13-1.52) in old smokers (>50 years) (Li et al., 2016). Mortality differences (/10,000 person-years) were 15.99 (95% CI: 15.34-16.64) in the young and 74.61(68.57-80.65) in the old. Compared with current smokers, the HRs of all-cause deaths for former smokers in younger and older adults were 0.57 (95% CI: 0.23-1.42) and 0.96 (95% CI: 0.73-1.26), respectively.

Among 20 033 individuals participating in the Health Effects of Arsenic Longitudinal Study (HEALS) in Bangladesh, cigarette/bidi smoking was positively associated with all-cause (HR 1.40, 95% CI 1.06 1.86) and cancer mortality (HR 2.91, 1.24 6.80), and there was a dose-response relationship between increasing intensity of cigarette/bidi consumption and increasing mortality. An elevated risk of death from ischemic heart disease (HR 1.87, 1.08 3.24) was associated with current cigarette/bidi smoking. Among women, the corresponding HRs were 1.65 (95% CI 1.16 2.36) for all-cause mortality and 2.69 (95% CI 1.20 6.01) for ischemic heart disease mortality. Cigarette/bidi smoking accounted for about 25.0% of deaths in men and 7.6% in women (Wu et al., 2013).

3.6.1.1 Effects of quitting smoking
Effects of quitting smoking on all-cause mortality were measured in a cohort of 1 494 Chinese people (961 men, 533 women) followed for 18 years (1976-1994) to assess changes in smoking behavior and then for an additional 17 years (1994-2011) to examine the relationships of continuing to smoke and new quitting with mortality risk. Ever smokers had increased risks of lung cancer, coronary heart disease, thrombotic stroke, and COPD, with dose-response relationships. For all tobacco-related mortality, the relative risk for new quitters compared with continuing smokers was 0.68 (95% confidence interval: 0.46, 0.99) for those who had quit two to seven years previously and 0.56 (95% confidence interval: 0.37, 0.85) for those who had quit eight years or more previously. The corresponding relative risks were 0.69 and 0.45 for lung cancer, 0.78 and 0.51 for coronary heart disease, 0.76 and 0.84 for thrombotic stroke, and 0.89 and 0.61 for COPD, respectively (He et al., 2014).

In the Singapore Chinese Health Study, a cohort study of middle-aged and elderly Chinese in Singapore (n=48 251), compared with current smokers, the adjusted HR (95% CI) for total mortality was 0.84 (0.76 to 0.94) for new quitters, 0.61 (0.56 to 0.67) for long-term quitters and 0.49 (0.46 to 0.53) for never-smokers. New quitters had a 24% reduction in lung cancer mortality (HR: 0.76, 95% CI 0.57 to 1.00), and long-term quitters had a 56% reduction (HR: 0.44, 95% CI 0.35 to 0.57). The risk for coronary heart disease mortality was reduced in new quitters (HR: 0.84, 95% CI 0.66 to 1.08) and long-term quitters (HR: 0.63, 95% CI 0.52 to 0.77), although the result for new quitters was of borderline significance due to the relatively small number of cardiovascular deaths. The risk for chronic pulmonary disease mortality was reduced in long-term quitters but increased in new quitters. The authors concluded that significant reduction in the risk of total mortality, specifically for lung cancer mortality, can be achieved within five years of smoking cessation (Lim, Tai, Yuan, Yu, & Koh, 2013).

References

Ando, M., Wakai, K., Seki, N., Tamakoshi, A., Suzuki, K., Ito, Y., . . . Grp, J. S. (2003). Attributable and absolute risk of lung cancer death by smoking status: Findings from the Japan Collaborative Cohort Study. *International Journal of Cancer,* 105(2), 249-254. doi:10.1002/ijc.11043

Baba, S., Iso, H., Mannami, T., Sasaki, S., Okada, K., Konishi, M., . . . Grp, J. S. (2006). Cigarette smoking and risk of coronary heart disease incidence among middle-aged Japanese men and women: The JPHC Study Cohort I. *European Journal of Cardiovascular Prevention & Rehabilitation,* 13(2), 207-213. doi:10.1097/01. hjr.0000194417.16638.3d

Bae, J. M., Lee, M. S., Shin, M. H., Kim, D. H., Li, Z. M., & Ahn, Y. O. (2007). Cigarette smoking and risk of lung cancer in Korean men: The Seoul male cancer cohort study. *Journal of Korean Medical Science,* 22(3), 508-512.

Castelli, W. P. (1984). Epidemiology of coronary heart disease: The Framingham study. *Am J Med,* 76(2a), 4-12.

Critchley, J. A., & Capewell, S. (2003). Mortality risk reduction associated with smoking cessation in patients with coronary heart disease: A systematic review. *Journal of the American Medical Association,* 290(1), 86-97. doi:10.1001/ jama.290.1.86

D'Elia, L., De Palma, D., Rossi, G., Strazzullo, V., Russo, O., Iacone, R., . . . Galletti, F. (2014). Not smoking is associated with lower risk of hypertension: Results of the Olivetti Heart Study. *European Journal of Public Health,* 24(2), 226-230. doi:10.1093/eurpub/ckt041

Ding, J., Sigurdsson, S., Garcia, M., Phillips, C. L., Eiriksdottir, G., Gudnason, V., . . . Launer, L. J. (2015). Risk factors associated with incident cerebral microbleeds according to location in older people: The Age, Gene/Environment Susceptibility (AGES)-Reykjavik Study. *JAMA Neurology,* 72(6), 682-688. doi:10.1001/ jamaneurol.2015.0174

Freedman, N. D., Leitzmann, M. F., Hollenbeck, A. R., Schatzkin, A., & Abnet, C. C. (2008). Cigarette smoking and subsequent risk of lung cancer in men and women: Analysis of a prospective cohort study. *Lancet Oncology,* 9(7), 649-656. doi:10.1016/51470-2045(08)70154-2

Fukuchi, Y., Nishimura, M., Ichinose, M., Adachi, M., Nagai, A., Kuriyama, T., . . . Zaher, C. (2004). COPD in Japan: The Nippon COPD Epidemiology Study. *Respirology*, 9(4), 458-465. doi:10.1111/j.1440-1843.2004.00637.x

Global Burden of Disease. (2016). *Global Health Data Exchange*. Retrieved from: http://ghdx.healthdata.org/gbd-results-tool

Halbert, R. J., Isonaka, S., George, D., & Iqbal, A. (2003). Interpreting COPD prevalence estimates: What is the true burden of disease? *Chest*, 123(5), 1684-1692.

Halbert, R. J., Natoli, J. L., Gano, A., Badamgarav, E., Buist, A. S., & Mannino, D. M. (2006). Global burden of COPD: Systematic review and meta-analysis. *Eur Respir J,* 28(3), 523-532. doi:10.1183/09031936.06.00124605

Halimi, J. M., Giraudeau, B., Vol, S., Caces, E., Nivet, H., & Tichet, J. (2002). The risk of hypertension in men: Direct and indirect effects of chronic smoking. *Journal of Hypertension,* 20(2), 187-193. doi:10.1097/00004872-200202000-00007

He, Y., Jiang, B., Li, L. S., Li, L. S., Sun, D. L., Wu, L., . . . Lam, T. H. (2014). Changes in smoking behavior and subsequent mortality risk during a 35-year follow-up of a cohort in Xi'an, China. *Am J Epidemiol,* 179(9), 1060-1070. doi:10.1093/aje/kwu011

Huxley, R. R., & Woodward, M. (2011). Cigarette smoking as a risk factor for coronary heart disease in women compared with men: A systematic review and meta-analysis of prospective cohort studies. *Lancet,* 378(9799), 1297-1305. doi:10.1016/s0140-6736(11)60781-2

Jensen-Urstad, M., Viigimaa, M., Sammul, S., Lenhoff, H., & Johansson, J. (2014). Impact of smoking: All-cause and cardiovascular mortality in a cohort of 55-year-old Swedes and Estonians. *Scandinavian Journal of Public Health,* 42(8), 780-785. doi:10.1177/1403494814550177

Johannessen, A., Omenaas, E., Bakke, P., & Gulsvik, A. (2005). Incidence of GOLD-defined chronic obstructive pulmonary disease in a general adult population. *Int J Tuberc Lung Dis,* 9(8), 926-932.

Kojima, S., Sakakibara, H., Motani, S., Hirose, K., Mizuno, F., Ochiai, M., & Hashimoto, S. (2007). Incidence of chronic obstructive pulmonary disease, and the relationship between age and smoking in a Japanese population. *J Epidemiol,* 17(2), 54-60.

Kwon, Y., Norby, F. L., Jensen, P. N., Agarwal, S. K., Soliman, E. Z., Lip, G. Y. H., . . . Chen, L. Y. (2016). Association of smoking, alcohol, and obesity with cardiovascular death and ischemic stroke in atrial fibrillation: The Atherosclerosis Risk in Communities (ARIC) Study and Cardiovascular Health Study (CHS). *PLoS One,* 11(1), 13. doi:10.1371/journal.pone.0147065

Lerner, D. J., & Kannel, W. B. (1986). Patterns of coronary heart disease morbidity and mortality in the sexes: A 26-year follow-up of the Framingham population. *Am Heart J,* 111(2), 383-390.

Li, K. B., Yao, C. H., Di, X., Yang, X. C., Dong, L., Xu, L., & Zheng, M. L. (2016). Smoking and risk of all-cause deaths in younger and older adults: A population-based prospective cohort study among Beijing adults in China. *Medicine (Baltimore),* 95(3), 5. doi:10.1097/md.0000000000002438

Lim, S. H., Tai, B. C., Yuan, J. M., Yu, M. M. C., & Koh, W. P. (2013). Smoking cessation and mortality among middle-aged and elderly Chinese in Singapore: The Singapore Chinese health study. *Tobacco Control,* 22(4), 235-240. doi:10.1136/tobaccocontrol-2011-050106

Lu, L., Mackay, D. F., & Pell, J. P. (2014). Meta-analysis of the association between cigarette smoking and peripheral arterial disease. *Heart,* 100(5), 414-423. doi:10.1136/heartjnl-2013-304082

Lubin, J. H., Couper, D., Lutsey, P. L., Woodward, M., Yatsuya, H., & Huxley, R. R. (2016). Risk of cardiovascular disease from cumulative cigarette use and the impact of smoking intensity. *Epidemiology*, 27(3), 395-404. doi:10.1097/ede.0000000000000437

Mannami, T., Iso, H., Baba, S., Sasaki, S., Okada, K., Konishi, M., . . . (2004). Cigarette smoking and risk of stroke and its subtypes among middle-aged Japanese men and women: The JPHC Study Cohort I. *Stroke*, 35(6), 1248-1253. doi:10.1161/01.STR.0000128794.30660.e8

Mons, U., Muezzinler, A., Gellert, C., Schottker, B., Abnet, C. C., Bobak, M., . . . Consortium, C. (2015). Impact of smoking and smoking cessation on cardiovascular events and mortality among older adults: Meta-analysis of individual participant data from prospective cohort studies of the CHANCES consortium. *British Medical Journal,* 350, 12. doi:10.1136/bmj.h1551

Moreira, G. L., Gazzotti, M. R., Manzano, B. M., Nascimento, O., Perez-Padilla, R., Menezes, A. M. B., & Jardim, J. R. (2015). Incidence of chronic obstructive pulmonary disease based on three spirometric diagnostic criteria in Sao Paulo, Brazil: A nine-year follow-up since the PLATINO prevalence study. *Sao Paulo Medical Journal,* 133(3), 245-251. doi:10.1590/1516-3180.2015.9620902

Nielsen, R. (2009). Present and future costs of COPD in Iceland and Norway: Results from the BOLD study. *Eur Respir J,* 34(4), 850-857. doi:10.1183/09031936.00166108

Notara, V., Panagiotakos, D. B., Kouroupi, S., Stergiouli, I., Kogias, Y., Stravopodis, P., . . . Investigators, G. S. (2015). Smoking determines the 10-year (2004-2014) prognosis in patients with Acute Coronary Syndrome: The GREECS observational study. *Tobacco Induced Diseases,* 13, 9. doi:10.1186/s12971-015-0063-6

Peters, S. A. E., Huxley, R. R., & Woodward, M. (2013). Smoking as a risk factor for stroke in women compared with men: A systematic review and meta-analysis of 81 cohorts, including 3 980 359 individuals and 42 401 strokes. *Stroke,* 44(10), 2821-2828. doi:10.1161/strokeaha.113.002342

Poulter, N. R. (2002). Independent effects of smoking on risk of hypertension: Small, if present. *Journal of Hypertension,* 20(2), 171-172. doi:10.1097/00004872-200202000-00002

Prescott, E., Bjerg, A. M., Andersen, P. K., Lange, P., & Vestbo, J. (1997). Gender difference in smoking effects on lung function and risk of hospitalization for COPD: Results from a Danish longitudinal population study. *Eur Respir J*, 10(4), 822-827.

Sandhu, R. K., Jimenez, M. C., Chiuve, S. E., Fitzgerald, K. C., Kenfield, S. A., Tedrow, U. B., & Albert, C. M. (2012). Smoking, smoking cessation, and risk of sudden cardiac death in women. *Circulation-Arrhythmia and Electrophysiology,* 5(6), 1091-1097. doi:10.1161/circep.112.975219

Shinton, R., & Beevers, G. (1989). Meta-analysis of relation between cigarette smoking and stroke. *BMJ*, 298(6676), 789-794.

Song, Y. M., & Cho, H. J. (2008). Risk of stroke and myocardial infarction after reduction or cessation of cigarette smoking: A cohort study in Korean men. *Stroke*, 39(9), 2432-2438. doi:10.1161/strokeaha.107.512632

Terzikhan, N., Verhamme, K. M. C., Hofman, A., Stricker, B. H., Brusselle, G. G., & Lahousse, L. (2016). Prevalence and incidence of COPD in smokers and non-smokers: The Rotterdam Study. *European Journal of Epidemiology,* 31(8), 785-792. doi:10.1007/s10654-016-0132-z

Tesfaye, F., Byass, P., & Wall, S. (2009). Population-based prevalence of high blood pressure among adults in Addis Ababa: Uncovering a silent epidemic. *BMC Cardiovasc Disord,* 9, 39. doi:10.1186/1471-2261-9-39

Tesfaye, F., Byass, P., Wall, S., Berhane, Y., & Bonita, R. (2008). Association of smoking and khat (Catha edulis Forsk) use with high blood pressure among adults in Addis Ababa, Ethiopia, 2006. *Prev Chronic Dis*, 5(3), A89.

Tolstrup, J. S., Hvidtfeldt, U. A., Flachs, E. M., Spiegelman, D., Heitmann, B. L., Balter, K., . . . Feskanich, D. (2014). Smoking and risk of coronary heart disease in younger, middle-aged, and older adults. *American Journal of Public Health,* 104(1), 96-102. doi:10.2105/ajph.2012.301091

Virdis, A., Giannarelli, C., Neves, M. F., Taddei, S., & Ghiadoni, L. (2010). Cigarette smoking and hypertension. *Curr Pharm Des,* 16(23), 2518-2525.

Wakai, K., Marugame, T., Kuriyama, S., Sobue, T., Tamakoshi, A., Satoh, H., . . . Tsugane, S. (2007). Decrease in risk of lung cancer death in Japanese men after smoking cessation by age at quitting: Pooled analysis of three large-scale cohort studies. *Cancer Science,* 98(4), 584-589. doi:10.1111/j.1349-7006.2007.00423.x

Wannamethee, S. G., Shaper, A. G., Whincup, P. H., & Walker, M. (1995). Smoking cessation and the risk of stroke in middle-aged men. *JAMA*, 274(2), 155-160.

World Health Organization. (2012). *WHO global report on mortality attributable to tobacco.*

Wu, F., Chen, Y., Parvez, F., Segers, S., Argos, M., Islam, T., . . . Ahsan, H. (2013). A prospective study of tobacco smoking and mortality in Bangladesh. *PLoS One,* 8(3), 11. doi:10.1371/journal.pone.0058516