

# Planning Tool Guidance

*Project Big Life*

*2019-09-26*



# Contents

<b>1</b>	<b>Welcome to Project Big Life’s Planning Tool</b>	<b>5</b>
<b>2</b>	<b>Introduction</b>	<b>7</b>
<b>3</b>	<b>Getting Started</b>	<b>9</b>
<b>4</b>	<b>How To</b>	<b>11</b>
4.1	Customize Data . . . . .	11
4.2	Load data . . . . .	13
4.3	Select measure (calculation) . . . . .	14
4.4	Filter data . . . . .	14
4.5	Stratify data . . . . .	15
4.6	Run scenarios . . . . .	16
4.7	Calculate results . . . . .	17
4.8	View Results . . . . .	17
4.9	Download results . . . . .	18
4.10	Resolve warning or error messages (To be developed) . . . . .	18
<b>5</b>	<b>Case examples</b>	<b>19</b>
5.1	Health Status Report . . . . .	19
5.2	Adjusting gears . . . . .	21
5.3	Healthy Cities . . . . .	22
5.4	Canada bikes like the Dutch . . . . .	24
5.5	Smoking by-law, 2004 Ottawa: retrospective . . . . .	26
<b>6</b>	<b>Key Concepts</b>	<b>29</b>
6.1	Data and sample files . . . . .	29
6.2	Multivariable predictive risk algorithms . . . . .	30
6.3	Calculations . . . . .	32
6.4	Calculations: specific health outcomes . . . . .	32

6.5	Scenario: Intervention . . . . .	34
6.6	Scenario: Health behaviour attribution . . . . .	35
6.7	Results . . . . .	39
6.8	Assumptions and Limitations . . . . .	40
<b>7</b>	<b>Glossary</b>	<b>43</b>
<b>A</b>	<b>Mortality Population Risk Tool (MPoRT)</b>	<b>47</b>
A.1	Data requirements for MPoRT . . . . .	48
A.2	Descriptions of MPoRT versions . . . . .	51
<b>B</b>	<b>Health behaviour attributable calculations</b>	<b>55</b>
B.1	Calculation . . . . .	55
B.2	Additional considerations . . . . .	57
<b>C</b>	<b>Bibilography</b>	<b>59</b>

# Chapter 1

## Welcome to Project Big Life's Planning Tool

### 1.0.1 What is the Project Big Life Planning Tool

The Project Big Life Team received numerous requests to generate, replicate, and updated health status reports such as Public Health Ontario's Seven More Years which examined the impact of health behaviours on health outcomes: mortality and life expectancy.

To address these requests and better support public health professionals the Project Big Life Team decided to create Project Big Life Planning Tool.

The Project Big Life Planning Tool gives public health professionals access to the algorithms used by the Project Big Life Team, and can be used to support your research, planning, development, and evaluation of health interventions.

The tool can help you:

- Determine who in the population is at risk
- Predict number of new cases of the disease for a population
- Determine how much a risk factor contributes to a health outcome
- Assess health benefits of potential interventions

### 1.0.2 Who made the Project Big Life Planning Tool?

The Project Big Life Planning Tool was developed by the Project Big Life Team based at the Ottawa Hospital and the Bruyère Research Institute.

The planning tool was made in collaboration with Public Health Ontario, the Population Health Analytics Lab at the University of Toronto and Ottawa Public Health.

Algorithms used in the planning tool were developed using linked population data housed at ICES and validated at Statistics Canada. All algorithms have been published in peer-review journals. To learn more about the specific algorithms used in the Project Big Life Planning Tool see our appendices to the specific algorithms or Project Big Life website.



# Chapter 2

## Introduction

The Project Big Life Planning Tool was developed in order to support health professionals: research, plan, develop, and evaluate evidence-based health interventions.

For instance Project Big Life Planning Tool helps:

- Public health professionals: assess the impact of interventions that target health behaviours
- Health planners: assess the need for palliative care (*To be developed*)

### What types of questions can it answer?

The Project Big Life Planning Tool can answer the following types of questions:

- What is the burden of smoking on life expectancy?
- How many deaths would be prevented if everyone increased their daily exercise by 10%?

### How does it work?

- This tool provides health planners with access to multivariable predictive risk algorithms, created and housed by the Project Big Life Team.
- The multivariable predictive risk algorithms use distinct characteristics and health profiles of groups of people to assess the risk of a health outcome (e.g. deaths).
- The multivariable predictive risk algorithms are developed and validated using data routinely collected by Statistics Canada and provincial health agencies, and the algorithms have been published in various journals.
- More information about multivariable predictive risk algorithms can be found in the key concepts or appendices.

### Why should I use it?

- It is **easy** and **flexible** to use.
  - The user only needs to upload their data and choose which calculation to run.
  - It can be used to assess the future risk of a health outcome.
  - It can be used to assess the effectiveness of different intervention scenarios (e.g. policy) on a health outcome.

- It generates **accurate** predictions.
  - It can be used to accurately assess the risk of a health outcome in populations that were not used in its development, and groups of people that account for only a fraction of the population.
- It is **private**.
  - Loaded data remains on your computer and is not uploaded or sent anywhere.



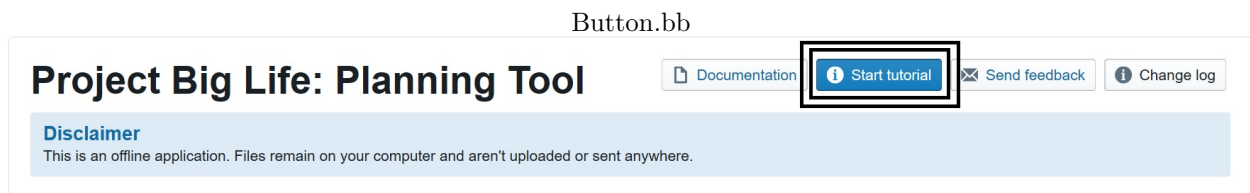
## Chapter 3

# Getting Started

To help you get started quickly with Project Big Life's Planning Tool, we built a Tutorial directly onto the platform.

The tutorial takes you through step-by-step how to use Project Big Life's Planning Tool. The tutorial will not explain the steps in detail) nor will it provide terminology glossary or how the tool operates, but it will give you an understanding of how easy it is to use the Planning Tool!

**To access the tutorial**, go onto Project Big Life's Planning Tool and click on the Tutorial button in the top right corner!





# Chapter 4

## How To

These guides will cover the topics covered in the tutorial but in greater detail.

- Customize data
- Load data
- Select measure (calculation)
- Filter data
- Stratify data
- Run scenarios: Intervention and Health behaviour attribution
- Calculate results
- Download results (To be developed)
- Resolve error messages (To be developed)

### 4.1 Customize Data

**Note:** The Project Big Life Planning Tool can only be used with the 2013/2014 Canadian Community Health Survey (CCHS) Public Use Microdata File (PUMF) or Shared versions in ‘.csv’ format.

Other population health surveys can be used only if the required variables are transformed to the specifications of the 2013/2014 CCHS data. Please note that the Project Big Life team is in the process of supporting CCHS versions from 2001 to 2013.

Prior to using the Project Big Life Planning Tool you may want to transform your data set. Reasons include: customized filter(s) and/or customized stratification(s), and transforming variables from other population health surveys to the 2013/2014 Canadian Community Health Survey (CCHS) format. Any variable that you create can be selected within the planning tool in the filter and/or stratify options.

Data manipulation can occur on any programming software: R, SAS, STATA, etc, as long as you output your data set as a ‘.csv’ file.

#### 4.1.1 Customize filter and/or stratification

An example of customizing your data set is converting the variable: Body Mass Index (CCHS 2013/2014 variable HWTGBMI) from a continuous variable into four distinct categories:

- Underweight: BMI less than 18.5
- Normal or Healthy Weight: BMI of 18.5 to 24.9

- Overweight: BMI of 25.0 to 29.9
- Obese: BMI greater or equal to 30.0

### Steps

The following steps show the R code that would be used to create these strata:

1. Convert observations “Not stated” from 999.99 to NA

```
data[data == 999.99] <- NA
```

2. Load the R package dplyr. This package is used for data manipulation.

```
library(dplyr)
```

3. Create a new column that contains four categories for BMI

```
data$newcolumn <- cut(data$HWTGBMI, breaks = c(0,18.5,25,30,Inf), labels=c("Underweight", "Healthy", "Overweight", "Obese"))
```

4. The output will be your data set + a new column with the corresponding category (“Underweight”, “Healthy”, “Overweight”, “Obese”) for that individual.

	HWTGBMI	newcolumn
1	22.68	Healthy
2	26.99	Overweight
3	NA	<NA>
4	34.44	Obese
5	23.77	Healthy
6	17.23	Underweight

This new column can now be used with the Project Big Life Planning Tool for the purpose of filtering or stratification.

### 4.1.2 Transforming variables to 2013/2014 CCHS format

Population health surveys other than the 2013/2014 CCHS PUMF or Shared files can be used if the variables: sex, age and survey weight (required variables) are transformed to the 2013/2014 CCHS format.

For example the 2005/2006 CCHS data the age variable is titled DHHEGAGE and in the 2013/2014 CCHS data it is DHHGAGE. Although the variable’s name changes the data is captured identically in the 2005/2006 and 2013/2014 surveys. Therefore you can use the 2005/2006 CCHS data if you transform this variable to the 2013/2014 format.

```
library(dplyr)
```

```
CCHSdata.2005.2006 <- CCHSdata.2005.2006 %>%
  rename(DHHGAGE = DHHEGAGE)
```

All required variables **must** be transformed to the 2013/2014 CCHS format. It is also recommended that you transform the recommended variables as well to increase the accuracy of the prediction(s).

When transforming required or recommended variables pay attention to how the variables are captured. There may be differences limit the ability to use the data set or that will affect your results. For instance from the 2001-2004 CCHS survey there are 15 age categories but from 2005-2013 there are 16 age categories.

## 4.2 Load data

**Note:** The Project Big Life Planning Tool can currently only support .csv data files from the 2013/2014 Canadian Community Health Survey (CCHS). Both the 2013/2014 CCHS PUMF and Shared file are accepted.

Data loaded to the Project Big Life Planning Tool:

- **Must include all the required** variables: DDH\_SEX (sex), DHHGAGE (age), and WTS\_M (survey weight).
- **Should include** some or all of the **recommended** variables. Recommended variables aren't required for calculations but will increase the accuracy of the predicted health outcome.

More information on the required and recommended variables can be found Appendix.

Only **one data file** can be used on the planning tool at a time; the planning tool cannot perform a calculation across multiple data files. If you wish to use data from different data sets, you first merge the data sets together and use the 2013/2014 CCHS format (e.g. variable names). Once data sets are merged into one data file you can load your data on to the Planning Tool.

**Note:** Data you load to the Project Big Life Planning Tool remains on your computer and is not uploaded or sent anywhere.

There are two options of data files: a sample file or your own file.

### 4.2.1 Load sample files

If you don't have your own data or want to explore the platform's capabilities before using your data, you can use the sample file on the Project Big Life Planning Tool. There are two sample files included on the planning tool:

- **sample.csv (Sample)** this is fabricated data set that includes all variables required for calculation, recommend for calculation, and needed for the case examples. This data set is based on the 2013/2014 CCHS PUMF but it is **entirely fabricated and can only be used for exemplary purposes**.
- **sample\_small.csv (Sample)** is the same data set as above but has fewer observations. Use this data set if you are testing out the platform as it will take less time to compute.

#### Steps

1. Click on the checkmark beside the "Sample file" to select it.

### 4.2.2 Load your own file

#### Steps

1. Click the **Browse** button under "Select a file to use in calculations".
2. Locate the file on your computer, select, and open.
3. Choose the file type (2013/2014 PUMF or 2013/2014 Shared) for your uploaded file, in the table that showcases all of the files.

Note:

- If the loaded file has all of the variables required and recommended for calculation, you will be able to continue with the planning tool.
- If the loaded file does **not** have all the variables **required** for the calculation you will not be able to continue with the planning tool.
- If the loaded file does **not** have all the variables **recommended** for calculation you will be able to continue with the planning tool, however the calculations may be less accurate.

### 4.3 Select measure (calculation)

There are two general types of calculations: summary measures and by row measures.

- **Summary measures:** When selected, the result will be a single measure for the entire dataset. For instance when Summary Measure - Life Expectancy (Summary Measure) is selected the result is a single life expectancy at birth for the given for the population.
- **By row measures:** When selected, the result will be the measurement for each individual (e.g. row) in the dataset.

Summary measures must be selected for calculations that have stratifications, intervention scenarios, and health behaviour attribution scenario.

#### Steps

1. Check the box beside the calculation's name under "Calculations" to select it. A single calculation or multiple calculations may be selected.
- Once selected the name(s) of the calculation(s) will appear to the right of "Calculations".

More details of what the calculations are and how they are performed can be found in key concepts - "Calculations".

### 4.4 Filter data

Use filters when you want to analyze only a subset of your data.

#### Steps

1. Click on the + **New** button under "Filters".
2. Select the variable that you want to filter on by typing its variable name into the "search variables" text bar.
3. Filter **in** the categories/levels within the variable by:
  - **Categorical:** Clicking on the "Search categories" text bar. Scroll and select the category you want to **keep** in your data. Repeat this step to add additional categories.
  - **Continuous:** Click the "cycle button" found under the variable you have selected. Two new boxes will appear.

- Select the minimum value for your subset data in the box on the left by typing the values or using the arrows.
  - Select the maximum value for your subset data in the box on your right by typing the value or using the arrows.
4. To add another filter repeat the steps above. A maximum of three filters are recommended to maintain statistical power (added filters reduce sample sizes and reduces statistical power).
- Once selected, the name(s) of the filtered variable(s) will appear to the right of “Filters”.

You are able to filter on all types of variables: required for calculation, recommended for calculation, and ignore variables (includes customized variables).

#### 4.4.1 Remove a filter

- To remove a filter entirely, click on the trash can beside the variable you want to delete.
- To remove a level within a filtered variable - categorical, click on the ‘x’ beside the variable level.

## 4.5 Stratify data

Use stratifications when you want to get a result for multiple strata (levels or classes).

A summary measure must be selected for stratifications, as only a summary measure will be outputted for each strata. By row measurements may also be selected but they will not be stratified.

### Steps

1. Click the box beside either “Life Expectancy (Summary)” or “Deaths (Five Years)” under the “Calculation” drop down.
  2. Select the variables you want to stratify on under the “Stratifications”. You are only able to stratify on categorical variables.
  3. To add another variable for stratification repeat the steps above. A maximum of 3 stratifications are recommended to maintain statistical power (added strata reduce strata sample size and reduces statistical power).
- Once selected, the name(s) of the stratified variable(s) will appear to the right of “Stratifications”.

You are able to stratify on all types of categorical variables: required for calculation, recommended for calculation, and ignore variables (includes customized variables).

#### 4.5.1 Remove a stratification

- To remove a stratification variable click on the ‘x’ beside the variable level.

## 4.6 Run scenarios

Scenarios can be used to predict the health outcomes when unhealthy behaviours:

- are modified in the population: **Intervention**, or
- were never present in the population: **health behaviour attribution**

Scenarios can be used to inform potential health policies or programs.

### 4.6.1 Intervention scenarios

Interventions provide you with the ability to customize the scenarios. For example you can answer the questions:

- what if we only had 15% of the population smoked rather than the current 20%?
- what if everyone increased their physical activity by 10%?
- what if everyone ate 4 fruit servings each day?
- what if everyone drank 2 fewer drinks per week?

These intervention scenarios allow you to predict and compare the effectiveness of policies.

There are 3 types of intervention scenarios that you can select:

- **Absolute:** each individual in the population **changes** their health behaviour **by a value of x**.
- **Relative:** each individual in the population **changes** their health behaviour **by a ratio of y**.
- **Target:** each individual in the population **has a set value of z**.

More information on the specifics of each type of intervention scenario and how they are calculated can be found in key concepts - “Scenarios: Interventions”.

#### Steps

1. Click the box beside either “Life Expectancy (Summary)” under the “Calculation” drop down.
2. Check the button beside: “Intervention” under the “Scenario” drop down.
3. Click on the health behaviour you want to modify: e.g. Diet.
  - A drop down menu of all the possible variables of that health behaviour you can modify will appear.
4. Click the box beside the variable you want to modify: e.g. Daily consumption - fruit - (D).
  - A drop down menu of the scenario types: absolute, relative, and target, will appear.
5. Check the button beside the type of intervention you want to modify: e.g. Target.
6. Use the arrows beside the text box: “Decrease by” or “Increase by” to add the value you are modifying: e.g. The value of 4 is used for the scenario what if everyone ate 4 fruit servings each day?

Multiple health behaviours and variables within the health behaviours can be selected for a single calculation.

- Once selected, the name(s) of the health behaviour(s) that have been selected for the intervention will appear to the right of “Scenario”.



### 4.6.2 Health behaviour attribution scenarios

Health behaviour attribution scenarios provide you with the ability to see the best case scenario for the population. For example:

- what if no one in the population ever smoked?
- what if everyone in the population met their recommended physical activity levels (3.00 METs/week)?

More information on health behaviour attribution calculations and how they are calculated can be found in key concepts[[#keyconcepts](#)] - “Scenarios: Health behaviour attribution”.

#### Steps

1. Click the box beside either “Life Expectancy (Summary)” under the “Calculation” drop down.
2. Check the button beside: “Health behaviour attribution” under the “Scenario” drop down.
3. Check the box beside the health behaviour that you want to have a health behaviour attribution calculation.

Multiple health behaviours can be selected for a single calculation.

## 4.7 Calculate results

1. Name your calculation in the text box: Calculation name.
  - Be specific when naming the calculation as it will make it easier to distinguish after running multiple calculations.

**Note:** the larger the data set is the longer the calculations will take. Depending on the size of the data set and the type of calculation being preformed it could take an hour or more.

2. Click the “calculate” button.
3. A new table will appear under the calculate button. The table includes the status of the calculation, the calculation: name, measures, filters and stratifications that were select.

A green check mark will appear in the status column.

## 4.8 View Results

1. Click on the calculation name, in the table under the green calculation button.
2. There are three tabs for the results:
  - **Overview:** provides a summary of the calculation that was preformed including: measure, filters used, errors, and warnings.
  - **Stratifications:** This will show the results for the measurement you choose to calculate. This screen will either have:
    - **a table** when you stratifications are selected.
    - **a graph** when stratification are selected. The graph is further described below.
  - **Log:** will give you a summary of the warnings and errors that occurred during the calculation.

### 4.8.1 Graph

The graph produce through plotly is interactive. Hover over the graph to see the results for each strata, exclude strata, download a .png of the graph, and much more.

## 4.9 Download results

Click on the **Download results** button under the **Results** section.

A pop-up will appear.

Select which measures you would like to download.

Select which calculations you'd like to download.

Scroll down in the pop up window and select "Download File: Your\_file\_and\_calculation\_name".

Two files will be downloaded.

- **\*\*Text file:\*\*** will give you a summary of the calculation preformed.
- **\*\*CSV file:\*\*** will provide you a table with the measures you preformed.

## 4.10 Resolve warning or error messages (To be developed)

There are different types of warning and error messages that may appear.

Below describes some of the messages that are likely to occur and steps to resolve them.

### 4.10.1 Invalid category

### 4.10.2 Out of range

### 4.10.3 Not a number

### 4.10.4 Sample size is too small

# Chapter 5

## Case examples

This chapter provides you with examples of how Project Big Life's Planning Tool can be used in your day-to-day operations. The case examples will cover:

Description

Name

Type

Statistics for health status reports

Health status report

Filter, Stratification, Scenario - Health behaviour attributable (smoking and diet)

Deaths prevented from increasing cycling infrastructure

Adjusting gears

Filter, Scenario - Intervention: Absolute (Physical activity)

Impact on life expectancy by modifying all health behaviours

Healthy cities

Filter, Stratification, Scenario - Intervention: Target (All)

Impact on life expectancy by increasing cycling rates

Canada bikes like the Dutch

Filter, Scenario, Intervention: Absolute (Physical activity)

Prevented deaths from enacting a 2004 smoking by-law

Smoking by-law, Ottawa 2004: retrospective

Transform variables, Filter, Scenario - Intervention: Target (Smoking)

### 5.1 Health Status Report

Health status reports are a way to report the health state for a population and the factors that influence a population's health. Information from health status reports are used to inform policy, planning, and resource allocation.

Using the Project Big Life Planning Tool you can replicate or update existing health status reports e.g. Public Health Ontario's: Seven More Years Report or generate a new health status report.

In this example we will show you how to create some health statistics to include in your health report. We will calculate:

- The predicted number of deaths by strata (e.g. by sex [males and females])
- The impact of eliminating unhealthy behaviours on life expectancy

For this example we will focus on the population of Alberta.

### 5.1.1 Predicted number of deaths stratified by sex and level of education

By showing the number of deaths by strata, the reader can see the distribution of deaths across specific population characteristics. Any categorical variable can be used for stratification but in this example we will use sex and level of education.

#### Steps

1. Select the sample file data.sample.csv under "Sample files".

**Note:** Although the data.sample.csv is based on the 2013/2014 Canadian Community Health Survey Public Use Microdata File, data.sample.csv is a completely synthetic data set and can only be used for exemplary purposes.

2. Select the calculation: Summary Measure – Deaths (Five years)
3. Add filter: GEOGPRV – 48, which is the corresponding code for Alberta.
4. Add two stratification: DDH\_SEX and EDUDR04
5. Title the calculation: Deaths by sex and education level
6. Click the calculate button

7. *To do: Results – walk through the results*

### 5.1.2 Impact of eliminating unhealthy behaviours: smoking and poor diet, on life expectancy

To show how much an unhealthy behaviour impacts life expectancy we use the scenario: health behaviour attributable. Health behaviour attributable scenarios can be used for the health behaviours: alcohol consumption, diet, physical activity, and smoking, individually or in any combination. In this example we will evaluate the impact of smoking and poor diet, in combination, on life expectancy.

#### Steps

1. Select the sample file data.sample.csv under "Sample files".

**Note:** Although the data.sample.csv is based on the 2013/2014 Canadian Community Health Survey Public Use Microdata File, data.sample.csv is a completely synthetic data set and can only be used for exemplary purposes.

2. Select initial calculation: Summary Measure – Life Expectancy (Summary)
3. Add filter: GEOGPRV – 48, which is the corresponding code for Alberta.
4. Click the text: Scenario.
5. Select health behaviour attribution.
6. Select the health behaviours: smoking and diet
7. Title the calculation: Alberta: smoking and diet attributable life expectancy
8. Click the calculate button

#### 9. To Do: Walk through results

**Note:** Physical inactivity and poor diet attributable life expectancy lost  $\neq$  Physical inactivity-deleted life expectancy + Poor diet-deleted life expectancy

as individuals in the population may be both physically inactive and have a poor diet.

## 5.2 Adjusting gears

The City of Hamilton has decided to implement a new initiative *Adjusting Gears* which will increase cycling infrastructure in Hamilton. As part of the plan they have decided to develop 100 km of new cycling infrastructure in the next 3 years (2019-2021).

The city has asked you to evaluate the health impact of this initiative.

As part of the analysis you want to determine how the 100km of new cycling infrastructure will impact the health of the Hamilton area, and how many deaths will be prevented.

### 5.2.1 Part A: Determine the how much a Hamiltonian will increase their physical activity by using these new cycling tracks

Assumption #1: After having done your research, you assume that with the new tracks the average person in Hamilton will cycle an additional 10 minutes/day (duration).

1. Convert the increased duration of cycling into daily energy expenditure (MET-hours):

MET-hours = Duration \* MET value for cycling

- The MET value for cycling in Canada is 4.0

### 5.2.2 Part B: Determine the impact on life expectancy

Use the Project Big Life Planning Tool for the following steps:

1. Load your data file: data.sample.csv to the Project Big Life Planning Tool.
2. Select initial calculation: Summary Measure – Deaths (5-years)
3. Add Filter: GEODPMF - 35937
4. Click: Scenarios, and select Intervention

5. Click Physical Activity then Select “Average daily leisure time energy expenditure in METs”

Assumption #2: Although the scenario for physical activity is for leisure energy expenditure, we assume 1) individuals that are active in their leisure time also use active transportation, and 2) cycling as a form of active transportation is minor part of their energy expenditure. Therefore we will adjust the average daily leisure time energy expenditure in METs.

6. Select Absolute
7. Type in the absolute change calculated in Part 1, into the text box.
8. Name your calculation: Adjusting gears - Deaths in Hamilton
9. Click Calculate

## 5.3 Healthy Cities

What would be Ottawa’s life expectancy, if the region had the same health behaviours as the healthiest region in Canada?

In this example, we will determine:

- A. who is the healthiest region in Canada,
- B. what are the health behaviours of Ottawa vs the healthiest region, and
- C. what would be the mortality outcomes be if Ottawa’s health behaviours were the same as the healthiest region.

Here we will use the average life expectancy to determine the health of regions: the higher the life expectancy for the region the healthier that region is.

For this example we will use the sample data set **data.sample.csv**, which can be downloaded at <https://github.com/Big-Life-Lab/PBL-Planning-Tool-Case-Examples>. This link also contains all of the R code for this case example. Although we have included the R code any other statistics software (SAS, STATA, etc) can be used to complete this case example.

The following steps include the R coding for each of the steps but you can preform the steps in any other software program.

### 5.3.1 Part A: Find the healthiest region in Canada

Using the Project Bid Life Planning Tool, we will find the healthiest region in Canada: the region with the highest life expectancy.

1. Load the data file: data.sample.csv.

**Note:** Although the data.sample.csv is based on the 2013/2014 Canadian Community Health Survey Public Use Microdata File, data.sample.csv is a completely fabricated data set and can only be used for exemplary purposes.

2. Select calculation: Summary Measure – Life Expectancy (Summary)
3. Add Stratification: GEODPMF. This variable represents the health regions in Canada
4. Title the calculation: Life expectancy by health region

5. Click the calculate button
6. Download the results - *TO DO*

Using your preferred statistic software (Excel, SAS, STATA, R Studio, etc) identify the healthiest region and the health of the Ottawa health region.

The healthiest region is XXX.

### 5.3.2 Part B: Determine the health behaviours of the healthiest region and the Ottawa region.

Calculate the average of each health behaviour in the healthiest region: XXX and the Ottawa region.

- **Smoking:** Calculate the prevalence of current smokers in the XXX and Ottawa region.
- **Physical Activity:** Calculate the average of physical activity (variable PACDEE) in the XXX and Ottawa region.
- **Alcohol:** Calculate the average of weekly alcohol consumption (variable ALWDWKY) in the XXX and Ottawa region.
- **Diet:** Calculate the average daily consumption of each diet component (variables: FVCD CAR, FVCD-FRU, FVCDJUI, FVCDPOT, FVCD SAL, and FVCD VEG) in the XXX and Ottawa region.
- The more daily servings an individual has of:
  - carrots, fruits, salad, and vegetables = the more healthy the individual,
  - juice = the less healthy the individual,
  - potato < 1 = the more healthy the individual, and
  - potato > 1 = the less healthy the individual.

#### Comparing health behaviours

Compare the averages of all health behaviours in the Ottawa region to the XXX. When the health behaviour for Ottawa is unhealthier than XXX, record the avg/prevalence of that health behaviour for XXX.

### 5.3.3 Part C: What if Ottawa acted like XXX?

We'll now run a scenario where the Ottawa region acts like XXX. Ottawa will now have the health behaviours of XXX, when the health behaviour of XXX is healthier than Ottawa's.

1. Load your data file: data.sample.csv
2. Select initial calculation: Summary Measure – Life Expectancy (Summary)
3. Add Filter: GEODPMF – 35951 (the code for the Ottawa Health region)
4. Click: Scenarios
5. Click on each health behaviour where Ottawa is unhealthier than XXX and input XXX's value in the target scenario for that health behaviour
  - *TO DO:* Examples of where that is
6. Title the Calculation: Ottawa - Intervention - HB1, ....
7. Click calculate.
8. *TO DO:* results

## 5.4 Canada bikes like the Dutch

Cycling as a form of transportation, is an effective way of increasing individuals' daily physical activity levels and can lead to a decrease in the risk of disease and death.

In this example we will determine the impact on life expectancy if Canadians were to cycle like the cycling enthusiasts the Dutch!

In this example we determine:

- A. How much Canadians cycle per day for transportation purposes on average?
- B. How much Dutch cycle per day for transportation purposes on average?
- C. What is the average difference in daily cycling transportation for Canadians and the Dutch?
- D. Predict how Canadian life expectancy would change if Canadians biked like the Dutch.

For this example we will use the sample data set **data.sample.csv**, which can be downloaded at <https://github.com/Big-Life-Lab/PBL-Planning-Tool-Case-Examples>. This link also contains all of the R code for this case example. Although we have included the Rcode other statistics software (SAS, STATA, etc) can be used to complete this case example.

**Note:** Although the data.sample.csv is based on the 2013/2014 Canadian Community Health Survey Public Use Microdata File, data.sample.csv is a completely synthetic data set and can only be used for exemplary purposes.

### 5.4.1 Part A: How much do Canadians cycle per day for transportation purposes on average?

We will determine how much Canadians cycle per day by calculating the average daily energy expenditure from cycling in the current Canadian population.

We will use 3 variables in our sample data set which are based off of the 2013/2014 CCHS PUMF and measure cycling as a form of active transportation:

- PAC\_8: In the past 3 months did you cycle to and from work or school?
- PAC\_8A: How many times did you cycle to and from work or school, in the past 3 months?
- PAC\_8B: How much time did you spend on each occasion?

**Assumption #1:** we assume individuals 65 or older are retired and therefore not cycling to/from work/school. Therefore, we will only evaluate individuals that are < 65 years old.

In order to calculate the average Canadian daily energy expenditure from cycling for transportation, you will need to:

- Find the average number of times, that an individual cycled to/from work/school in the past 3 months (mean PAC\_8A)
- Find the average number of hours spent cycling to/from work/school, in the past 3 month (mean of PAC\_8B)
- Calculate the average duration of cycling:  $\text{Duration} = [(\text{Frequency}/3 \text{ months}) * (\text{Time}/\text{trip})] / (\text{Days}/3 \text{ months})$
- Calculate the average daily energy expenditure (MET-hours) for cycling to/from work/school:  $\text{MET-hours} = \text{Duration} * \text{MET value for cycling}$



**Note:** For Canadians to cycle like the Dutch, we need to use the Dutch MET cycling value of 5.8 and not the Canadian MET cycling value of 4.0. The Dutch MET value accounts for the speed of cycling of the Dutch.

- Calculation the average daily expenditure from cycling for transportation for all individuals in the population not only those that cycled to/from work/school

On average a Canadian only gets 0.024 MET-hours/day from cycling to/from work/school.

#### **5.4.2 Part B: How much do the Dutch cycle/day for transportation purposes on average?**

Using data collected as part of the Dutch National Travel Survey (2010 – 2012), Fisherman, 2015 (doi: 10.1371/journal.pone.0121871 - Table S1) reported the average additional daily energy expenditure from cycling as a form of transportation for the Dutch: males: 1.3 MET-hours, and females: 1.4 MET-hours.

The average Dutch daily energy expenditure from cycling as a form of transportation for both sexes is then 1.35 MET-hours.

#### **5.4.3 Part C: What is the average difference in daily cycling transportation for Canadians and the Dutch?**

Canadians need to increase their average daily energy expenditure from cycling from 0.07 MET-hours to 1.35 MET-hours, in order to be like the Dutch.

**Note:** Since the physical activity intervention scenario on the Project Big Life Planning Tool is for daily leisure energy expenditure and not daily active transportation, we will use the ‘absolute’ scenario for this example. We will increase each individual’s daily leisure energy expenditure by 1.28 MET-hours (1.35 MET-hours - 0.07 MET-hours).

#### **5.4.4 Part D: Predict how Canadian life expectancy would change if Canadians biked like the Dutch.**

Use the Project Big Life Planning Tool for the following steps:

1. Load your data file: data.sample.csv to the Project Big Life Planning Tool.
2. Select initial calculation: Summary Measure – Life Expectancy (Summary)
3. Add Filter: DDHGAGE – 1,2,3,4,5,6,7,8,9,10,11,12
4. Click: Scenarios, and select Intervention
5. Click Physical Activity then Select “Average daily leisure time energy expenditure in METs”

Assumption #2: Although the scenario for physical activity is for leisure energy expenditure, we assume 1) individuals that are active in their leisure time also use active transportation, and 2) cycling as a form of active transportation is minor part of their energy expenditure. Therefore we will adjust the average daily leisure time energy expenditure in METs.

6. Select Absolute

7. Type in the absolute change calculated in Part 3, into the text box.
8. Name your calculation: Canada bikes like the Dutch
9. Click Calculate
10. Interpret results

## 5.5 Smoking by-law, 2004 Ottawa: retrospective

The City of Ottawa has historically been a leader in municipal policy to reduce tobacco smoking at the population level. A key milestone was the introduction of a 2004 by-law that limited smoking in bars/restaurants.

We want to know how many smoking attributable deaths were prevented due to this by-law.

Assumption: We assume that the by-law lead to decrease in current smokers either through individuals quitting smoking or individuals not starting to smoke.

For this example we will:

- A. Transform 2003 CCHS PUMF variables to 2013 CCHS PUMF variables
- B. Determine the prevalence of current smokers in 2003
- C. Determine the prevalence of current smokers in 2013
- D. Estimate how many smoking attributable deaths were prevented due to enacting the by-law

For this example we will use the sample data set **data.sample2003.csv** and **data.sample2013.csv**, which can be downloaded at <https://github.com/Big-Life-Lab/PBL-Planning-Tool-Case-Examples>. This link also contains all of the R code for this case example. Although we have included the R code any other statistics software (SAS, STATA, etc) can be used to complete this case example.

### 5.5.1 Part A: Transform 2003 CCHS PUMF variables to 2013 CCHS PUMF variables

The Project Big Life Planning Tool only works on 2013-2014 CCHS variable names only. Therefore if you are using a CCHS data set from previous years you will have to transform the variables to the 2013-2014 format.

1. Use **data.sample2003.csv**

**Note:** Although the `data.sample2003.csv` is based on the 2003/2004 Canadian Community Health Survey Public Use Microdata File, `data.sample2003.csv` is a completely fabricated data set and can only be used for exemplary purposes.

With the `data.sample2003` data set you will need to:

- Rename 2003 CCHS PUMF variables to their corresponding 2013 CCHS PUMF variables,
- Create variable `CCC_091` from variables `CCCC_91A` and `CCCC_91B`, and
- Recode `DHHGAGE` variable

Note: The Project Big Life team is developing a R package that will make this step easier for you! The package transforms CCHS variables from 2001 to 2014.

We'll update the guidance as soon as the package is ready!

### 5.5.2 Part B: Determine the prevalence of current smokers in 2003

Current smokers are defined by variable SMKDSTY: 1 – Daily, 2 – Occasional, 3 – Always Occasional.

In 2003, current smokers accounted for 26% of the Ottawa population.

### 5.5.3 Part C: Determine the prevalence of current smokers in 2013.

The prevalence of smokers in 2013 will be the scenario target in Part D.

1. Use **data.sample.csv**

**Note:** Although the data.sample.csv is based on the 2013/2014 Canadian Community Health Survey Public Use Microdata File, data.sample.csv is a completely fabricated data set and can only be used for exemplary purposes.

In 2013, current smokers accounted for 21% of the Ottawa population.

### 5.5.4 Part D: Estimate how many smoking attributable deaths were prevented due to enacting the by-law

Use the Project Big Life Planning Tool for the following steps:

1. Load your data file: data.sample2003.csv to the Project Big Life Planning Tool.
2. Select initial calculation: Summary Measure – Deaths (5 years)
3. Add Filter: GEODPMF - 35951
4. Click: Scenarios, and select Intervention
5. Click Smoking then Select “Prevalence of current smokers”
6. Select Target
7. Type in 21 (prevalence of smokers from 2013) into the text box.
8. Name your calculation: Smoking attributable deaths 2003
9. Click Calculate



# Chapter 6

## Key Concepts

This section explains some key concepts in Project Big Life’s Planning Tool. This section will explain how it works rather than how to do things.

- Data and sample files
- Multivariable predictive risk algorithms
- Calculations: general
  - Summary vs. By Row
  - Weighted vs. unweighted
- Calculations: specific health outcomes
  - Risk of health outcome
  - Number of health outcomes
  - Life expectancy
- Scenario: Intervention
- Scenario: Health behaviour attribution
- Results
  - Why are there more deaths for the healthy behaviour?
- Assumptions and limitations

### 6.1 Data and sample files

The Project Big Life Planning Tool currently accepts **2013/2014 Public Use Microdata File and Shared File of the Canadian Community Health Survey (CCHS)** in ‘csv’ format.

### 6.1.1 What is the Canadian Community Health Survey?

The CCHS is an annual cross-sectional survey preformed by Statistics Canada. The CCHS collects information related to health status, health care utilization, and health determinants for the Canadian population. Data is shared at the the sub-provincial geographic level (health region or combination of health regions).

- Details about the survey and its design can be found on Statistic Canada website (<http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&Id=144170>).
- Details and access to the Public Use Microdata file (PUMF) can be found through the Odesi website (<https://search2.odesi.ca/#/details?uri=%2Fodesi%2Fcchs-82M0013-E-2013-2014-Annual-component.xml>)

### 6.1.2 Why can the Project Big Life Planning Tool only accept the 2013/2014 Canadian Community Health Survey data?

Each year the CCHS changes a few variables it captures. This makes CCHS data sets from every year other then 2013/2014 partly incompatible with the algorithms used by the Project Big Life Planning Tool.

The Project Big Life Planning Team is currently working to adjust the algorithms so that they accept CCHS year: 2001 to 2014, and will update this guidance.

### 6.1.3 Can I use other population health survey data sets?

Yes you can use other population health survey data sets but you have to transform these data set to the 2013/2014 CCHS format.

### 6.1.4 Sample Files

- `data.sample.csv` this is synthetic data set that includes all variables required for calculation, recommend for calculation, and needed for the case examples (found in Chapter 5). While this data set was model on the response rates of the 2013/2014 CCHS PUMF (e.g. approximately 80% of respondents did not have high blood pressure (CCC\_071)), the data is fabricated. **This data set cannot be used for real analysis.**

## 6.2 Multivariable predictive risk algorithms

Multivariable predictive risk algorithms predict the future risk of health outcomes (e.g. Life Expectancy) for a population using routinely collected health data.

Multivariable predictive risk algorithms can be used to:

- Project the number of new cases of the health outcome
- Estimate the contribution of specific risk factors of the health outcome
- Evaluate effectiveness of health interventions
- Describe the distribution of risk in the population (diffused or concentrated)

Multivariable predictive risk algorithms are able to assess equity issues compared to competing population risk methods (e.g. World Health Organization Global Burden of Disease).

More information on what multivariable predictive risk algorithms are and how they can be used can be found the journal article: *Predictive risk algorithms in a population setting: an overview* (Manuel D, 2012)

### 6.2.1 Development of multivariable predictive risk algorithms

Data:

- Multivariable predictive risk algorithms are created using routinely collected data that includes information about risk factors (exposure) and health events (outcomes).
- Data is collected at an individual level through population health surveys (e.g. Canadian Community Health Survey) and administrative databases (e.g. Vital Statistics). Data sources are linked together when the individual has given permission too.
- Individuals are followed overtime until the health event (e.g. death or disease) occurs.
- Separate data is collected to create a derivation cohort and validation cohort(s).
  - Note: The risk factors that are collected are from population health surveys and are self-reported; no clinical data (e.g. blood pressure) is collected. Risk factors focus on health behaviours (e.g. smoking) and sociodemographic factors, commonly associated with health outcome.

#### Algorithm generation:

- Multivariable predictive risk algorithms are cox proportional hazard models that analyze time to health outcome (e.g. death) *Question for Carol - The models are not cox-proportional hazard models but they are similar?*
- Multivariable predictive risk algorithms are developed and validated in 4 stages:
  - Algorithm derivation: the predictive risk algorithm is created using data from the derivation cohort
  - Algorithm validation: the predictive risk algorithm is applied to the validation cohort
  - Final algorithm generation: validation and derivation cohorts are combined to estimate the final application of the predictive risk algorithm
  - Derivation of the application algorithm: creation of a parsimonious (fewer predictors) algorithm that maintained discrimination, calibration, and overall algorithm performance
- In each stage of the algorithm development and validation, algorithm performance is assessed using measures of discrimination and calibration.

### 6.2.2 Multivariable predictive risk algorithms built in Project Big Life Planning Tool

- There is currently 1 multivariable predictive risk algorithm is built into to Project Big Life planning tool.

Title

Outcomes

Information

Mortality Population Risk Tool

5 year risk of death, Life Expectancy, Cause deleted

Appendix A

## 6.3 Calculations

### 6.3.1 Summary vs By Row

There are two general types of calculations Summary Measures and By Row Measures.

Summary measures: When selected, the result will be a single measure for the entire dataset. For instance when Summary Measure - Life Expectancy (Summary Measure) is selected the result is a single life expectancy at birth for the given for the population.

By row measures: When selected, the result will be the measurement for each individual (e.g. row) in the dataset.

**Note:** Summary Measures are not the same as taking the average of By Row Measures. Summary measures account for the survey weights in their calculations. Only averaging the By Row Measures does not account for the survey weights and will result in an incorrect outcome.

### 6.3.2 Weighted vs unweighted

Weights are used in complex surveys like the Canadian Community Healthy Survey (CCHS). A weight is given to each respondent in the survey and the weight corresponds to the number of individuals in the population the respondent represents.

When a data set has weights, like the CCHS PUMF data set, the weights are used to calculate the population's outcome e.g. number of deaths.

When a survey does not have weights, the the population's outcome is not calculated with weights.

In either case, weights are not used in the calculation of an individual's outcome (e.g. an individual's 5 year risk of mortality).

## 6.4 Calculations: specific health outcomes

### 6.4.1 Risk of health outcome

Risk of the health outcome (e.g. risk of dying) is the outcome of a multivariable predictive risk algorithm. An example of the multivariable risk algorithm is:

$$\text{Risk} = \sum_t h_0(t) * e^{\beta_{pred.smoking} * x_{smoking} + \beta_{pred.cancer} * x_{cancer} + \beta_{pred.age} * x_{age} + \dots}$$

Where:

- $t$  = survival time
- $h_0(t)$  = the baseline hazard
- $\beta_{pred}$  = predictive hazard ratios for the exposures
- $x$  = the exposure. The exposure can be continuous (e.g. age) or categorical (e.g. smoking status).

Categorical exposures are represented by dummy/factor variables. Each category has its own beta and the exposure is binary. For example smoking status is categorical variable with categories: current, former  $\leq 5$  years, former  $>5$  years, or never smoked. For  $\beta_{pred.current.smoker}$  the exposure:  $x_{current.smoker} = 1$  if the individual currently smokes or 0 if the individual is another type of smoker.



### 6.4.2 Number of health outcomes

The number of health outcomes (e.g. summary - deaths) is calculated through the following steps:

1. Risk of the health outcome is calculated for each individual (row) in the data set using the multivariable predictive risk algorithm.
2. Each individual's (row) risk is weighted with their corresponding survey weight (CCHS PUMF = WTS\_M and CCHS shared file = WTS\_S).
3. The weighted mean of the health outcome (e.g. mean risk of death) is calculated.
4. The weighted mean is then multiplied with the total number of individuals in the population to generate the number of health outcomes (e.g. number of deaths in 5 years).

### 6.4.3 Life expectancy

Life expectancy is calculated using abridge life tables using a modified Chiang approach (Chiang, 1984) (Hsieh, 1991). *(link to reference for Chiang and Hsieh and also one of our papers)*. Life expectancy is calculated by two methods: one for summary life expectancy, and a second for by row life expectancy.

#### 6.4.3.1 Summary Life Expectancy

Life expectancy is calculated separately for males and females.

##### Males:

1. The mortality risk for each male individual is calculated using the male mortality multivariable predictive risk algorithm for mortality (MPoRT). Details about the MPoRT can be found in Appendix A.
2. Male individuals are grouped into the 5-year age groups that are used in the 5-year abridge life tables (e.g. 40-44 years old).
3. The weighted average risk of death for each age group is calculated.
4. A male 5-year abridge life table is created using the weighted average risks of death ( $q(x)$ ) for each age group and the median age for the age group.

##### Females

Steps 1-4 used to calculate life expectancy for males, are repeated for females using the female MPoRT and a female 5-year abridge life table.

#### Summary life expectancy

5. The summary life expectancy, or life expectancy of the entire population, is calculated by adding the male life expectancy with the female life expectancy, and taking its average.

#### Summary life expectancy by strata

Steps (1-4) are repeated for each strata. There will be strata specific weighted risk of death and strata specific life tables.

Step 5 is repeated with the average life expectancy calculated across all strata.

### 6.4.3.2 By row life expectancy

An individual's life expectancy is calculated by creating a new life table specific to that individual.

These life tables are 1-year abridge life tables, and begin at the individual's age (e.g. an individual that is 43 years old, will have the life table start at 43 years).

1. The probability of death for a person's current age is calculated using the respective MPoRT (male or female), and the individual's health profile (e.g. never smoked, 15 drinks weekly, has hypertension, is a Canadian Citizen, etc) (e.g.  $q_x$ , where  $x = 43$ ).
2. The probability of death is recalculated for incremental older ages (additional rows of the life table) up to age 90 years (  $q_{(x+1)}, q_{(x+2)} \dots q_{90}$ ). For each life table row, age is the only variable that is changed for MPoRT risk calculation.

**Note:** The result for By Row Life Expectancy are the predicted years of life that the individual still has left to live based on thier age (e.g. 43). To calculate age of death, add the respondent's age to their by row life expectancy estimate.

**Note:** By row life expectancy does not account for the individual's weight, and therefore can not be used to estimate the summary life expectancy.

## 6.5 Scenario: Intervention

How would life expectancy change if everyone increased their physical activity levels by 10%?

The health intervention scenario allows you to predict how changing the health behaviours: alcohol consumption, diet, physical activity, and smoking, of a population will affect the population health outcome (e.g. life expectancy). This feature can be used to predict the effectiveness of proposed policies or programs.

There are three types of scenarios: **absolute**, **relative**, and **target**.

- **Absolute:** each individual in the population **changes** their health behaviour **by a value of x**.
- **Relative:** each individual in the population **changes** their health behaviour **by a ratio of y**.
- **Target:** each individual in the population **has a set value of z**.

For target scenarios if the individual's value is already at the target value or beyond the target value then their value is not changed. E.g. If the target value for physical activity is 2.5 METs/day, then any individual that already has METs/day  $\geq 2.5$  METs/week then there value will not be adjusted.

The changes for each type of scenario for **alcohol**, **physical activity**, and **diet** are described in the following table:

Health.Behaviour

Absolute.change

Relative.change

Target

Each individual changes ...

Each individual changes ...

Each individual has the value ...

Alcohol Consumption

the number of drinks they have per week by x

the number of drinks they have per week by y %

z drinks per week

Physical Activity

their physical activity level by x METs per day

their physical activity level by y % METs per day

z METs per day

Diet

the number of servings of fruits and vegetables they eat by x per day

the number servings of fruits and vegetables they eat by y % per day

z fruits and vegetables per day

the number of glasses of juice they drink by x per week

the number of glasses of juice by y % per day

z glasses of juice per day

the number of servings of potatoes the eat by x per week

the number of potatoes they eat by y % per day

z potatoes per day

The **smoking** health intervention scenario is different then the other types of health intervention scenarios as they adjust the prevalence of the health behaviour.

Health.Behaviour

Absolute.change

Relative.change

Target

Smoking

The prevalence of smokers decreases by x %

The prevalence of smokers decreases by y %

The prevalence of smokers is z %

Although each type of health intervention for smoking: absolute, relative and target, changes the prevalence of current smokers they are different. The following figure shows how each is different from one another.

To adjust the prevalence of smokers, the change is applied to every current smoker in the population; individuals are not selected at random.

## 6.6 Scenario: Health behaviour attribution

What would be the life expectancy of a population if be no one in the population ever smoked? This scenario is a health behaviour attribution scenario.

A health behaviour attribution scenario:

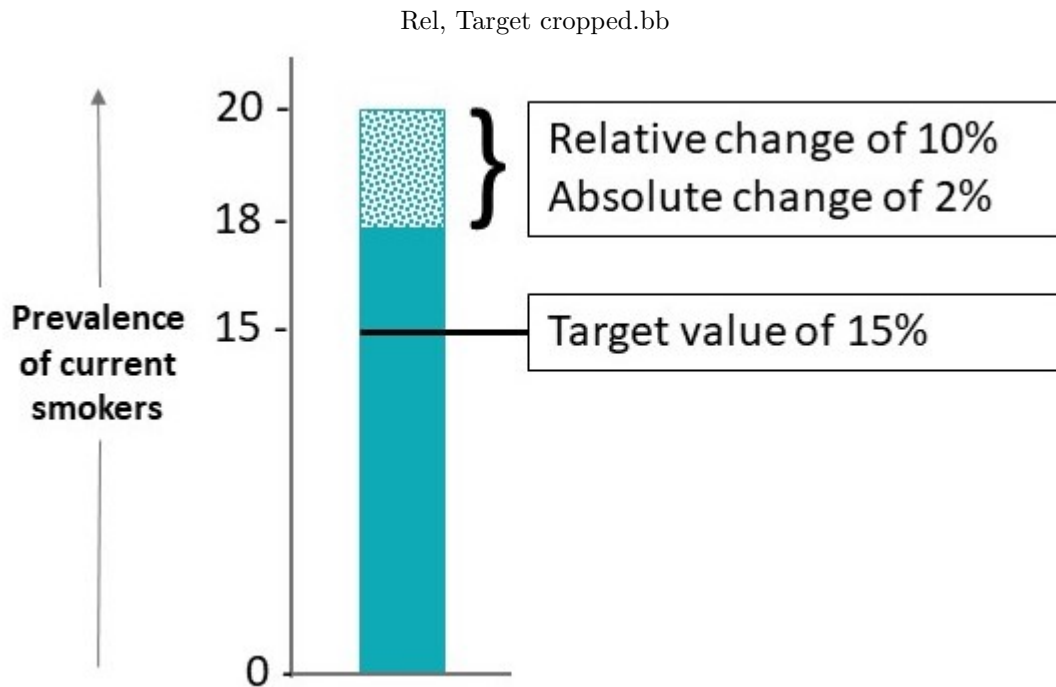


Figure 6.1: Comparison of health intervention scenario types

- predicts the health outcome if no one was exposed to the health behaviour = **health behaviour-deleted outcome**, and
- estimates the effect the health behaviour has on the health outcome for the population = **health behaviour attributable outcome**

To understand health behaviour attribution scenarios let's use the example of smoking and deaths for a population.

The number of deaths for a population is calculated twice:

- with the population that has all smokers = **baseline deaths**, and
- with a counterfactual scenario where everyone in the population is non-smokers = **smoking-deleted deaths**

The difference between the number of deaths for the two populations is the **smoking attributable deaths**.

In this example smoking attributes approximately 15,000 deaths in the population.

This can also be done with life expectancy. Life expectancy is calculated:

- with the population that has all smokers = **baseline life expectancy**, and
- with the counterfactual population where there are no smokers = **smoking-deleted life expectancy**.

The difference between the life expectancy of the two populations is the **life expectancy lost due to smoking**.

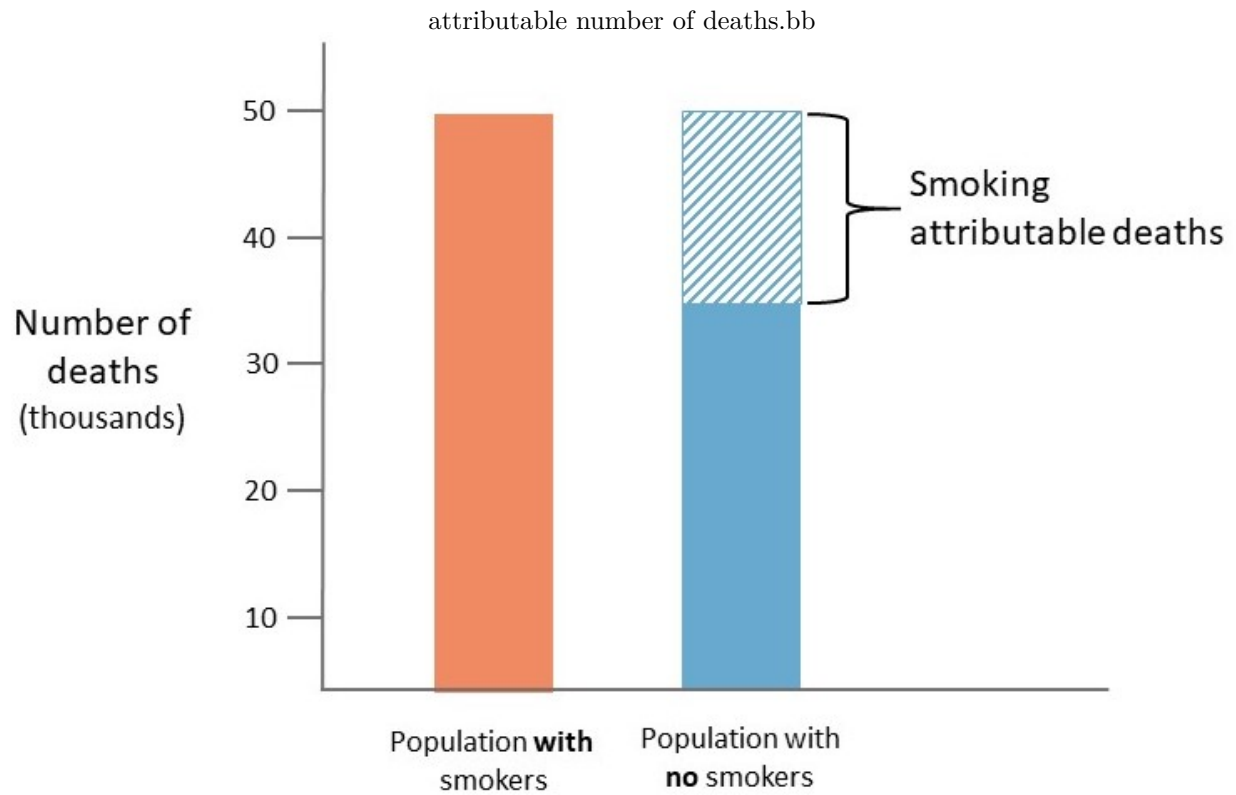


Figure 6.2: Health behaviour attributable scenario - Deaths

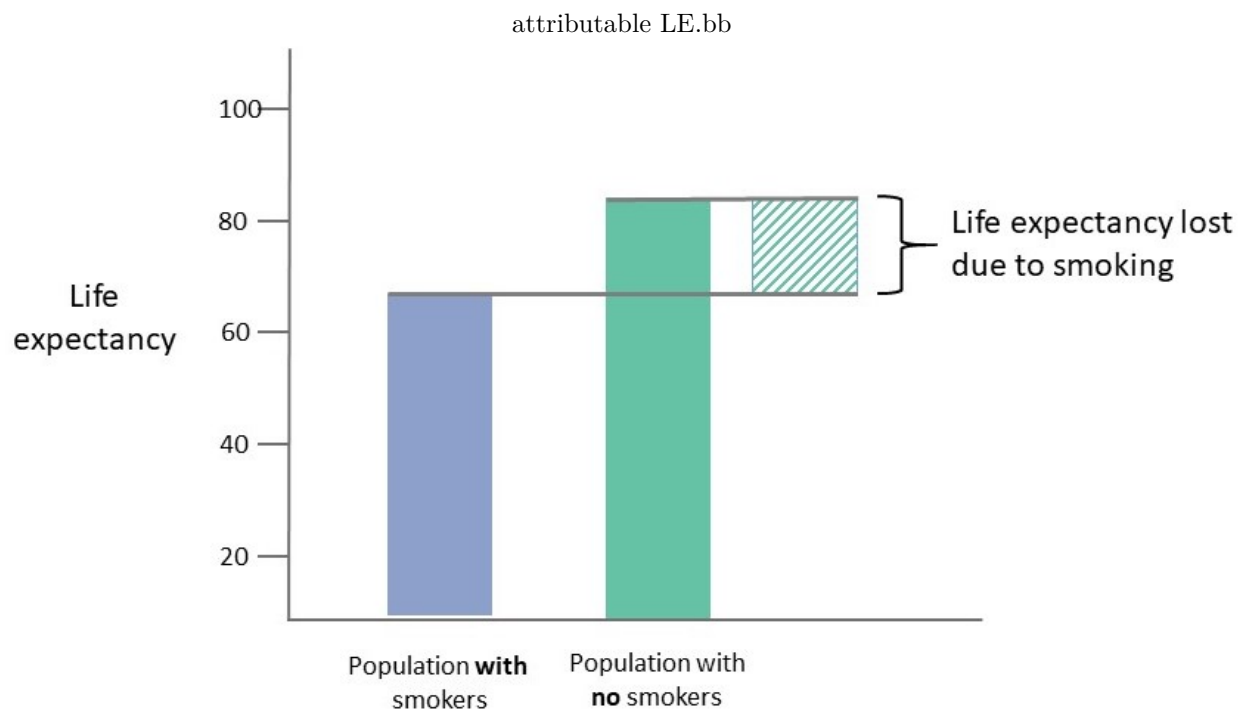


Figure 6.3: Health behaviour attributable scenario - Life expectancy

### 6.6.1 Multiple selected health behaviours

If multiple health behaviours are selected the:

- **health behaviour-deleted outcome** calculated accounts for **both** of these health behaviours, and
- **health behaviour attributable outcome** is calculated **seperately** for each health behaviour.

We will use smoking (SMK) and physical inactivity (inPA) to demonstrate what happens when multiple behaviours are selected for health behaviour attribution scenario:

Baseline

Counterfactual

Difference

HealthBehaviour1

HealthBehaviour2

Number of deaths

Deaths

SMK and inPA - deleted deaths

SMK and inPA attributable deaths

SMK attributable deaths

SMK attributable deaths

150,000

100,000

50,000

40,000

15,000

Life expectancy

Life expectancy

SMK and inPA - deleted life expectancy

Life expectancy lost due SMK and inPA

inPA attributable life expectancy lost

inPA attributable life expectancy lost

75

80

5

2

4

**Note:** SMK and inPA - deleted outcome  $\neq$  SMK attributable outcome + inPA attributable outcome as individuals in the population may be both smokers and physically inactive.

### 6.6.2 Reference health behaviours

The health behaviours (references) used to calculate the health behaviour deleted outcomes are:

Health.Behaviour

Reference

Smoking

Never smoker

Alcohol Consumption

0 drinks/week

Physical Activity

3.0 METs/day

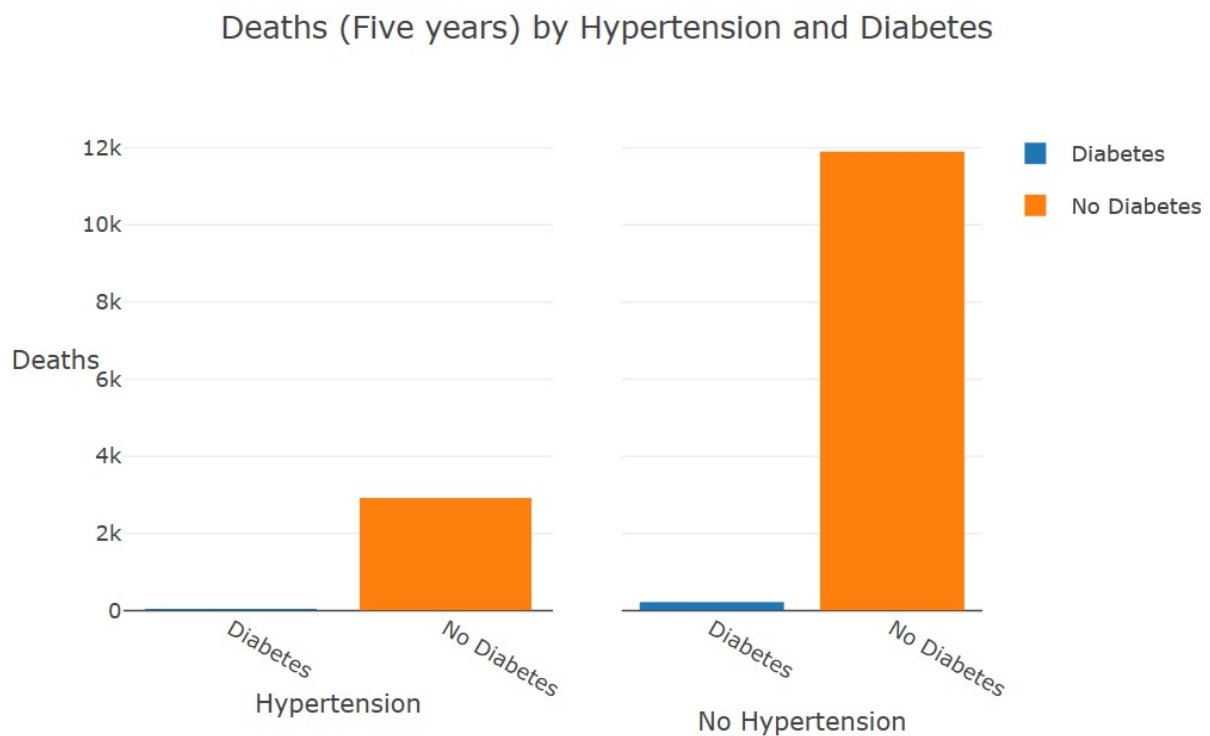
Diet

A total of: 5 fruit and vegetables, 0 juice, and 0 potato, servings/day

## 6.7 Results

### 6.7.1 Why are there more deaths for the healthy behaviour?

You may get a result where there are more deaths for the “healthier” exposure rather than an “unhealthier” exposure, for instance:



This result occurs as there are more people in the population that do not have diabetes. You can standardized your data set prior to the calculation to get a different result.

## 6.8 Assumptions and Limitations

### 6.8.1 Limitation: Data source: Canadian Community Health Survey

#### 6.8.1.1 Included population

The CCHS survey used to develop the mortality algorithm in the Project Big Life Planning Tool, include only individuals living in the community setting 20 years of age or older. Therefore individuals that live: in long-term care facilities, First Nations reserves, incarcerated populations, and full-time members of the Canadian Forces were excluded.

The exclusion of these individuals is likely to lead to a small under-estimation of the burden of a health behaviour on a health outcome. Excluded populations have higher prevalence of unhealthy behaviours like smoking: incarcerated and First Nations have higher prevalence of current smoking and individuals in long-term care facilities are more likely to be former smokers than the general population. However this under-estimation is small as the excluded population only accounts for  $< 3\%$  of the Canadian population.

#### 6.8.1.2 Reporting of health behaviours

Self-reported data is used to health surveys use self-reported data to predict risk and health outcomes. With self-reported surveys there is the possibility of social desirability bias, where respondents tend to over-report what they perceive to be healthy behaviours and under-report what they perceive to be unhealthy behaviours.

This is particularly the case with alcohol, where there is a consistent under-reporting of alcohol consumption in most population health surveys. In Ontario, the sum of self-reported alcohol consumption is about half the volume of alcohol sold {Rehm J (2006)}.

In addition the full spectrum of the behaviour has not been captured due to limited questions about the health behaviour in surveys. With diet there were few questions on the CCHS survey and these questions only asked about frequency of diet components. Information on diet factors like quantity, sodium intake, trans fat, or other healthy/unhealthy behaviours were not captured. It is likely that the burden of poor diet is under-estimated.

Although more accurate risk factor ascertainment could improve discrimination and calibration of the algorithms used in the Project Big Life Planning Tool, these algorithms already have high discrimination and calibration.

### 6.8.2 Limitation: Multivariable predictive risk algorithms

#### 6.8.2.1 Hazard ratios and exposures are assumed constant

The hazard ratios in the algorithms of the Project Big Life Planning Tool, and an individual's exposures are assumed to be constant.

Constant hazard ratios do not account for period effect which could lead to an over-estimation of the burden of a health behaviour e.g. the differential mortality between smokers and non-smokers may have been greater for older periods, Therefore the algorithms do not account for any period effects.



**6.8.3 Limitation: Calculations**

**6.8.3.1 Uncertainty**

**6.8.3.2 Attributable risk calculations**

**6.8.4 Limitation: Using the Project Big Life Planning Tool**



# Chapter 7

## Glossary

### **5-year mortality risk**

The probability that an individual will die in the next 5 years.

### **Body Mass Index (BMI)**

A weight-to-height ratio used as an indicator of obesity and underweight. BMI is calculated by dividing an individual's body weight in kilograms by the square of height in metres ( $\text{kg}/\text{m}^2$ ).

### **By Row Measures**

When selected, the result will be the measurement for each individual (e.g. row) in the dataset.

### **Calibration**

The agreement between predicted risk generated from the model and observed risk generated from the data.

### **Canadian Community Health Survey**

An annual survey performed by Statistics Canada that collects information related to health status, health care utilization and health determinants for the Canadian population. Details about the survey can be found on Statistic Canada website (<http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&Id=144170>).

### **Discrimination**

The ability of the model to differentiate between high risk individuals and low risk individuals.

### **Error Message**

Error messages will occur when variables that are “**Required for Calculation**” are missing in the data. If the entire column for the variable is missing then the calculation cannot be performed on the data. If there are missing row entries for the variable then the entire row will not be used in the calculation.

### **Filter**

Chooses part of your dataset for analysis. If you filter on ‘Sex’ and then ‘Male’, calculations will only be performed on individuals that are ‘Male’ and ‘Females’ will be excluded. For example, when calculating Life Expectancy on the filter variable ‘Sex’ then ‘Male’ there will be a Life Expectancy estimate for ‘Males’ and *no* Life Expectancy estimate for ‘Females’.

### **Health Behaviour**

Actions people do that may affect their health, positively or negatively. Health behaviours are among the determinants of health and are influenced by the social, cultural and physical environments in which people live and work.(Statistics Canada, 2010) They are also shaped by individual choices and external

constraints.(Statistics Canada, 2010) The four health behaviours of **smoking**, **alcohol consumption**, **diet**, and **physical activity** are specified in Project Big Life’s planning tool.

### **Health behaviour-deleted outcome**

The health behaviour-deleted outcome is the estimated health outcome in a counterfactual population where a specific health behaviour (e.g. smoking) never existed. For instance, everyone in the counterfactual population were are never smokers.

### **Health behaviour attributable outcome**

The health behaviour attributable outcome represents the effect the health behaviour (e.g. smoking) has on the health outcome for a population. For example, smoking attributes to 30,000 deaths in the population or 30,000 deaths in the population are caused by smoking.

### **Ignored Variables**

Are not included in the calculation. It does not matter if your dataset includes these variables or not. Ignored variables can used for filter and stratification.

### **Life Expectancy (LE)**

Life expectancy is a calculation of how long a person or population would be expected to live, on average, given unchanging risk of death from a specific point in time.

### **Metabolic Equivalent of Task (MET)**

The metabolic equivalent of task (MET) is a measure of the rate of energy expenditure from an activity; a measure of calories burned by type, duration and frequency of physical activity. The reference value of 1 MET is defined as the energy expenditure rate at rest which is equal to 1kcal/kg/day.

### **Predictor**

A variable that is used in the algorithm to predict the outcome.

### **Recommend for calculation**

Variables that are included in the calculation but not necessary for the calculation to run. Rather these variables increase the accuracy of the results.

### **Required for calculation**

Variables that are included in the calculations and are necessary for the calculation to run. If a dataset does not have these variables then the calculation will not run.

### **Risk**

The probability of a health event occurring at some point of time in the future.

### **Socioeconomic Position**

People in poorer socioeconomic circumstances generally have poorer health. Deprivation measures identify those who experience material or social disadvantage compared to others in their community. The Deprivation Index for Health in Canada developed by the Institut national de santé publique du Québec (INSPQ)(Pampalon R, Raymond G, 2000) is used in this planning tool. The index includes education, employment and income as measures of material deprivation; and single-parent families, living alone, or being divorced, widowed or separated as measures of social deprivation. The deprivation index was used to assign geographical areas into socioeconomic position groups (low, middle and high) based on material and social quintiles. High-deprivation neighbourhoods were those in the top two quintiles for both social and material deprivation. Low-deprivation neighbourhoods were those in the bottom two quintiles.

### **Stratification**

The separation of data into smaller strata (levels or classes which individuals are assigned too). If the variable ‘Sex’ is stratified it creates two strata: ‘Male’ and ‘Female’. Calculations are performed on each strata (level

or class) and the outcome will be specific to that strata. For example, when calculating Life Expectancy on the stratified variable 'Sex' there will be a Life Expectancy estimate for 'Males' and a different Life Expectancy estimate for 'Females'.

Stratification can only occur on categorical variables.

### **Summary Measures**

When selected, the result will be a single measure for the entire dataset. For instance when Summary Measure - Life Expectancy (Summary Measure) is selected the result is a single life expectancy at birth for the given for the population.

When stratifications are selected, the summary measure will be given for each strata. For instance when Summary Measure - Life Expectancy and Stratification - Sex are selected, then two life expectancy measures will be given one for males and one for females.

### **Warning Message**

Warning messages will occur when variables that are **“Recommended for Calculation”** are missing in the data. If the entire column for the variable is missing the calculation will still be performed on the data. If there are missing row entries for the variable the row will still be used in the calculation

### **Weights**

A weight is given to each respondent in the survey and the weight corresponds to the number of individuals in the population the respondent represents.



# Appendix A

## Mortality Population Risk Tool (MPoRT)

**\*Outcomes: Life Expectancy, Number of deaths, 5-year mortality risk**

### Calculations

Using MPoRT you are able to calculate:

- For each individual (By row)
- 5 year mortality risk
- Life Expectancy
- For the population (Summary)
- Number of deaths
- Life Expectancy
- Health behaviour-deleted life expectancy
- Health behaviour-deleted number of deaths
- Health behaviour attributable life expectancy
- Health behaviour attributable number of deaths

### Types of Questions

- What is the burden of smoking on life expectancy?
- How many deaths would be prevented if everyone met their daily exercise requirements?

### Description

A multivariable predictive risk model that estimates the future risk of all-cause death in Canada. It adjusts for health behaviours: smoking, unhealthy alcohol consumption, poor diet, and physical inactivity, and a wide range of other risk factors.

## A.1 Data requirements for MPoRT

In order to use the MPoRT algorithm on the Project Big Life Tool your data set **must include** the following CCHS 2013/2014 variables:

PUMF\_CCHS\_variable

Shared\_CCHS\_variable

Description

DDH\_SEX

DDH\_SEX

Sex

DHHGAGE\*

DHH\_AGE\*

Age

WTS\_M\*

WTS\_S\*

Survey weight

\* Variable names are different between the two files

In addition the following variables **should** be included to increase the prediction accuracy:

PUMF\_CCHS\_variable

Shared\_CCHS\_variable

Description

ALWDWKY

ALWDWKY

Weekly alcohol consumption

CCC\_071

CCC\_071

Do you have high blood pressure

CCC\_091

CCC\_091

Do you have chronic obstructive pulmonary disease

CCC\_101

CCC\_101

Do you have diabetes

CCC\_121

CCC\_121

Do you have heart disease

CCC\_131



CCC\_131

Do you have active cancer

CCC\_151

CCC\_151

Do you suffer from effects of stroke

CCC\_280

CCC\_280

Do you have a mood disorder

EDUDR04

EDUDR04

Highest level/education of respondent

FVCD CAR

FVCD CAR

Daily consumption - carrot

FVCD FRU

FVCD FRU

Daily consumption - fruit

FVCD JUI

FVCD JUI

Daily consumption - juice

FVCD POT

FVCD POT

Daily consumption - potatoes

FVCD SAL

FVCD SAL

Daily consumption - green salad

FVCD VEG

FVCD VEG

Daily consumption - vegetables

HWTGBMI\*

HWTDBMI\*

BMI self reported

PACDEE

PACDEE

Daily energy expenditure

SDCFIMM

SDCFIMM

Immigrant

SDCGCGT\*

SDCDCGT\*

Cultural or racial origin

SDCGRES\*

SDCDRES\*

Length of time in Canada since immigration

SMK\_01A

SMK\_01A

In lifetime, smoked 100 or more cigarettes

SMK\_05B

SMK\_05B

# of cigarettes smoked daily - occasional smoker

SMK\_05C

SMK\_05C

Past month, smoked 1+ cigarettes daily - occasional smoker

SMK\_09A

SMK\_09A

When did you stop smoking daily - occasional smoker

SMK\_204

SMK\_204

# of cigarettes smoked daily - daily smoker

SMK\_208

SMK\_208

# of cigarettes smoke each day - occasional smoker

SMKDSTY

SMKDSTY

Type of smoker

SMKG01C\*

SMK\_01C\*

Age smoked first whole cigarette

SMKG09C\*

SMK\_09C\*

Years since stopped smoking daily - occasional smoker

SMKG203\*

SMK\_203\*

Age started to smoke daily

\* Variable names are different between the two files

## A.2 Descriptions of MPoRT versions

Versions of MPoRT have been developed since 2012 and used in various studies. Each version of MPoRT (v1.0, v1.2, v2.0) used the Ontario subset of the Canadian Community Health Survey (CCHS) for development and the survey respondents were linked to personal death records. In later versions of MPoRT (v1.2, v2.0) the following changes were made: (a) algorithm variables were adjusted to improve predictions, and (b) the algorithms were validated using: the Ontario subset of CCHS of the years that were not used in development and the National CCHS data set (excluding Ontario).

### MPoRTv1.0

Was used in the Seven More Years report, a joint report with Public Health Ontario and ICES . In summary, the algorithm estimated the risk of death associated with health behaviours: smoking, unhealthy alcohol consumption, poor diet, physical inactivity and stress. There were approximately 550,000 person-years of follow up and over 6000 deaths in the development data set. The algorithm used categorical predictor variables for health behaviours and sociodemographic factors.

### MPoRTv1.2

Was published in PLoS. In summary, the algorithm estimated the risk of death associated with health behaviours: smoking, unhealthy alcohol consumption, poor diet, and physical inactivity (stress was removed due to its low prediction ability). There were approximately 1 million person-years of follow up and over 9000 deaths in the development and validation data sets. The algorithm used multiple continuous predictor variables, and added chronic disease predictor variables and interaction terms.

### MPoRTv2.0 - Version in Project Big Life's Planning Tool

This version of MPoRT has not yet been published.

*Development:* This predictive risk model was developed using Ontario subsets of the 2001 to 2008 CCHS and participants were linked to personal health records. There were approximately 1.3 million person-years of follow-up and over 15,000 deaths in the developmental data set.

*Validation:* This predictive risk model was validated using three different data sets: Ontario subset of the 2009 to 2012 CCHS, National data set (except Ontario) of the 2003 to 2008 CCHS, and the National data set of the 2000 and 2005 National Health Interview Survey in the United States of America. In all validation data sets individuals were linked to personal health records.

*Model:* Two MPoRTs have been created: one for males and one for females. Each model is a cox-proportional hazard model that looks similar to:

$$\text{Mortality risk} = \sum_t h_0(t) * e^{\beta_{pred.smoking} * x_{smoking} + \beta_{pred.cancer} * x_{cancer} + \beta_{pred.age} * x_{age} + \dots}$$

Where:

- $t$  = survival time
- $h_0(t)$  = the baseline hazard
- $\beta_{pred}$  = predictive hazard ratios for the exposures
- $x$  = the exposure. The exposure can be continuous (e.g. age) or categorical (e.g. smoking status).

*Parameters:* The parameters used in this multivariable predictive risk model are:

Category

Variable

Scale

Description

Demographic

Age\*

Continuous

5 knot spline. Valid range 20 to 102

Sex

Dichotomous

Stratified Female/Male

Health Behaviour

Pack years of smoking

Continuous

3 knot spline. Valid range: 0 to 78 (Female), 0 to 112.5 (Male)

Smoking Status

Categorical

Non-smoker

Current Smoker

Former Smoker  $\leq 5$  years

Former  $> 5$  years

Alcohol (number of drinks per week)

Continuous

4 knot spline (Females) and 3 knot spline (Males). Valid range: 0 to 25 (Female), 0 to 50 (Male)

Former/non-drinker

Dichotomous

Yes/No

Simplified diet score

Continuous

3 knot spline. Valid range: -18.9 to 20.7 (Female), -16.8 to 18.4 (Male)

Leisure physical activity (MET)

Continuous

3 knot spline. Valid range: 0 to 12.4 (Female), 0 to 16 (Male)

Socio-demographic

Ethnicity

Categorical

White

Black

Chinese

Arab; South Asian; West Asian

Filipino; Japanese; Korean; Southeast Asian

Other; Indigenous; Latin American; Multiple origin; unknown

Immigrant

Dichotomous

Yes/No

Fraction of lifetime in Canada

Continuous

3 knot spline<sup>†</sup>. Valid range: 0 to 1

Education

Categorical

Less than secondary

Secondary School Graduation

Some Post-Secondary

Post-Secondary Graduation

Neighbourhood social and material deprivation

Ordinal

Low (1st or 2nd quantile)

High (4th or 5th quantile)

Moderate (all others)

Chronic Conditions

Diabetes

Dichotomous

Yes/No

High Blood Pressure

Dichotomous

Yes/No

Chronic Respiratory Disease

Dichotomous

Yes/No

Mood Disorder

Dichotomous

Yes/No

Cancer

Dichotomous

Yes/No

Dementia

Dichotomous

Yes/No

Heart Disease

Dichotomous

Yes/No

Stroke

Dichotomous

Yes/No

Epilepsy

Dichotomous

Yes/No<sup>‡</sup>

BMI

Continuous

3 knot spline. Valid range: 8.9 to 47.2 (Female), 8.6 to 43.7 (Male)

\* Age interaction included for all variables except immigrant, fraction of time in Canada, and ethnicity <sup>†</sup> Excluded in the male model, remains in the female model <sup>‡</sup> Excluded in the female model, remains in the male model

### A.2.1 Derived parameters

In MPoRTv2 there are some parameters that are derived from variables within the data set. These include: pack-years of smoking, and the simplified diet score.

#### Pack-years of smoking

Pack-years of smoking is generated from: the type of smoker, age in which they started smoking daily, how many cigarettes they smoke daily, and if applicable the age in which they stopped smoking daily. Pack-years of smoking also includes: their age, the age of their first cigarette, and whether throughout their life they have smoked 100+ cigarettes.

In summary, the more an individual smokes or the longer they smoke, the greater the pack-years of smoking.

#### Simplified Diet Score

The simplified diet score essentially adds the healthy diet variables together (daily servings of carrots, fruit, salad, vegetables) and subtracts the unhealthy diet variables (daily servings of juice and potato).

## Appendix B

# Health behaviour attributable calculations

Here we will explain the details behind Project Big Life Planning Tool health behaviour attributable scenarios including: the calculations and additional considerations.

### B.1 Calculation

There are two parts to the calculations performed in health behaviour attributable scenarios: (A) calculate the risk, and (B) calculate the health outcome: life expectancy or number of deaths.

#### B.1.1 Part A: Risk calculations

The original multivariable predictive risk algorithm is:

$$\text{Risk} = \sum_t h_0(t) * e^{\beta_{pred.smoking} * x_{smoking} + \beta_{pred.cancer} * x_{cancer} + \beta_{pred.age} * x_{age} + \dots}$$

**Step 1.** Modify the original algorithm to include the external coefficient(s). This means replacing all predictive hazard ratios/betas related to the health behaviour to the causal hazard ratios/betas.

- Remove the original regression coefficient(s) for the health behaviour.
- Add the new external coefficient(s) to the algorithm. External coefficients are generated from either: causal models, or from systematic reviews or meta-analysis.

$$\text{External coefficient risk} = \sum_t h_0(t) * e^{\beta_{causal.smoking} * x_{smoking} + \beta_{causal.cancer} * x_{cancer} + \beta_{pred.age} * x_{age} + \dots}$$

**Step 2.** Risk is calculated using the modified algorithm created in Step 1 and the respondent's original profile (e.g. current smoker). This is the “external coefficient risk”.

$$\text{External coefficient risk} = \sum_t h_0(t) * e^{\beta_{causal.smoking} * (\text{current smoker}) + \beta_{causal.cancer} * x_{cancer} + \beta_{pred.age} * x_{age} + \dots}$$

only -cbf.bb

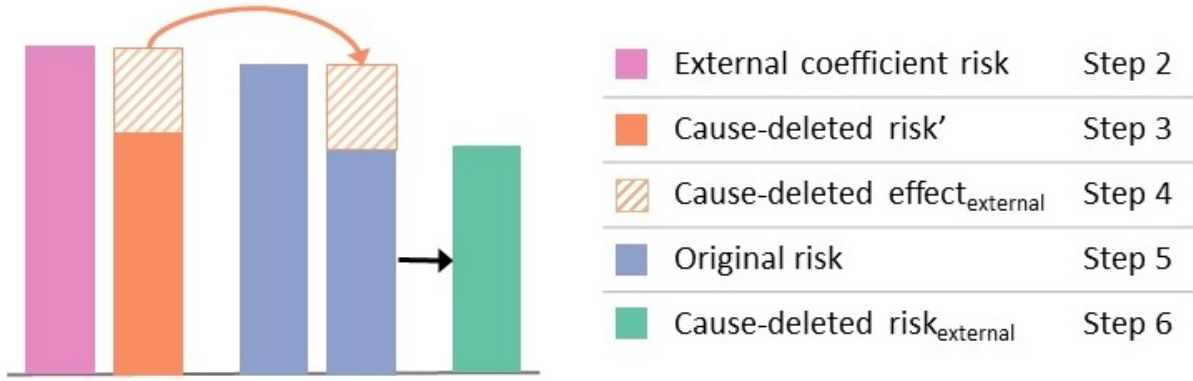


Figure B.1: Risk portion of the health behaviour attribution calculations

**Step 3.** “Health behaviour-deleted risk” is calculated by setting an exposure to a reference (non-exposed) value (all other risk exposures remain unchanged).

$$\text{Health behaviour-deleted risk}' = \sum_t h_0(t) * e^{\beta_{\text{causal.smoking}} * (\text{never smoker}) + \beta_{\text{causal.cancer}} * x_{\text{cancer}} + \beta_{\text{pred.age}} * x_{\text{age}} + \dots}$$

**Step 4.** The “health behaviour attributable risk external” is calculated as “external coefficient risk” (Step 2) minus the “health behaviour-deleted risk” (Step 3).

$$\text{Health behaviour attributable risk}_{\text{external}} = \text{External coefficient risk} - \text{Health behaviour-deleted risk}'$$

**Step 5.** Original risk is calculated using the original algorithm and the original respondent’s profile.

$$\text{Original risk} = \sum_t h_0(t) * e^{\beta_{\text{pred.smoking}} * (\text{current smoker}) + \beta_{\text{pred.cancer}} * x_{\text{cancer}} + \beta_{\text{pred.age}} * x_{\text{age}} + \dots}$$

**Step 6.** The “health behaviour-deleted risk external” is calculated by “original risk” (Step 5) minus the “health behaviour-deleted effect external” (Step 4).

$$\text{Health behaviour-deleted risk}_{\text{external}} = \text{Original risk} - \text{Health behaviour attributable risk}_{\text{external}}$$

### B.1.2 Part B: Health outcome calculations

Using risks generated above you can then calculate:

- Health behaviour-deleted life expectancy or health behaviour attributable life expectancy lost
- Health behaviour-deleted number of deaths or health behaviour attributable number of deaths

#### Life expectancy calculations

Step I: Calculate the original life expectancy by using the original risk (Step 5 above) in the sex-specific 5-year abridge period life tables.



Step II: Calculate the health behaviour-deleted life expectancy by using the health behaviour-deleted risk external (Step 6 above) in the sex-specific 5-year abridge period life tables.

Step III: Calculate the health behaviour attributable life expectancy lost by: original life expectancy (Step I) minus health behaviour-deleted life expectancy (Step II):

Health behaviour attributable life expectancy lost = Health behaviour-deleted life expectancy – Original life expectancy

### Number of deaths calculations

Step I: Calculate the number of deaths that would occur using the original risk (Step 5 above).

Step II: Calculate the number of deaths that would occur using the health behaviour-deleted risk external (Step 6 above).

Step III: Calculate the health behaviour attributable deaths by: original number of deaths (Step I) minus health behaviour-deleted number of deaths (Step II):

Health behaviour attributable deaths = Original number of deaths – Health behaviour-deleted number of deaths

## B.2 Additional considerations

### B.2.1 Health behaviours with non-monotonic relationships

Health behaviours with relationships that are non-monotonic (always increasing or always decreasing) can be used in health behaviour attributable calculations but special consideration may be warranted for both policy and analytic reasons.

An example of a health behaviour with a non-monotonic relationship is alcohol. Some studies suggest that there is a “J” shaped risk relationship for outcomes such as cardiovascular disease.

*Is this still a consideration if we are using 0 as the reference value? Need to complete the thought at the end of the paragraph* Alcohol drinking guidelines usually do not recommend people: non-drinkers or former-drinkers, initiate drinking. In this situation, the target population for health behaviour attribution calculations can be restricted to for respondents with moderate or higher drinking, or multiple reference exposures could be created.

### B.2.2 Health behaviours with interactions or multiple coefficients

The health behaviour attributable risk and the health behaviour-deleted risk estimates can be calculated for health behaviours with interaction terms, including complex interactions or multiple coefficients. Examples include:

- health behaviour parameters with age interaction;
- spline functions with terms for each knot point; or,
- composite risks such as smoking which may coefficients for smoking status (current, former, never) and pack-years.

All coefficients that related to a common risk factor should be simultaneously considered.

### B.2.3 Exposures not in the original algorithm



## Appendix C

# Bibilography



# Bibliography

- Chiang, C. L. (1984). *The Life Table and Its Applications*. Robert E. Krieger Publ. Co., Malabar, Florida.
- Hsieh, J. J. (1991). Construction of expanded continuous life tables—a generalization of abridged and complete life tables. *Mathematical Biosciences*, 103(2):287–302.
- Manuel D, e. a. (2012). Predictive risk algorithms in a population setting: an overview. *Journal of Epidemiology & Community Health*, 66(10):859–865.
- Pampalon R, Raymond G (2000). A deprivation index for health and welfare planning in quebec. *Chronic Dis Can*, 21(3):104–13.
- Rehm J, e. a. (2006). Alcohol-attributable mortality and potential years of life lost in canada 2001: implications for prevention and policy. *Addiction*, 101:373–84.
- Statistics Canada (2010). Healthy people, healthy places. Technical report, Statistics Canada.