Planning Tool Guidance

Project Big Life

2019-06-10

Contents

Chapter 1

Welcome to Project Big Life's Planning Tool

To do: Include motivation: why someone should use the platform (video and text)

```
## PhantomJS not found. You can install it with webshot::install_phantomjs(). If it is installed, pleas
```

Chapter 2

Introduction

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

For instance Project Big Life Planning Tool helps:

- Public health practitioners: assess the impact of a preventative intervention on a health behaviour.
- Health planners: assess the need for palliative care.

What types of questions can it answer?

The Project Big Life Planning Tool can answer the following types of questions (this list is not exhaustive): - What is the burden of smoking on life expectancy? - How many deaths would be prevented if everyone met their daily excersize requirements?

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. Life Expectancy).
- The multivariable predictive risk algorithms are developed and validated using routinely collected data 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 (Chapter 6).

Why should I used 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 current or 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 developement, and groups of people that account for only a fraction of the population.

- It is **Private**.

  - Uploaded data remains on your computer and is not uploaded or sent anywhere.

Chapter 3

Getting Started

To help you get started with Project Big Life's Planning Tool quickly, 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 (Chapter 4) nor will it provide reference material (Chapter 7), 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 (http://policy.projectbiglife.ca/) and click on the Tutorial button in the top right corner!

Chapter 4

How To

These guides will cover similar topics as the tutorials but in greater detail.

- Import Data
- Filter Data
- Stratify Data
- Calculate Outcomes
- Visualize Data
- Export Data
- Generate Intervention Scenarios
- Resolve Error Messages

4.1 Import Data

4.2 Filter Data

4.3 Stratify Data

- Default Stratification
- Custom Stratification Include the coding steps for R/STATA/SAS

4.4 Calculate Outcomes

*Note that the larger the file the longer the calculations will take. Use LE or Risk of death as examples in the how-to

4.5 Visualize Data

- export
- create your own(?)

4.6 Export Data

4.7 Generate Intervention Scenarios

- Define target population (all or high-risk only), and specify intervention effect

4.8 Resolve Error Messages

- Out of range
- Sample Size is too small

Chapter 5

Applications

This chapter provides you with examples of how Project Big Life's Planning Tool can be used in your day-to-day operations. The examples will cover: generating a health status report and determining the impact of a local and national policy.

5.1 Health Status Report

5.2 Diet

5.3 Transportation

Chapter 6

Key Topics

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

- Multivariable Predictive Risk Algorithms
- Calculations
  - Risk
  - Life Expectancy
  - Cause-Deleted
- Health Intervention Scenarios

6.1 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)

4.1.1 Development of multivariable predictive risk models

Data:

- Multivariable predictive risk models 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.

Model Generation:

- Multivariable predictive risk models are developed and validated in 4 stages:
  - Model derivation: the predictive risk model is created using data from the derivation cohort
  - Model validation: the predictive risk model is applied to the validation cohort
  - Final Model Generation: validation and derivation cohorts are combined to estimate the final application of the predictive risk model
  - Derivation of the application model: creation of a parsimonous (fewer predictors) model that maintained discrimination, calibration, and overall model performance
- In each stage of the model development and validation, model performance is assessed using measures of discrimination and calibration.

4.1.2 Multivariable predictive risk algorithms in Project Big Life Planning Tool

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

Title

Outcomes

Information

Mortality Population Risk Tool

5 year risk of death, Life Expectancy, Life years lost

Appendix 1

6.2 Calculations

6.2.1 Risk

6.2.2 Life Expectancy Calculation

6.2.3 Cause-Deleted

6.3 Health Interventions Scenarios

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 (kg/m2).

Burden

The impact or size of a health problem in an area, measured by cost, mortality, morbidity or other indicators. The burden of unhealthy behaviour is calculated by the differences in life expectancy based on individuals' exposure to four health behavioural risks for poor health relative to the healthy category.

By Row Measures

When selected, the output '.csv' file will include the result of the calculation for each row (e.g. individual) of 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 preformed 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=795204).

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.

Exposure

In the risk algorithms the exposure refers to the level of the predictor varaible, e.g. for the predictor variable EDUDR04 there are four levels of exposure: (1) Post-Secondary, (2) Some Post-Secondary, (3) High School, (4) Less then High School.

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.

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 expediture 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.

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 desanté publique du Québec (INSPQ)(Pampalon R, Raymond G, 2000) is used in this plannning 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 seperation 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'.

Summary Measures

When selected, the output .csv file will include the calculation result for the entire population of the dataset.

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.

Appendix A

Mortality Population Risk Tool (MPoRT)

Outcomes: 5-yr risk of death, Life Expectancy, Cause-deleted Life Expectancy

Calculations

Using MPoRT you are able to calculate:

- 5 year mortality risk
- Number of deaths
- Life Expectancy
- Cause-deleted deaths and life expectancy
- Burden of health behaviour in deaths and on life expectancy

Types of Questions

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

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 dataset (excluding Ontario).

MPoRTv1.0 Was used in the "Seven More Years" report, a joint report with Public Health Ontario and IC/ES (https://www.ices.on.ca/Publications/Atlases-and-Reports/2012/Seven-More-Years). 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 dataset. The algorthim used categorical predictor variables for health behaviours and sociodemographic factors.

MPoRTv1.2 Was published in PLoS (https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1002082). 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 datasets. The algorithm used multiple continous predictor variables, and added chronic disease predictor variables and interaction terms.

MPoRTv2.0 - The version used 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 dataset.

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

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

Category

Variable

Scale

Description

Demographic

Age*

Continous

5 knot spline. Valid range 20 to 102

Sex

Dichotomous

Stratified Female/Male

Health Behaviour

Pack years of smoking

Continous

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

Smoking Status

Categorical

Non-smoker

Current Smoker

Former Smoker <= 5 years

Former > 5 years

Alcohol (number of drinks per week)

Continous

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

Continous

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

Leisure physical activity (MET)

Continous

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; Southeach Asian

Other; Indigenous; Latin American; Multiple origin; unknown

Immigrant

Dichotomous

Yes/No

Fraction of lifetime in Canada

Continous

3 knot spline†. 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‡

BMI

Continous

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

* Age interaction included for all variables exept immigrant, fraction of time in Canada, and ethnicity † Excluded in the male model, remains in the female model ‡ Excluded in the female model, remains in the male model

Bibliography

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.

Statistics Canada (2010). Healthy people, healthy places. Technical report, Statistics Canada.