

EPIC model reference

Mohsen Sadatsafavi

Saturday, August 13, 2016

Contents

This document is automatically generated using a markdown tool. It is a work in progress.	1
BACKGROUND	2
Structure of EPIC	2
Global parameters	2
Demographic module	3
Smoking module	3
COPD module	4
Lung Function (disease progression) module	4
Exacerbation module	5
Cost and health state utility values (payoffs)	6
Model validation results	6
Sanity checks (see if all tests have passed)	6
Validating population module and mortality	7
Validating the smoking module	12
Validating of COPD module	14
Validating of lung function module	15
Model outputs	16
Predicted COPD incidence and prevalence	16
Conclusions	18

**This document is automatically generated using a markdown tool.
It is a work in progress.**

```
## Loading required package: gsubfn
## Loading required package: proto
## Loading required package: RSQLite
## Loading required package: DBI

##
## > errors <- c(ERR_INCORRECT_SETTING_VARIABLE = -1, ERR_INCORRECT_VECTOR_SIZE = -2,
## +   ERR_INCORRECT_INPUT_VAR = -3, ERR_EVENT_STACK_FULL = -4,
## +   .... [TRUNCATED]
##
## > record_mode <- c(record_mode_none = 0, record_mode_agent = 1,
## +   record_mode_event = 2, record_mode_some_event = 3)
##
## > agent_creation_mode <- c(agent_creation_mode_one = 0,
```

```
## +      agent_creation_mode_all = 1, agent_creation_mode_pre = 2)
##
## > medication_classes <- c(MED_CLASS_SABA = 1, MED_CLASS_LABA = 2,
## +      MED_CLASS_LAMA = 4, MED_CLASS_ICS = 8, MED_CLASS_MACRO = 16)
##
## > events <- c(event_start = 0, event_fixed = 1, event_birthday = 2,
## +      event_smoking_change = 3, event_COPD = 4, event_exacerbation = 5,
## +      ev .... [TRUNCATED])
##
## Initializing the session
## [1] 0
```

Welcome to EPIC! EPIC is a microsimulation model of COPD to inform research, practice, and policy. This document contains details of model structure and results of model validation and calibration.

BACKGROUND

EPIC is a ‘Whole Disease Model’ of COPD from its pre-clinical stages to advanced disease and death. The promise of EPIC is to enable evaluation of clinical or policy decisions at any level of care in the context of COPD. The purpose of EPIC is to enable the evaluation of many different decisions potentially in interaction with each other, at different levels of care. Examples of decisions include policies that would change the future rate of smoking, implementation of biomarkers for the diagnosis of COPD or its acute exacerbations, the use of different medications for management of patients with COPD, strategies towards reducing the rate of readmissions to hospital, and end of life care. The breadth of the decision set means EPIC needs to model the entire pathway of COPD progression (including its incidence).

Structure of EPIC

EPIC is a Discrete Event Simulation (DES) model. It simulates the progression of individuals one at a time, by moving time continuously forward until the individual dies or the time horizon of the analysis is finished. The structure of EPIC is organized in modules as follows - Demographic module: models the creation of simulated individuals and their characteristics - Smoking module: models the smoking behavior as an important risk factor for COPD - COPD module: models the occurrence of COPD in non-COPD population - Lung function module: models the progression of COPD once it develops - Payoff module: assign costs and quality of life weights to other components of the model and calculates the overall costs and quality-adjusted life years.

Global parameters

Global parameters control the overall setting of the analysis. They include baseline (minimum age), time horizon of the analysis, and discount values. By default, EPIC models individuals 40 years or older and evaluates the outcomes from 2015 to 2034 (time horizon of 20 years).

Table 1 : Global parameters

Parameter	Description
age0	Starting age in the model
time_horizon	Model time horizon
discount_cost	Discount value for cost outcomes
discount_qaly	Discount value for QALY outcomes

Demographic module

This module is responsible for the creation, aging, and death (due to causes other than COPD) of patients. EPIC is an open (dynamic) prevalent-incident population of individuals 40 years or older at year 2015 and onwards. That is, the entire Canadian population 40 years or older in year 2015 (prevalent cohort) is generated; as well, individuals who reach age 40 during the time horizon (e.g., 2015-2034) are also created (incident cohort). Variables currently include sex, age at creation, current age, weight and height, smoking status, and pack-years.

Relative prevalence of individuals by age in 2015 is taken from Statistics Canada (CanSIM Table 052.0005). All projections are taken from the main (M1) scenario for the year 2015. The resulting prevalent cohort will consist of individuals representing the Canadian population (40 years of age or older) in year 2015. Relative incidence of individuals by age during the time horizon decides the age of each newly created individual. Individuals who are born in Canada or immigrated into the country before age 40 will show up in the model upon the time they reach age 40. The model also accommodates arrival of older immigrants.

Please refer to the model validation/calibration section to evaluate the performance of this module.

Table 2 : parameters of the demographic module

Parameter	Description
p_female	Proportion of females in the population
height_0_betas	Regressoin coefficients for estimating height (in meters) at baseline
height_0_sd	SD representing heterogeneity in baseline height
weight_0_betas	Regressoin coefficients for estimating weight (in Kg) at baseline
weight_0_sd	SD representing heterogeneity in baseline weight
height_weight_rho	Correlaiton coefficient between weight and height at baseline
p_prevalence_age	Age pyramid at baseline (taken from CanSim.052-0005.xlsm for year 2015)
p_incidence_age	Discrete distribution of age for the incidence population (population arriving after the first date) - ger
p_bgd_by_sex	Life table
l_inc_betas	Ln of incidence rate of the new population - Calibration target to keep populatoin size and age pyram
ln_h_bgd_betas	Effect of variables on background mortality

Smoking module

The smoking module is responsible for modeling the smoking behaviour in individuals. Note that the smoking history at baseline (pack-years) is handled in the demographic module.

The events associated with smoking are handled through smoking_change event in the model. At the time a patient is created, pack-year and smoking status are modeled through regression equations based on CCHS and CanCOLD data. During their lifetime, patients can initiate smoking for the first time, quick smoking, or experience smoking relapse. The log-hazard of these events are modeled as regression equations relating the hazard to patient demographics.

In baseline, the model is calibrated to generate declining smoking patterns from 2015 onwards which is the continuation of ~4% annual decline observed in the past decades.

Please refer to the model validation/calibration section to evaluate the performance of this module.

Table 3 : parameters of the smoking module

Parameter	Description
logit_p_current_smoker_0_betas	Probability of being a current smoker at the time of creation
logit_p_ever_smoker_con_not_current_0_betas	Probability of being an ever smoker conditional on not being current s
pack_years_0_betas	Regression equations for determining the pack-years of smoking at the

Parameter	Description
pack_years_0_sd	Standard deviation for variation in pack-years among individuals (current)
ln_h_inc_betas	Log-hazard of starting smoking (incidence or relapse)
ln_h_ces_betas	Log-hazard of smoking cessation

COPD module

This module is responsible for simulating prevalent and incident COPD.

Note that EPIC does not model lung function (e.g., FEV1) in individuals without COPD. Instead, the development of COPD is modeled as a binary status. The definition of COPD is that of the GOLD criteria ($FEV1/FVC < 0.7$).

The core equations are prevalence equations and incidence equations. The former decides whether individuals have COPD at the time of creation. The latter determines if, and at what time, an individual without COPD at baseline develops COPD.

Prevalent COPD: The best source of evidence currently seems to be the BOLD study. This international study recruited XXX individuals from Vancouver, Canada, and reported a prevalence of X% in men and X% in women. Importantly, this study reported the odds ratio (OR) of COPD over decades of age and pack-years of smoking. This provided us with all the quantities needed to formulate the logit of the probability of having COPD as a function of individual characteristics and an intercept factor.

Incident COPD: There is much less information published on the incidence of COPD. This is natural given that the moment a person satisfies the definition of COPD based on spirometric criteria is no a catastrophic one (like myocardial infarction) in the life of the COPD patient. EPIC models COPD incidence on the basis of the core assumption that within the strata of COPD risk factor, COPD incidence is such that it keeps COPD prevalence constant over time (a concept embedded in the SIR epidemiological models). This means we could, through calibration, formulate a regression coefficient for the log-hazard of COPD as a function of patient characteristics that results in stationary prevalence across risk factor strata. The fix prevalence was tested by running the model for a large number of patients, extracting the data, and estimating a logit model for COPD prevalence. The calibration was deemed successful if a) the regression coefficients matched those seen in BOLD, and b) the coefficient for calendar time was close to zero.

Table 4 : parameters of the COPD module

Parameter	Description
logit_p_COPD_betas_by_sex	Logit of the probability of having COPD ($FEV1/FVC < 0.7$) at time of creation (separate for men and women)
ln_h_COPD_betas_by_sex	Log-hazard of developing COPD ($FEV1/FVC < 0.7$) for those who did not have COPD at baseline

Lung Function (disease progression) module

EPIC models FEV1 as the core variable representing the natural history of COPD. FEV1 is only modeled when the individual develops COPD. Modeling FEV1 is done in two stages: modeling baseline FEV1 at the time individuals are created, and modeling their progression over time.

Lung function at creation time

We rely on the patient-level data of the nationally representative COLD and the Canadian Cohort of Obstructive Lung Disease (CanCOLD) studies. Both studies are population-based with multi-stage sampling to ensure representativeness. Because FEV1 is only modeled among those with COPD, we needed to estimate its distribution with respect to patient characteristics conditional on $FEV1/FVC < 0.7$. In order to do so,

we ran an OLS on COLD data for individuals with FEV1/FVC<0.7. For the incident COPD, we need to construct a probability distribution of FEV1 condition on FEV1/FVC=0.7. Again, we used COLD data, this time including all individuals. The OLS model also had as an independent variable $x=FEV1/FVC$. We predict the distribution of FEV1 by setting $x=0.7$.

Lung function decline over time

Lung function decline is modeled based on our published work using data from the Lung Health Study. In brief, we fitted a random-intercept, random-slope model for baseline and future FEV1. Conditional estimates of slope of FEV1 decline as a function of patient characteristics and baseline FEV1 (estimates as explained above).

GOLD severity grades

Much of the information on COPD outcomes (e.g., costs, quality of life, mortality rates) is available according to the GOLD severity grades which are based on FEV1 to its predicted ratio. EPIC quantifies GOLD grades based on the modeled FEV1 combined with predicted FEV1s estimated based on age, sex, and height in the Canadian population (Tan et. al.)

Table 5 : parameters of the lung function module

Parameter	Description
fev1_0_prev_betas_by_sex	Regression (OLS) coefficients for mean of FEV1 at time of creation for those with COPD (separately by sex)
fev1_0_prev_sd_by_sex	SD of FEV1 at time of creation for those with COPD (separately by sex)
fev1_0_inc_betas_by_sex	Regression (OLS) coefficients for mean of FEV1 at time of development of COPD(separately by sex)
fev1_0_inc_sd_by_sex	SD of FEV1 at time of development of COPD (separately by sex)
pred_fev1_betas_by_sex	Coefficients for calculation of predicted FEV1 based on individual characteristics
dfev1_betas	Regression equations (mixed-effects model) for rate of FEV1 decline
dfev1_re_sds	SD of random-effect terms in the mixed-effects model of FEV1 decline
dfev1_re_rho	Correlation coefficient between random-effect terms in the mixed-effects model of FEV1 decline

Exacerbation module

Exacerbations are identified by their time of onset, duration, and severity. Currently, EPIC models three levels of severity (mild, moderate, severe). There is heterogeneity with regard to exacerbations in two ways: individuals vary in their background rate of exacerbation (over and above the variation explained by observable characteristics), and in their tendency to experience more severe versus milder exacerbations. Parameters related to rate and severity have been largely estimated from the dedicated analysis of MACRO clinical trial.

The rate of exacerbation is modeled as a Poisson process (assumption of constant hazard was supported by our analysis of MACRO data). The log-hazard of exacerbation is a function of patient characteristics (sex, age, FEV1) and an individual-specific random-effect term. Similarly, the log-odds of more severe compared with less severe exacerbation is a function of patient characteristics (in a ordinal logit model) and a random-effect term that models between-individual variability in the severity of exacerbation.

Table 6 : parameters of the exacerbation module

Parameter	Description
ln_rate_betas	Regression coefficients for the random-effects log-hazard model of exacerbation (of any severity)
logit_severity_betas	Regression coefficients for the proportional odds model of exacerbation severity
ln_rate_intercept_sd	SD of the random intercept for log-hazard of exacerbation

Parameter	Description
logit_severity_intercept_sd	SD of the random intercept for proportional odds model of exacerbation severity
rate_severity_intercept_rho	Correlation coefficient between the random effect terms of rate and severity
exac_end_rate	Rate of ending of an exacerbation (inversely related to exacerbation duration) according to s
p_death	Probability of death due to exacerbation according to its severity level

Cost and health state utility values (payoffs)

The payoff module attaches costs and utility values (the latter is used to calculate quality-adjusted life year [QALY]) to the events occurring in the model. Currently, costs are associated with the following state/events:

- Background costs in the general non-COPD population - Background costs of each GOLD grade (not including exacerbations) - Costs associated with exacerbations according to their severity

Similarly, utilities are associated with background - Background utilities in the general non-COPD population

- Background disutility of each GOLD grade (not including exacerbations) - loss of QALY (referred to as disutility in the model) associated with exacerbations according to their severity

Table 7 : parameters of the payoff module (costs)

Parameter	Description
bg_cost_by_stage	Annual direct costs for non-COPD, and COPD by GOLD grades
exac_dcost	Incremental direct costs of exacerbations by severity levels

Table 8 : parameters of the payoff module (utility)

Parameter	Description
bg_util_by_stage	Background utilities for non-COPD, and COPD by GOLD grades
exac_dutil	Incremental change in utility during exacerbations by severity level

Model validation results

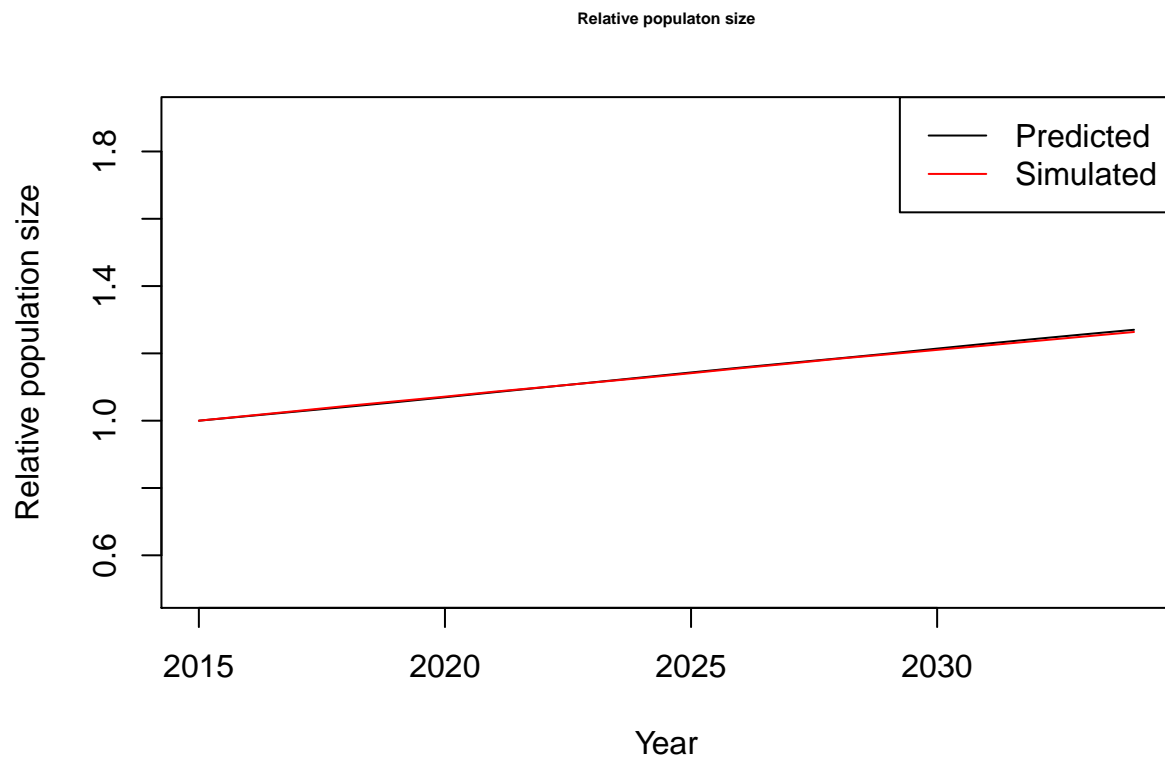
The output of several model validation routines is provided below.

Sanity checks (see if all tests have passed)

```
## Initializing the session
## test 1: zero all costs
## Test failed!
## test 2: zero all utilities
## Test passed!
## test 3: one all utilities ad get one QALY without discount
## Test passed!
## test 4: zero mortality (both bg and exac)
## Test passed!
```

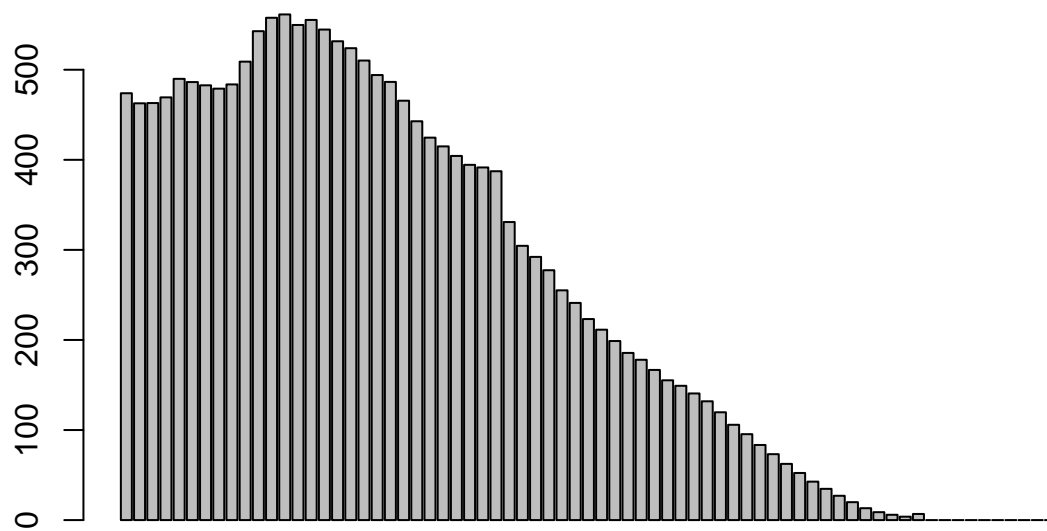
Validating population module and mortality

```
## Validate_population(...) is responsible for producing output that can be used to test if the populat
## Initializing the session
##
## Because you have called me with remove_COPD= 0 , I am NOT going to remove COPD-related mortality from
```



```
## The plot I just drew is the expected (well, StatCan's predictions) relative population growth from 2
## working...
## And the black one is the observed (simulated) growth
## Also, the ratio of the expected to observed population in years 10 and 20 are 102736.2 and 102354
## The observed population pyramid in 2015 is just drawn
```

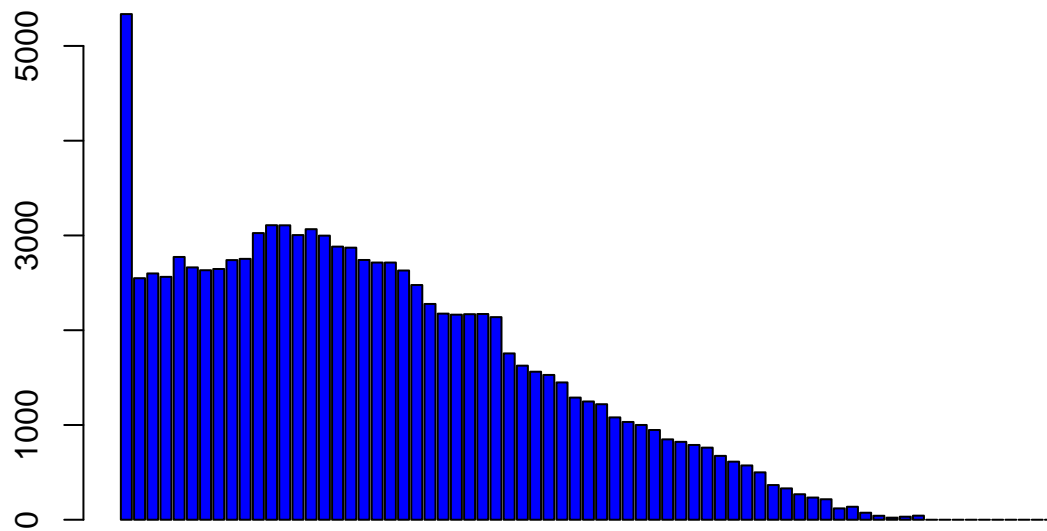
Predicted Pyramid - 2015



Age

PRedicted average age of those >40 y/o is 59.15069

Simulated Pyramid - 2015

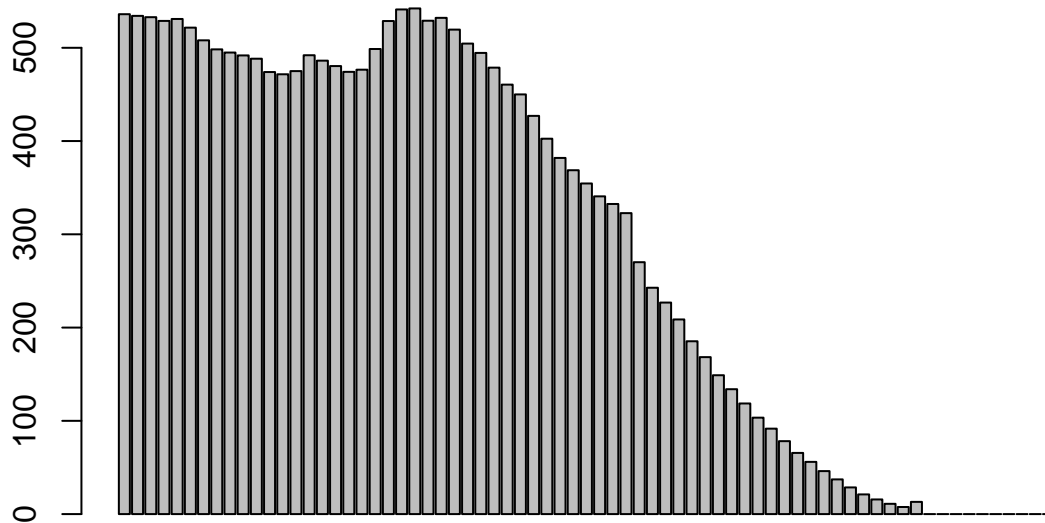


Age

Simulated average age of those >40 y/o is 58.65576

The observed population pyramid in 2025 is just drawn

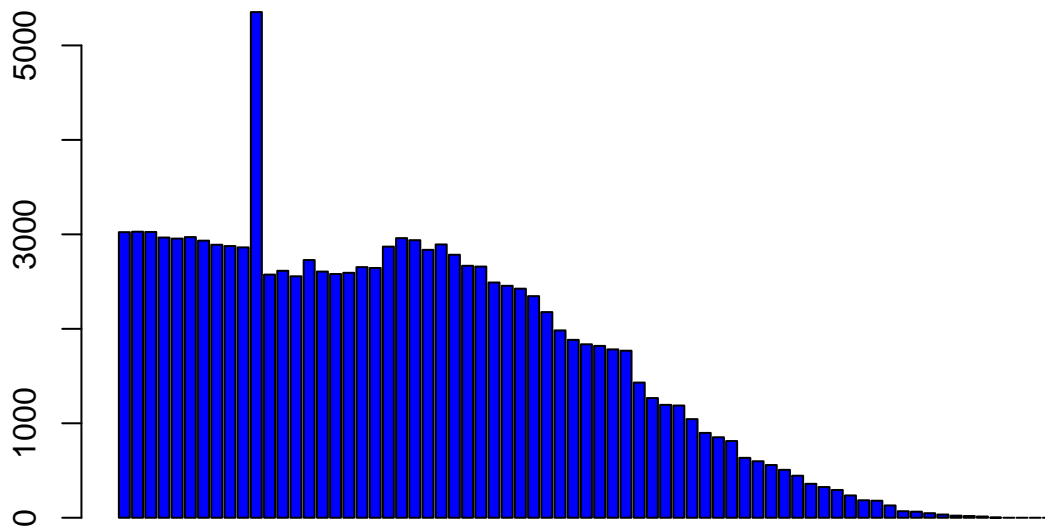
Predicted Pyramid - 2025



Age

PRedicted average age of those >40 y/o is 60.85587

Simulated Pyramid - 2025

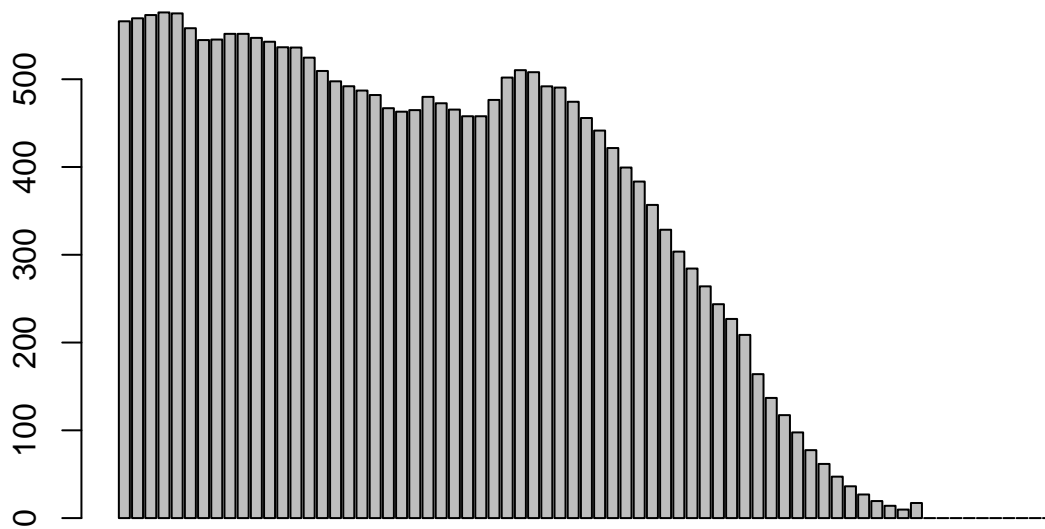


Age

Simulated average age of those >40 y/o is 60.65449

The observed population pyramid in 2034 is just drawn

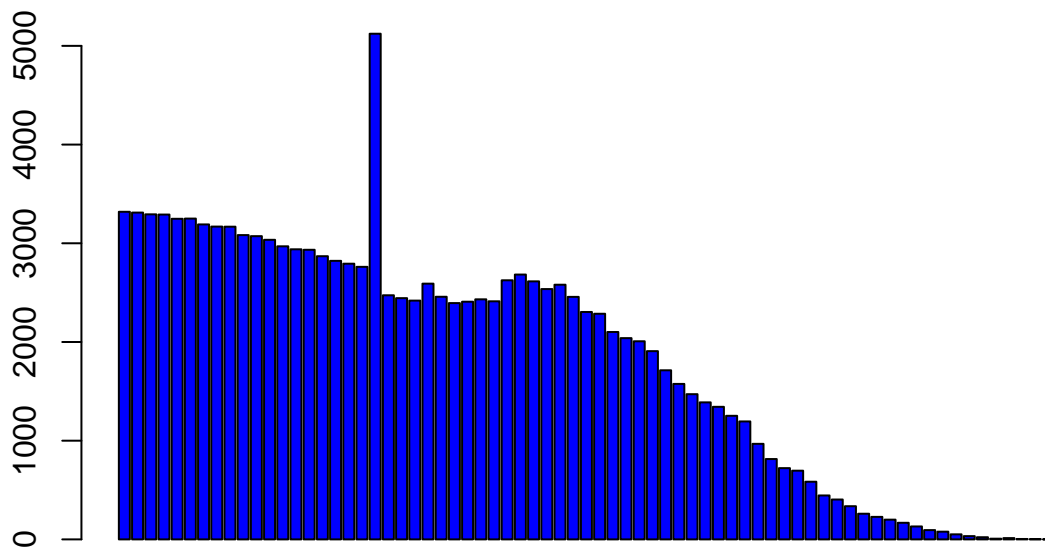
Predicted Pyramid - 2034



Age

```
## PRedicted average age of those >40 y/o is 62.04563
## Simulated average age of those >40 y/o is 61.97396
## This task is over... terminating
```

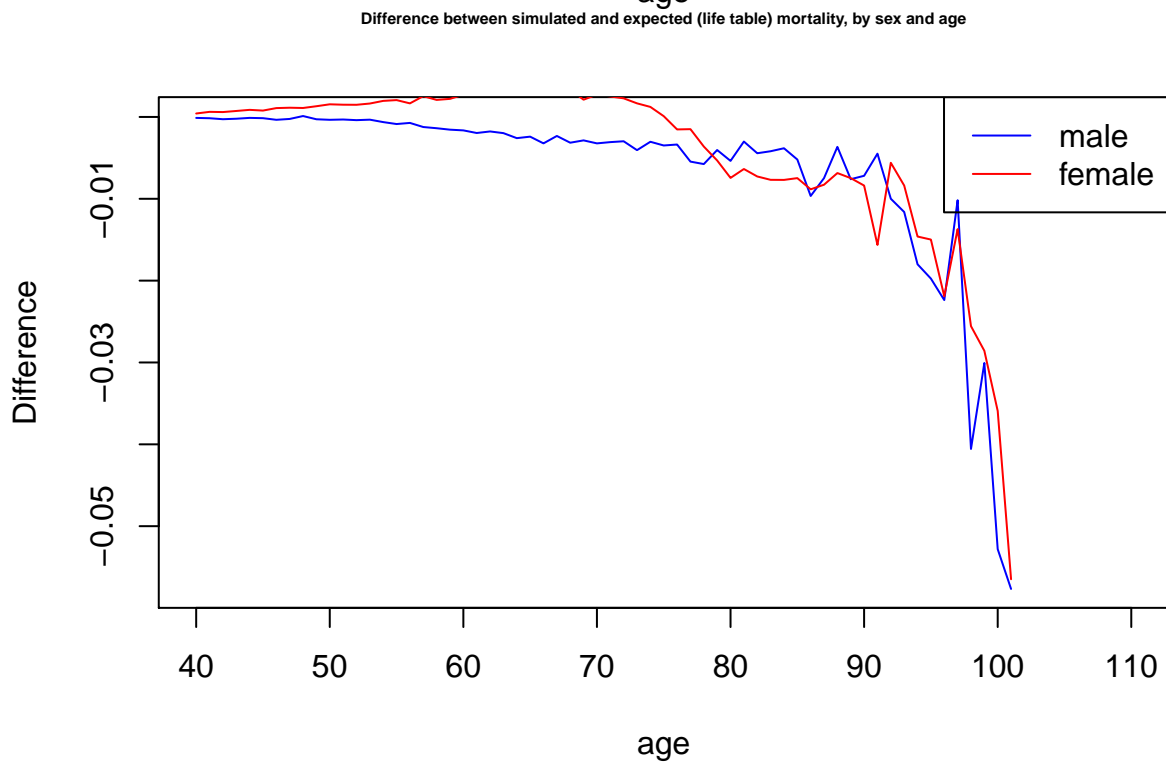
Simulated Pyramid - 2034



Age

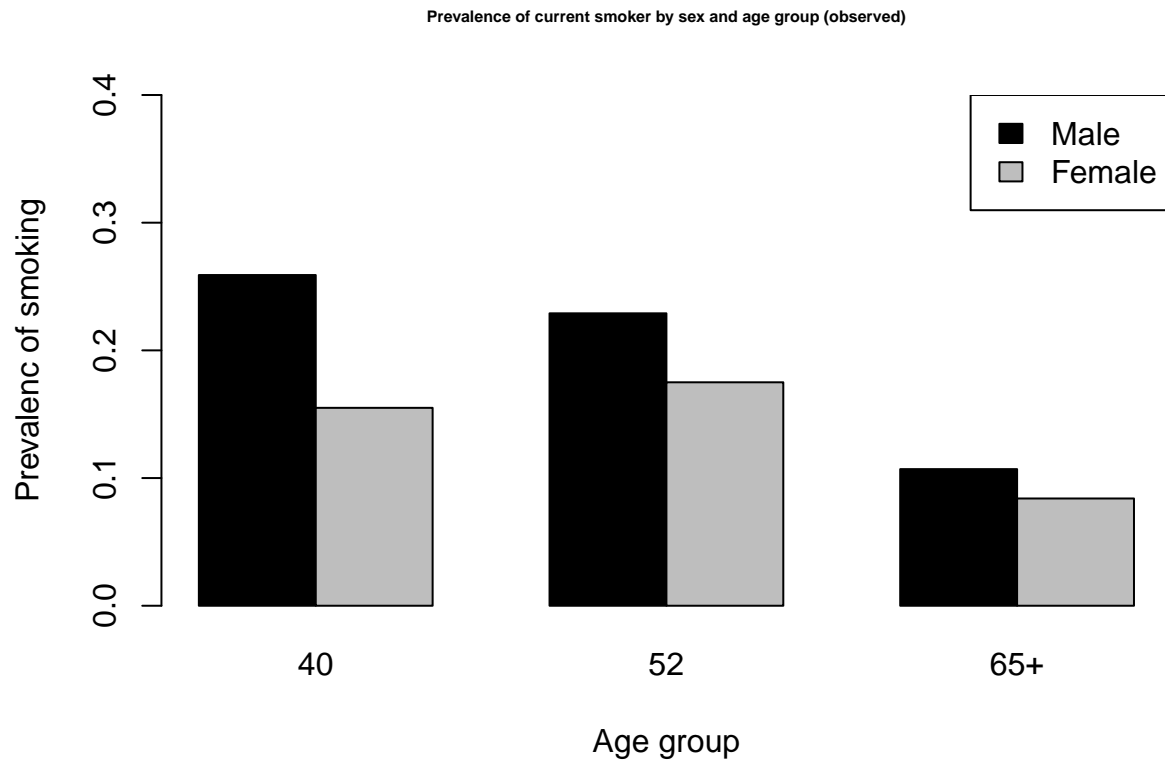
```
## Terminating the session
```

```
## Hello from EPIC! I am going to test mortality rate and how it is affected by input parameters
## Initializing the session
## working...
## Mortality rate was 0.0142484
```

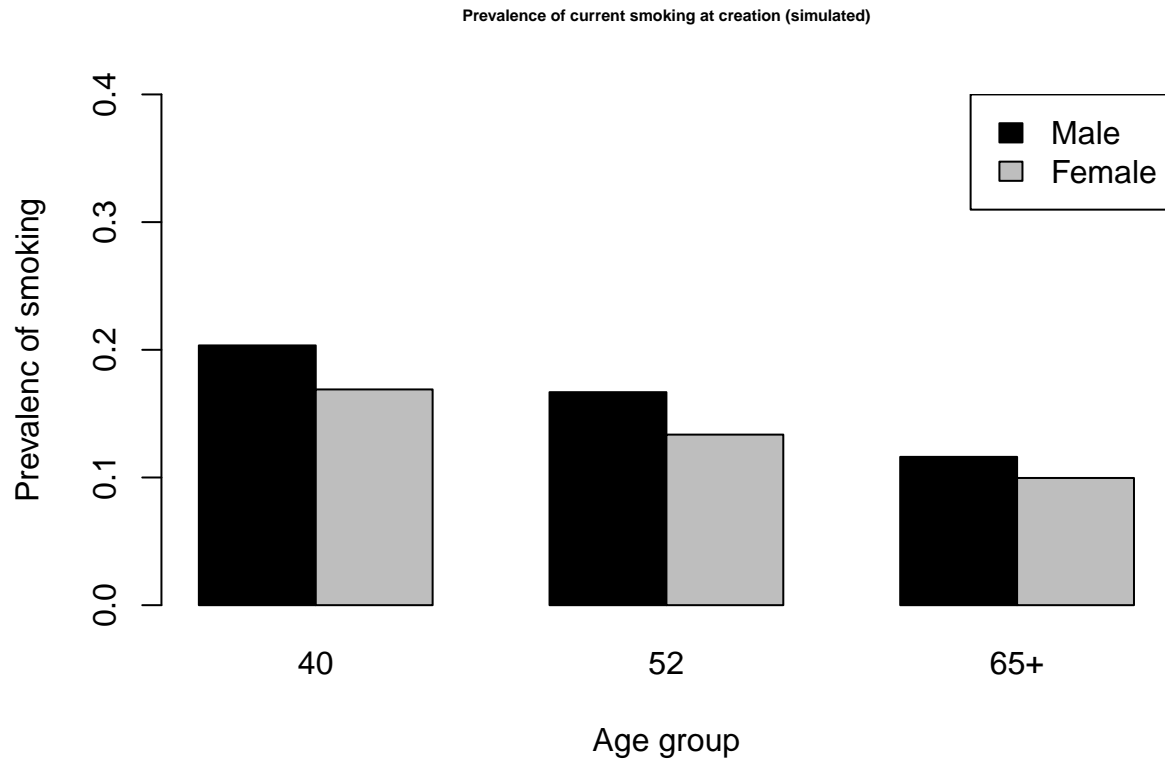


Validating the smoking module

```
## Welcome to EPIC validator! Today we will see if the model make good smoking predictionsInitializing ...
##
## Because you have called me with remove_COPD= 1 , I am indeed going to remove COPD-related mortality :
## Starting validation target 1: baseline prevalence of smokers.
## This is the observed percentage of current smokers in 2014 (m,f)
```



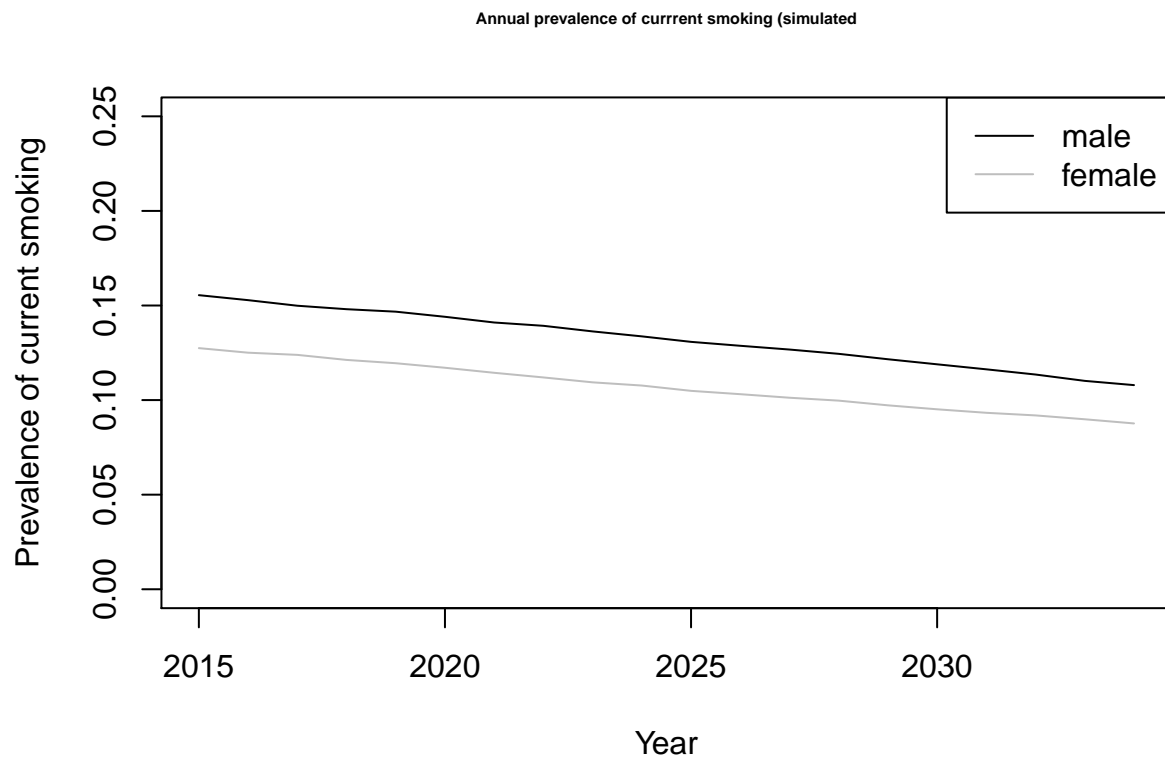
```
## Now I will run the model using the default smoking parametersrunning the model
## This is the model generated bar plot
```



This step is over; press enter to continue to step 2Now we will validate the model on smoking trends.

average decline in % of current_smoking rate is 0.01924859

This test is over; terminating the session



Terminating the session

Validating of COPD module

```
## Initializing the session
## working...

## Loading required package: tcltk

## Terminating the session

## $p_copd_at_creation
## [1] 0.137245
##
## $inc_copd
## [1] 0.0307521
##
## $inc_copd_by_sex
## [1] 0.0307521
##
## $p_copd_at_creation_by_sex
## [1] 0.1520532 0.1225075
##
## $p_copd_at_creation_by_age
##      40-50      50-60      60-70      70-80      80-111
## 0.04964626 0.12452304 0.21964380 0.36890951 0.55105309
##
## $p_copd_at_creation_by_pack_years
##      0-25      25-50      50-75      75-Inf
## 0.08651081 0.13100734 0.50398712 0.81756757
##
## $scalib_prev_copd_reg_coeffs_male
## (Intercept)      age  pack_years      year
## -6.03998535  0.06871236  0.02079582  0.02796011
##
## $scalib_prev_copd_reg_coeffs_female
## (Intercept)      age  pack_years      year
## -6.26649162  0.06888418  0.02033483  0.03136909
##
## $scalib_prev_gold2p_reg_coeffs_male
## (Intercept)      age  pack_years      year
## -6.374876256  0.054232592  0.022515938  0.005837331
##
## $scalib_prev_gold2p_reg_coeffs_female
## (Intercept)      age  pack_years      year
## -6.91493294  0.05921363  0.02695736  0.01587158
##
## $scalib_prev_gold3p_reg_coeffs_male
## (Intercept)      age  pack_years      year
## -10.94814242  0.08476280  0.02381830 -0.01042851
##
## $scalib_prev_gold3p_reg_coeffs_female
## (Intercept)      age  pack_years      year
## -12.016161654  0.098311090  0.033555550 -0.002229539
```

Validating of lung function module

This function examines FEV1 valuesInitializing the session

Trajectories of FEV1 in 100 individuals

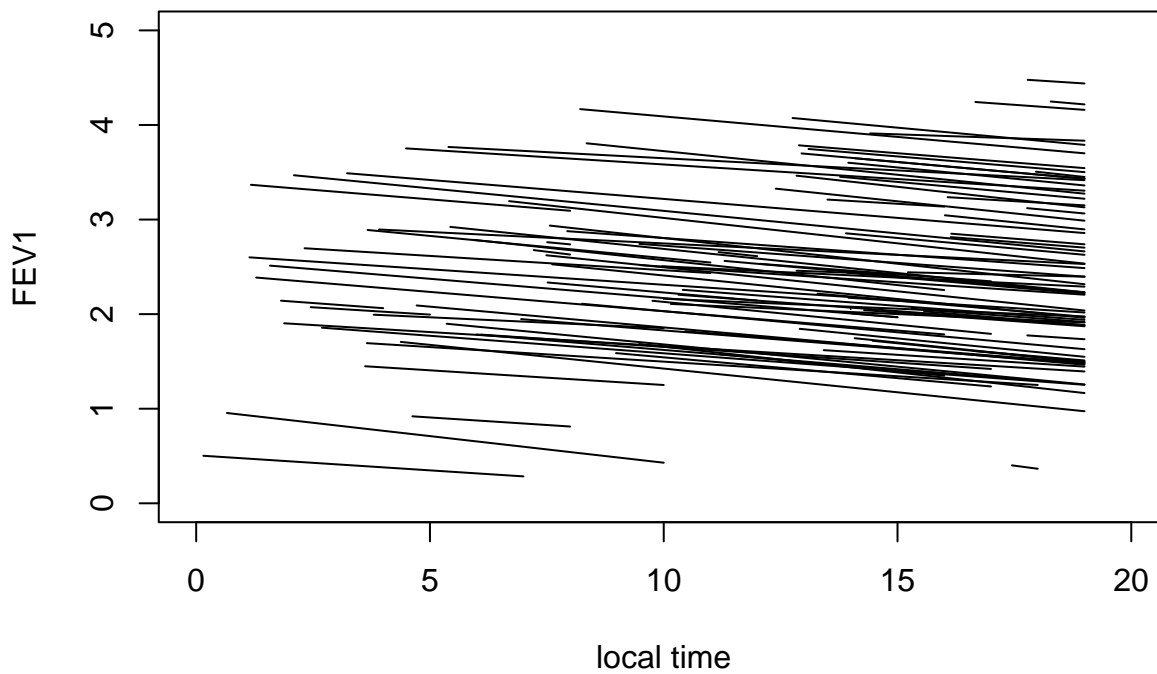


Table 9 : FEV1 for the prevalent population

gold	Mean	SD
0	0.0000000	0.0000000
1	2.7901408	0.7758154
2	1.8855108	0.6022888
3	1.0070443	0.3296322
4	0.3327707	0.2702329

Table 10 : FEV1 for the incident population

gold	Mean	SD
1	2.7351012	0.7802488
2	1.8959559	0.5665680
3	0.9686778	0.3360303
4	0.4012855	0.1990587

Table 11 : GOLD grades for the prevalent population

gold	N	Percent
0	70564	0.863
1	6102	0.075
2	4299	0.053
3	662	0.008

gold	N	Percent
4	172	0.002

Table 12 : GOLD grades for the incident population

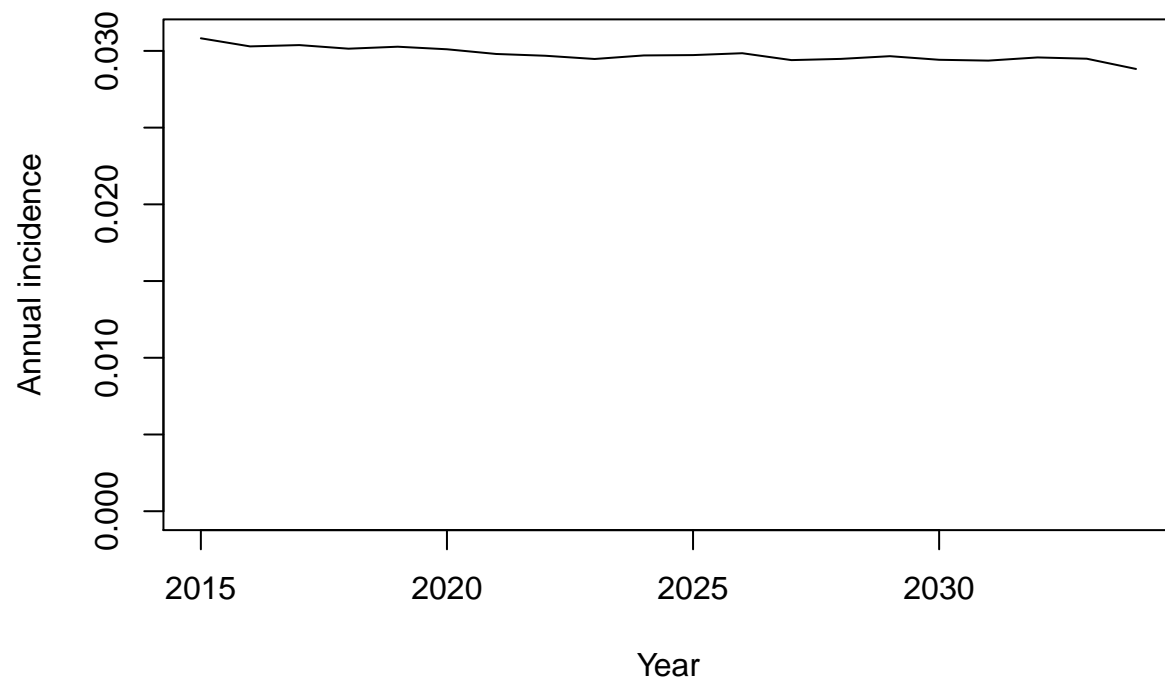
gold	N	Percent
1	19494	0.769
2	5527	0.218
3	296	0.012
4	24	0.001

Model outputs

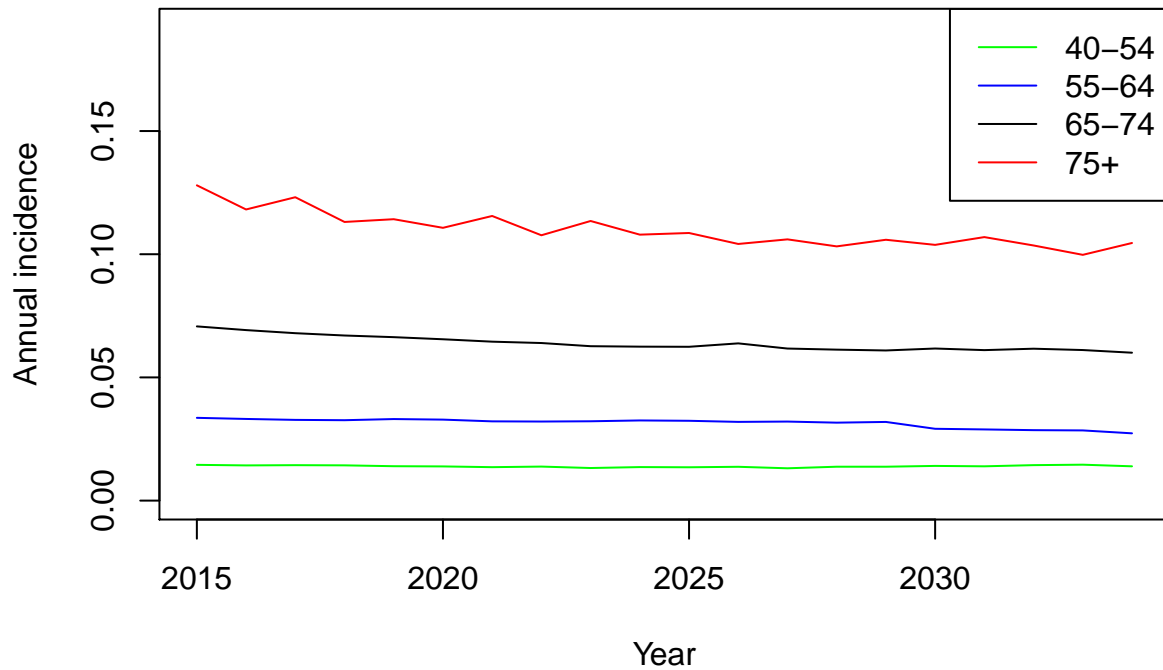
Predicted COPD incidence and prevalence

Initializing the session

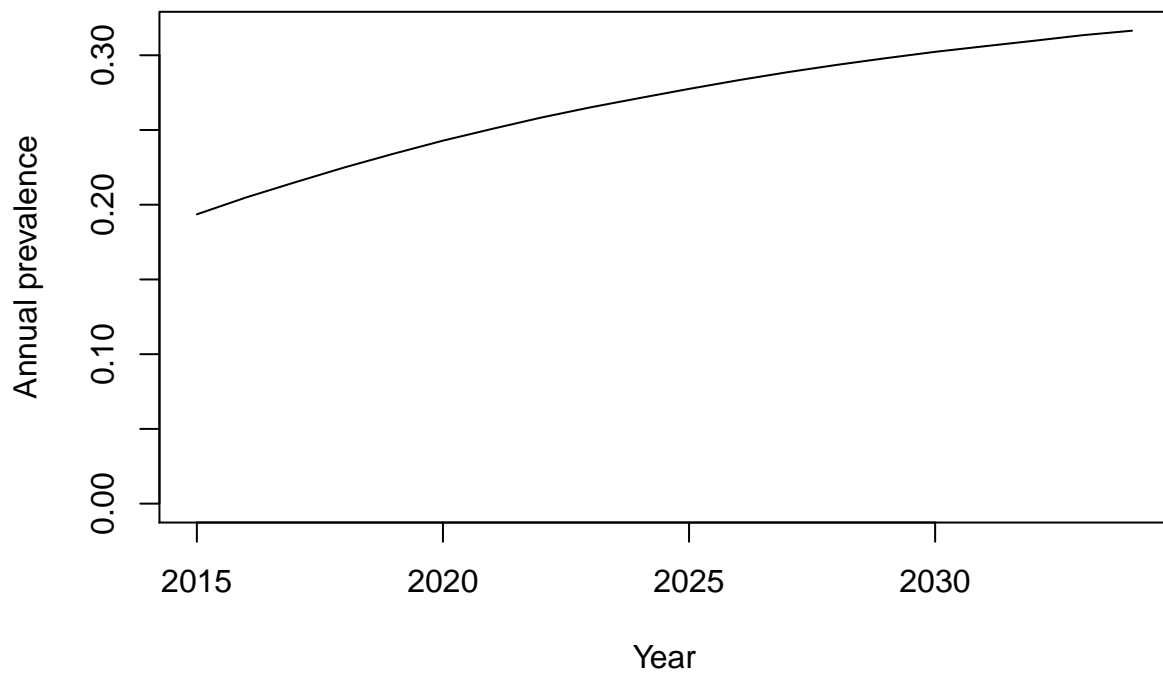
Incidence over time (simulated)



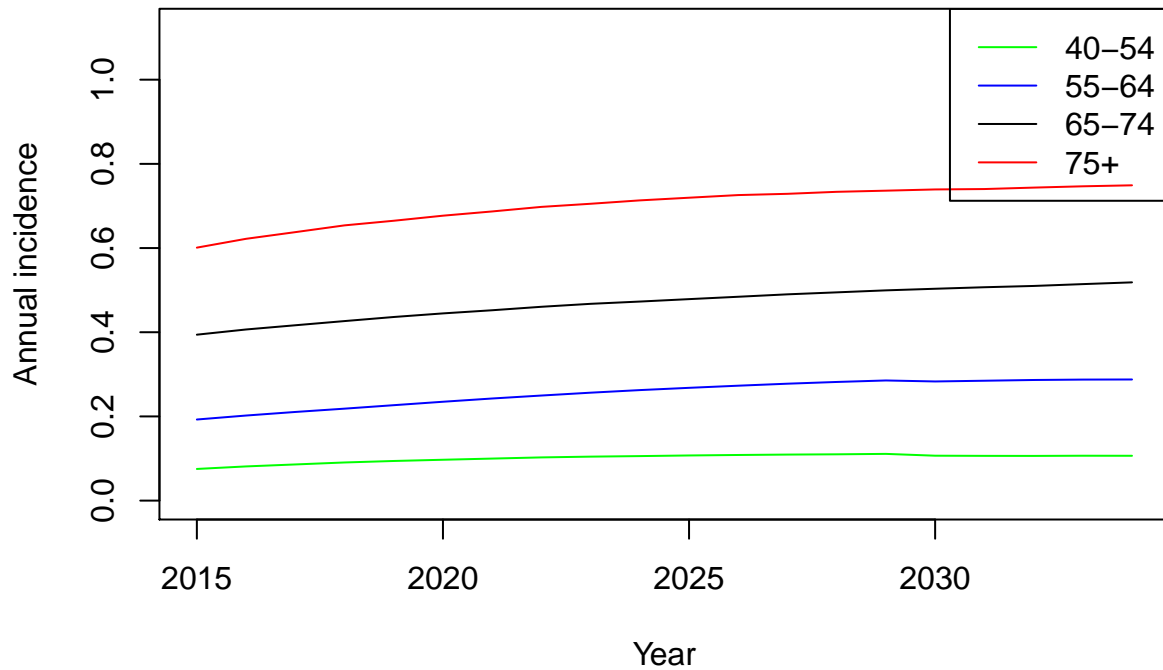
Incidence by age group (simulated)



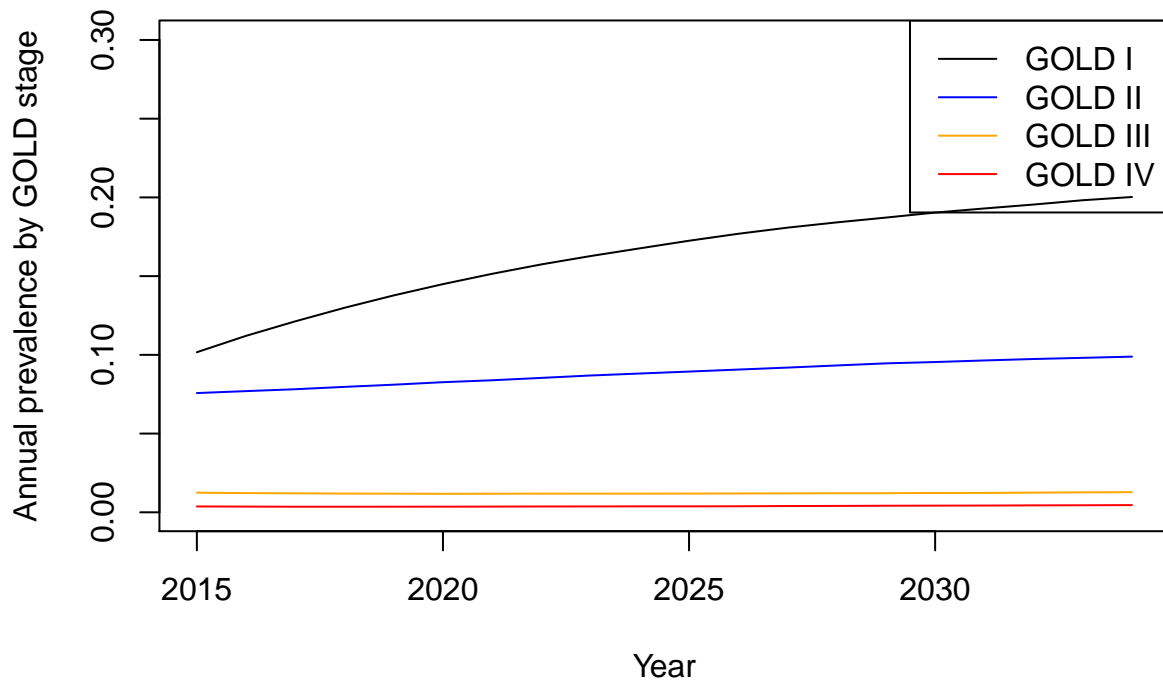
Prevalence over time (simulated)



Prevalence by age group (simulated)



Prevalence by GOLD stage (simulated)



Conclusions

The EPIC seems to perform well on many dimensions. More work is required to calibrate the model on mortality as currently the mortality rates for older groups is below the life table values in Canada.

The following items are on the agenda for the development team: 1. Setting new validation targets for model output. 2. Addition of comorbidity to the model