MASTER'S DEGREE THESIS

Interuniversity Master in Statistics and Operations Research UPC-UB

Discrete event simulation applied to prediction of future demand of colonoscopies

Author: Javier Louro Aldamiz-Echevarría

Advisor: Mercè Comas, Pau Fonseca

Department: Statistics and operational research

University: Universitat politècnica de Catalunya

Academic year: 2016

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Facultat de Matemàtiques i Estadística

UNIVERSITAT DE BARCELONA

Universitat Politècnica de Catalunya
Facultat de Matemàtiques i Estadística

MASTHER'S DEGREE THESIS

Discrete event simulation applied to prediction of future demand of colonoscopies

Javier Louro Aldamiz-Echevarría

Advisor: Mercè Comas, Pau Fonseca

Departament d'Estadística i Investigació Operativa

This dissertation is dedicated to the whole Epidemiology and Evaluation deparment of the Hospital del Mar, in particular to my office mates Laura, Anabel and Isa because of the incredible work enviroment created and support offered this months that led to enthusiasm and desire to go to work every morning and to my mentor Mercè, without her help and guidance this work would not have neither head nor tail.

Far better an approximate answer
to the right question, which is often vague,
than an exact answer to the wrong question,
which can always be made precise.
- John Tukey [1]

[1] See Tukey J. (1962)

Prologue

Popularity of simulation techniques keeps growing in the field of health services research. Discrete event simulation is being used more and more for economic evaluation, although it was initially applied for planning of healthcare resources. In the context of population-based screening programs, discrete event simulation models offer a tool to predict future needs of services according to the population structure and, thus, represent a useful tool for health care planners and decision-makers. The case of colorectal cancer screening is currently in the frontline, as the Catalan Director Plan of Oncology has carried out the extension of the Program to all the Catalan territory during 2015, which will imply that in 2016 the Program will be active and generating needs of colonoscopies in all Catalonia. Colorectal cancer presents a particularity in front of other cancers. The natural history of colorectal cancer presents premalignant stages in the form of polyps or adenomas, which can be detected and removed through colonoscopy to prevent them to reach a cancer stage. Even if all adenomas are removed, future surveillance under colonoscopy is required to remove new adenomas or adenomas not found in the previous colonoscopy. The intensity of surveillance and the complexity of the surveillance colonoscopies will depend on its number and size. In summary, the impact of extending the Colorectal Cancer Screening Program to the whole territory will have a great impact in the demand of colonoscopies, not only those colonoscopies needed to confirm the diagnostic after a positive screening test, but also the colonoscopies needed to make a long-term follow-up of those men and women with premalignant findings. Meeting this increased demand represents a challenge to the National Health Service and, thus, careful planning of resources and of training of specialized professionals is needed, taking into account population ageing and specificities of the health areas. The present simulation model and dissertation is a step forward the introduction of simulation techniques in the process of healthcare decision-making.

Mercè Comas
Epidemiology and Evaluation Department
Institut Hospital del Mar d'Investigacions Mèdiques (IMIM)

Abstract

Colorectal cancer is the most common cancer in both men and women in Catalonia, with more than 6,000 new cases each year. It also represents the second leading cause of cancer death. For this reason a colorectal cancer screening program is currently being extended to all Catalonia. The operational structure will consist of various territorial technical offices that include one or more endoscopic units. It is considered that the endoscopic units will assume both colonoscopies after a positive screening fecal inmunochemical test and the colonoscopies for surveillance of premalignant lesions detected in the colonoscopy after a positive FIT.

The aim of this project is to estimate the number of colonoscopies of both types that every endoscopic unit will have to provide through a discrete event simulation model, with a special focus on the probabilistic sensitivity analysis of the main factors affecting the demand of colonoscopies.

This model simulates a time horizon of 20 years (from 2015 to 2034) to allow a long term prediction of colonoscopy demand according to the program results. Individual persons were simulated. All persons underwent biennial screening from 50 to 69 years of age. The conceptual model was based on the European Guidelines for Colorectal Cancer Screening for both the screening process and the surveillance after premalignant findings process. The model was fed mainly with data from the first and second rounds of the "Barcelona Esquerra" and "Litoral Mar" Program and applied to the Catalan population aged 50-69 predicted for the future years.

For the population of all Catalonia the model predicted a total number of colonoscopies that increase from 21,286 in the first year, passing on 39.060 passed 10 years, to 45,319 after 20 years, making a total of almost 730,000 colonoscopies during these 20 years according to the model predictions. Of these 730,000 colonoscopies the 60% will be colonoscopies after a positive FIT, and 40% will be those for the surveillance of the detected pre-malignant lesions.

Results of the sensitivity analysis showed that FIT positivity is far the most sensitive variable. A change in a 1% of positivity has an impact similar to a 10% change in participation. Adherence affects to the number of surveillance colonoscopies only, and it is the variable with the lowest effect in magnitude, although its effect increases with time.

In order to deliver these results for every endoscopic unit in a friendly user way, an application was created using R software to present customized data for each unit. This application can be useful for planning the necessary resources in a 20-year horizon and will be presented to the Technical Screening Office of Catalonia.

In conclusion, this simulation model and its analysis have shown to be powerful tools for health services planning and to inform decision-making. Beyond the modelling/technical matters, this piece of research should facilitate reactions on the capacity of the health system to meet the demand of colonoscopies induced by the CRC screening program, and how endoscopy workforce should be subject to a conscious planning.

Key words: Colonoscopies, health service research, healthcare planning, probabilistic sensitivity analysis, applications.

MSC2000: 62P10

Contents

List of Figures

List of Tables

Chapter 1
Introduction

Cancer is a word, not a sentence
- John Diamond

1.1. About the problem

Colorectal cancer (also known as CRC for its acronym in English) is, in Spain and specifically in Catalonia, the most common cancer in incidence and the second in mortality, making it one of the most important to consider.

In this chapter we will do an introduction about the problem that this cancer represents and the problem we want to face in this project.

Despite colorectal cancer is a cancer that kills a lot of people worldwide yearly, it is true that this cancer has a low mortality if it is detected early, which makes very important the population awareness of screening programs.

The colorectal cancer screening program is extended in Spain, where is already implemented or piloted in all regions except in Castilla La Mancha. In Catalonia, it was established in two regions, has been extended to all Catalonia during 2015, and will be fully operative in 2016. This program it is focused to both men and women aged between 50 to 69 and it based on a biennial fecal occult blood inmunochemical test (FIT) and then a colonoscopy for positive ones.

With the extension of this screening program, arises the need of estimate the number of colonoscopies, both after a positive FIT and for surveillance of premalignant lesions detected, that every endoscopic unit will have to provide in the future for being prepared for it.

This problem will be solved using a discrete event simulation model that predicts the future demand of colonoscopies in each Endoscopic Unit of Catalonia in the next 20 years, that will be explained on chapter 2.

Participation in the screening program, FIT positivity and adherence to surveillance colonoscopies are crucial parameters in the model. All of them are included as probabilistic parameters in the model. In order to assess how the uncertainty

associated with the estimation of this three parameters affects the results of this simulation model, a sensitivity analysis was performed and explained in chapter 3.

On chapter 4 it will be explained what are the next steps on this project and how this model is being improved in order to be more accurate at predicting. Finally, the discussion and conclusions of this dissertation will be included on chapter 5.

1.2. A brief look to the epidemiology of CRC

All the data of incident, mortality and prevalence worldwide of colorectal cancer that is presented on this section has been collected from the GLOBOCAN project [1] created by the International Agency for Research on Cancer (IARC) which aim is to provide contemporary estimates of the incidence of, mortality and prevalence from major types of cancer, at national level, for 184 countries of the world. The data specifically from Spain and Catalonia is collected from the review "Las cifras del cáncer 2014 [2] that provided the Spanish Society of Oncology (SEOM)".

Colorectal cancer is the development of cancer in the colon or rectum (parts of the large intestine). It occurs due to abnormal growth of epithelial cells with the ability to invade other parts of the body. Signs and symptoms may include blood in the stool, a change in bowel movements, weight loss and fatigue. There are some factors that increase the risk that a person develops the disease, including:

History of cancer. Individuals who have previously been diagnosed and treated as having cancer have a higher risk of colorectal cancer than the general population as well as for other cancers.

Heredity. Family history of colorectal cancer, especially a close relative less than 55 years or multiple relatives. Familial adenomatous polyposis, involves nearly 100% risk of developing colorectal cancer by the age of 40, unless it has been treated. Lynch syndrome or colorectal cancer hereditary non-polyposis. Chronic ulcerative colitis or Crohn's disease, about 30% after 25 years if the entire colon is affected [3].

Age. The risk of colorectal cancer increases with increasing age. Most cases occur between 60 and 70 years, whereas before age 50 is rare, unless there is a family history of early onset of colorectal cancer, particularly adenomatous polyps [4].

Smoke. It is more likely that a person who smokes die of colorectal cancer than a non-smoking a person. The American Cancer Society did a study where it was found that women who smoke are 40% more likely to die of colorectal cancer than women who never smoked. Male smokers have a 30% higher risk of dying from the disease than their non-smoking counterparts [5].

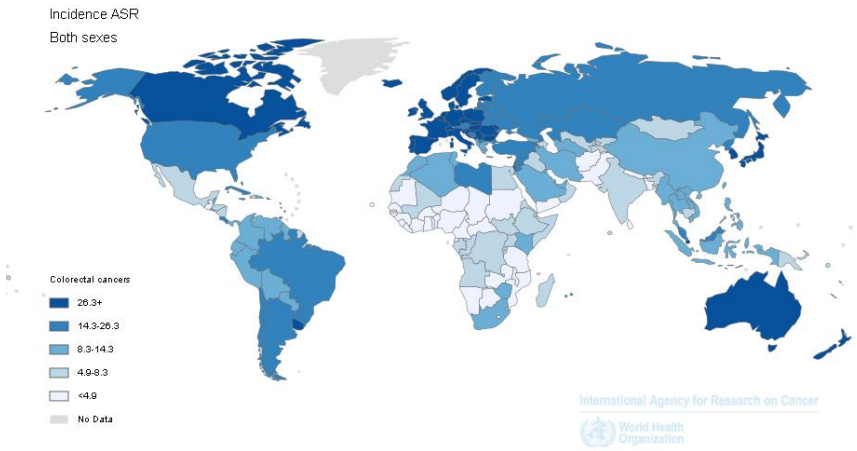
Diet. Studies show that a diet rich in meat and low in fruits, vegetables, poultry and fish increases the risk of colorectal cancer. However, other studies cast doubt on the claim that a diet high in fiber reduces the risk of colorectal cancer rather, the relationship between dietary fiber and risk of colorectal cancer is still in debate. An investigation by the European Prospective Investigation into Cancer and Nutrition suggested that diets high in red meat, as well as those low in fiber, are associated with a risk of colorectal cancer [6].

---

[1] See Globocan
[2] See Las cifras del cáncer 2014
[3] See Lutgens M.W., van Oijen M.G., van der Heijden G.J. et al (2013)
[4] See Cunningham D., Atkin W., Lenz H.J. et al (2010)
[5] See Botteri E., Iodice S., Bagnardi V., et al (2008)
[6] See Bouvard V., Loomis D., Guyton K.Z. et al (2015)

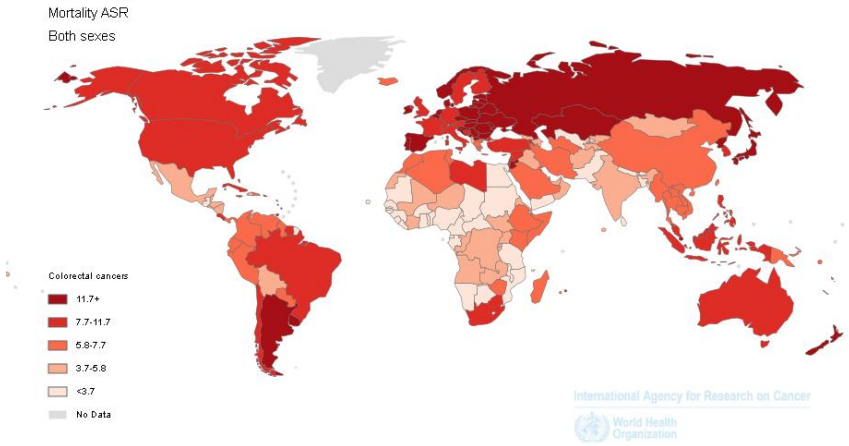
Physical activity. Those who are physically active have a lower risk of developing colorectal cancer [7].

FIG. 1.1. Estimated Colorectal Cancer incidence worldwide



Colorectal cancer was at 2012 the third most common cancer in men (746,000 cases, 10.0% of the total) and the second in women (614,000 cases, 9.2% of the total) worldwide. Almost 55% of the cases occur in more developed regions. There is wide geographical variation in incidence across the world and the geographical patterns are very similar in men and women: incidence rates vary ten-fold in both sexes worldwide, the highest estimated rates being in Australia/New Zealand (44.8 and 32.2 per 100,000 in men and women respectively), and the lowest in Western Africa (4.5 and 3.8 per 100,000) as shown in figure 1.1 [8].

FIG. 1.2. Estimated Colorectal Cancer mortality worldwide

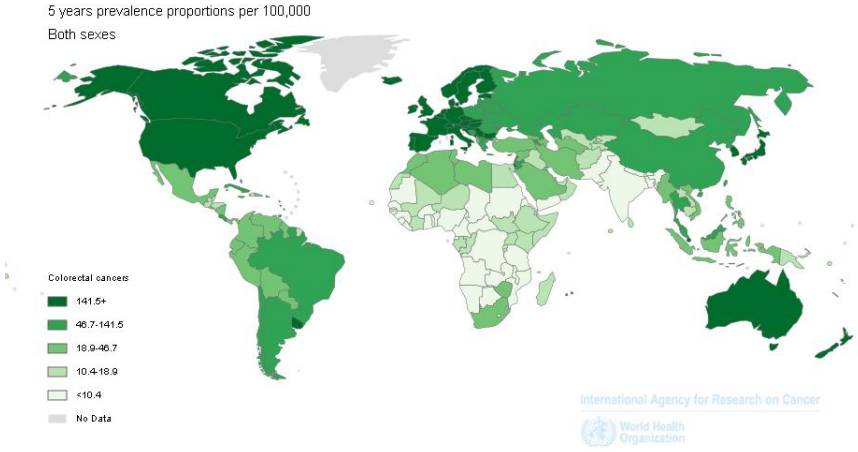


[7] See Giovannucci E., Ascherio A., Rimm E.B. et al (1995)
[8] See Ferlay J., Soerjomataram I., Ervik M. et al (2013)

Mortality is also lower in the less developed regions of the world, reflecting a poorer survival in these regions. There is less variability in mortality rates worldwide (six-fold in men, four-fold in women), with the highest estimated mortality rates in both sexes in Central and Eastern Europe (20.3 per 100,000 for men, 11.7 per 100,000 for women), and the lowest in Western Africa (3.5 and 3.0, respectively) as shown in figure 1.2 [9].

Figure 1.3 shows 5 year prevalence proportions per 100,000 inhabitants. As in terms of mortality and incidence, prevalence seems to be higher in the most developed regions of the world [10].

FIG. 1.3. Estimated Colorectal Cancer prevalence worldwide

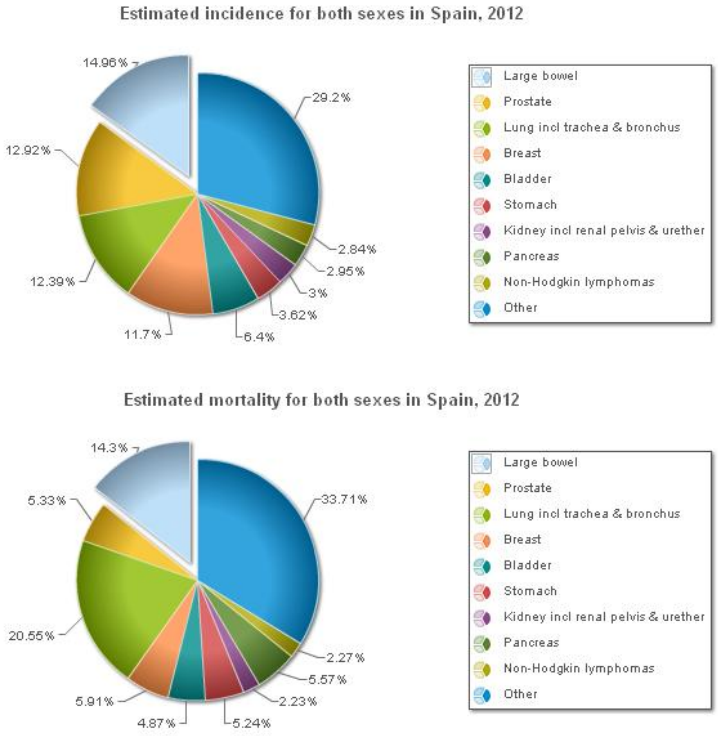


In Spain cancer is the second most common mortality cause (27.5% of deaths) just below the circulatory system diseases. Colorectal cancer is the most common cancer in terms of incidence and the second in terms of mortality if we look at both sexes. Figure 1.4 shows the estimated incidence and mortality of the different cancers in 2012.

That year, 15,604 persons of the 107,012 that died of cancer in Spain, died for colorectal cancer. Stratifying by sexes in terms of males is the third in incidence after lung and prostate cancer and second in mortality after lung cancer. In women is the second in incidence after breast cancer and the third in mortality after breast and lung cancer.

[9] See Ferlay J., Soerjomataram I., Ervik M. et al (2013)
[10] See Bray F., Ren J.S., Masuyer E. et al (2012)

FIG. 1.4. Estimated incidence and mortality for both sexes in Spain



The 5-year prevalence in Spain at 2012 was 581,688 cases with a rate of 1467.6 cases per 100,000 inhabitants, 3 years prevalence was 389,498 and 1 year prevalence was 151,257 cases.

1.3. Screening of colorectal cancer

Colorectal cancer is the most common cancer in all men and women in Catalonia, with more than 6,000 new cases each year. As in all Spain, it also represents the second leading cause of cancer death.

This cancer can be diagnosed by obtaining a sample of the colon during a sigmoidoscopy or colonoscopy. This is followed by medical imaging to determine whether the disease has spread. As over 80 % of colorectal cancers arise from adenomatous polyps, the screening of this cancer is effective not only for early detection, but also for prevention. The diagnosis of colorectal cancer by screening tends can occur 2-3 years before the diagnosis of cases with symptoms. All polyps that are detected can be removed, usually by colonoscopy, and so this prevents them from becoming cancer cells.

The three main detection tests are fecal occult blood tests (FOBT), flexible sigmoidoscopy and colonoscopy. Of the three, sigmoidoscopy is the only one that can not detect the right side of the colon, where there are 42 % of malignant tumours Virtual colonoscopy by CT scan is as good as standard colonoscopy for the detection of cancers and large adenomas, but is much more expensive, is associated with exposure to radiation, and can not eliminate the detected abnormal growths as can the standard colonoscopy.

FOBT is recommended every two years and may be guayac or inmunochemical. If results are abnormal FOBT, participants are advised to do a colonoscopy. Inmunochemical tests are very accurate and do not require changes in diet or medication before the test as guayacs does.

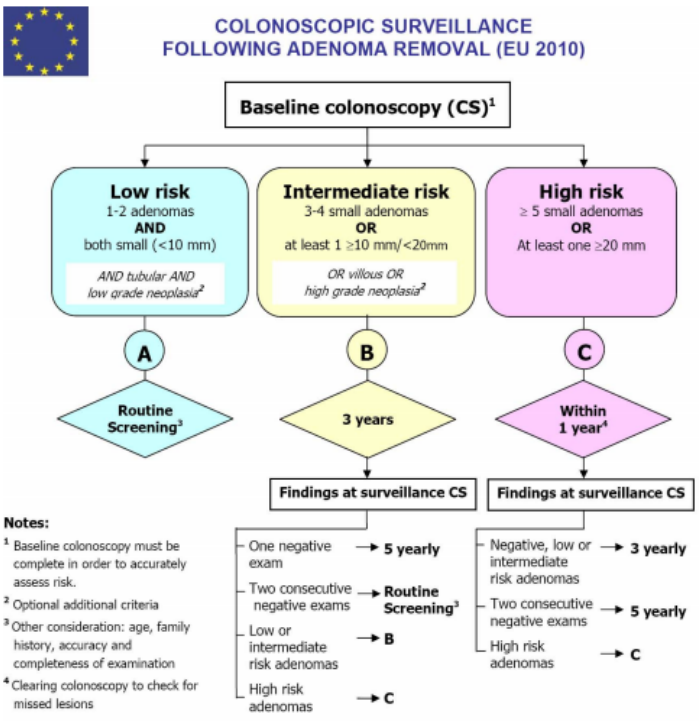
The United States Preventive Services Task Force and ather medical societies recommends screening for colorectal cancer using fecal occult blood testing, sigmoidoscopy, or colonoscopy in adults, beginning at age 50 years and continuing until age 75 years [11].

In Europe there is a guideline [12] about the process a patient has to follow when the result of the FIT is positive. Besides negative and positive for cancer, it divides the result of adenoma findings in 3 groups; low, intermediate and high risk depending on the number and size of the adenomas. This guideline is shown on figure 1.5.

Some countries have national screening programs that offer FOBT to all adults within a certain age group. Among them, in Spain the screening program is already implemented or piloted in all regions except in Castilla La Mancha. Specifically, the catalan program began in 2,000 and it includes people from 50 to 69 years old, it currently uses FIT and is based on the european guidelines.

[11] See U. S. Preventive Services Task Force.
[12] See Segnan N., Patnick J., von Karsa L. (2010)

FIG. 1.5. European Guidelines of the process of monitoring the adenoma findings of the colonoscopy after a positive FIT



The program in Catalonia is offered and managed in a decentralized manner. In particular it is organized in [13]:

- **Technical screening offices (OTC):** Their function is the management, monitoring and evaluation of the program in its territory: invitations management, coordination, collection and processing of screening tests, coordination of endoscopic units, activity monitoring and controlling and improving the quality of the screening program. These offices must have a coordinator and an adequate human team for the tasks to develop (professionals and experts in public health and preventive medicine, support for management and data analysis, administrative staff and other necessary staff). It is recommended one technical office for each health care region (RS) of the 7 which conform Catalonia, however due to the huge variation in population number between RS, those with a small volume of target population will be able to share OTC.
- **Endoscopic units:** These are the units that have to perform colonoscopies for those positive cases of screening and other tests necessary for the diagnosis process. There are actually 38 in Catalonia.
- **Pharmacies:** Act as health workers, providing information, giving people the kit and monitoring the storage and shipment of the sample to the laboratory.

[13] See Generalitat de Catalunya

- **The team of primary healthcare :** They have an important role in the program and in the dissemination of information and advice about the program and early detection of cancer.
- **Technical screening office of Catalonia:** Serving to define and ensure the application of the common organization of quality criteria and evaluation, coordinating the activities of the various OTC, developing a common system of minimum information that allows planning and evaluation of screening programs, assess the impact and results and finally propose corrective actions to improve quality and results.

The process of this program works as follows:

The OTC invites men and women between 50 and 69 years, through a letter sent to their address, to participate in the program.

The letter includes the device to collect stool sample or it is indicated where they can pick it up, either in the pharmacy or in the primary care centre. Once the samples is collected, it is returned by the patient, they return the samples by the means indicated and samples are analysed in a laboratory. In the event that the test is negative, i.e., that there are not traces of blood in the stool, or a concentration of blood below a determined threshold, a letter with the result is sent and the next invitation is issued after 2 years.

If the result is positive a colonoscopy is recommended to find out the possible causes of bleeding. Not all bleeding is due to cancer, most are due to other benign lesions. Depending on the outcome, the health care follow up will differ.

The Catalan program will be finally extended in 2016. The areas of “Barcelona Esquerra” and “Litoral Mar” began in 2009. Parameters of the present model are mainly based on of the first and second rounds of this areas.

1.4. Discrete event simulation

The science of simulation allows us to create new modelling methodologies that analyse complex systems through virtual experimentation for use to measure the impact of complex interventions in health services. Discrete-event simulation is a technique already known in operations research, that has been used in military research and manufacturing systems. In the context of health care, Markov models and decision trees have been used for several years, but they have many limitations to reproduce health problems, so the discrete-event simulation is used every day more in this area.

Markov models is conceptualized in terms of "states" and "transitions ". A patient can move from one state (this states can reproduce de health status of patient) to another, but these states are always discrete.

So it arises the need of a type of simulation which each state that represents the health status of the patient is measured by attributes, thus can be continuous or discrete, and here appears discrete-event simulation.

Discrete-event simulation has been defined as a flexible modelling method characterized by the ability to represent complex behaviours, interactions within and between individuals, populations and their environments.

You can also represent all the features of a real system, as facilities or resources. Moreover, although changes in the system are discrete, they occur on a continuous time scale, as each event is scheduled to happen at a time value drawn a continuous random distribution

Besides discrete-event simulation has a specific tool; the queues, that allows us to model a waiting list which can not be done with Markov models. Furthermore individual characteristics of each patient are simulated more straightforward with discrete event simulation.

Speaking of the model output, the output of the discrete-event simulation models is not only survival by state as Markov models, but also incidence, prevalence and evolution of health states through the time horizon of the simulation. Moreover, any simulation output can be reported at any time up to the time horizon, not just at the end as in Markov models. Furthermore, Markov models prior information is lost because of the Markov's assumption, that causes that only the current state is taken into account. In our discrete event simulation models, using events rather than states, dependence on prior events may be included, therefore the Markov's assumption is overcome.

If we use the discrete-event simulation models to analyse need and demand in terms of health services it is important to calculate the incidence and prevalence in the population all the time, taking into account survival of the population and that cohorts are dynamic, with entries and exits of individuals through time.. By contrast, the Markov models analyse patients in the initial cohort . The key point in the evaluation of health services is the prevalence of disease and the availability and consumption of resources over time. The resource capacity to meet the needs and demands is limited and there may be queues. As said before we can also analyse

waiting list that are a particular type of queue, patients are not physically lining up for the service, but are waiting to receive the medical service.

In summary, discrete event simulation was chosen to solve the present problem because a dynamic cohort, with entries and exits of entities (persons) through time was needed, because the paths of entities within the system were complex and because of the relevance of adjusting the results to the predicted population ageing.

1.5. Aim and objectives of this work

The aim and objectives we will seek in this project are:

Aim:

- Define a mathematical model to analyse through simulation the future demands of colonoscopies that has to provide each endoscopies unit to carry out the program for early detection of colorectal cancer throughout Catalonia

Objectives:

- Verify and validate the previous model.
- Make a sensitivity analysis to assess the impact of changing the participation of the population, the FIT positivity and adherence to surveillance colonoscopies, using a model and validating it.
- Define an application to predict the number of colonoscopies based on this three variables, with the model the models estimated in sensitivity analysis.

Chapter 2
Simulation model

All models are wrong,
but some are useful
- George E.P. Box [1]

To create a discrete event model we have to go through three stages.

First step we need to do before build the model is a process of observation of system in order to learn about the matter we are going to analyse.

This knowledge should then be reflected in a conceptual model, in which each member of this integrated model, events to occur as well as subjects, attributes and variables have to be estimated. It has to be included a definition of each component of the model, i.e., the events of the process studied, subjects, their attributes and parameters to be estimated. The study population must also be defined, as the level of detail and scope of the model.

The second step is to estimate the parameters that are needed to characterize the model given our ability to transform our conceptual model in a computational model that will be the next step. To do this we need to collect the necessary data for the estimation of these parameters.

As stated previously the third step is to translate the conceptual model into a computed model. We have to take care about the complexity of health systems because we're trying to model discrete events occurring at any point in time.

In this chapter I will explain briefly the model that was used and the validation of it.

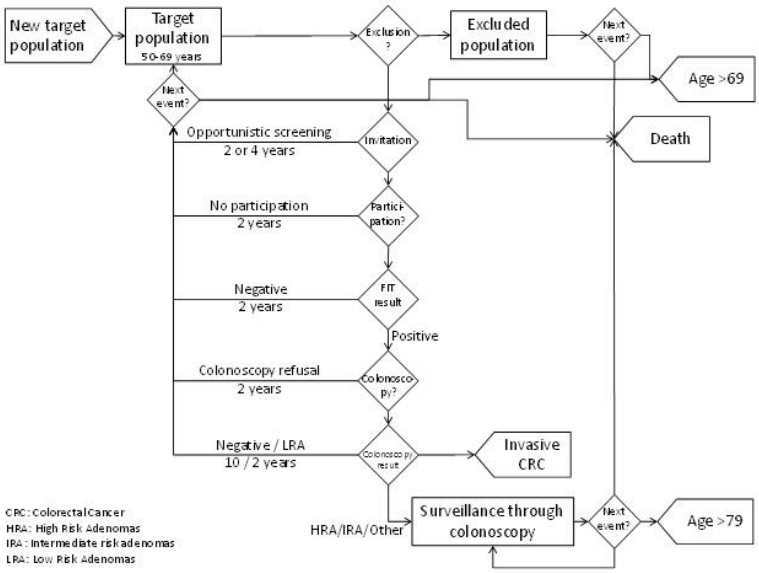
[1] See Box, G.E.P., Draper N. R. (1987)

2.1. Conceptual Model

This model has two stages; the screening process and the surveillance of adenomas.

The events that this model simulated were: Inclusion of a new person in the target population, exclusion process, invitation process, participation process, result of the FOBT test, colonoscopy after a positive FOBT test and surveillance colonoscopy. The exits of the model were detection of invasive CRC, death or exclusion from the target population. Individuals under surveillance through colonoscopy had an age limit of 80 years.

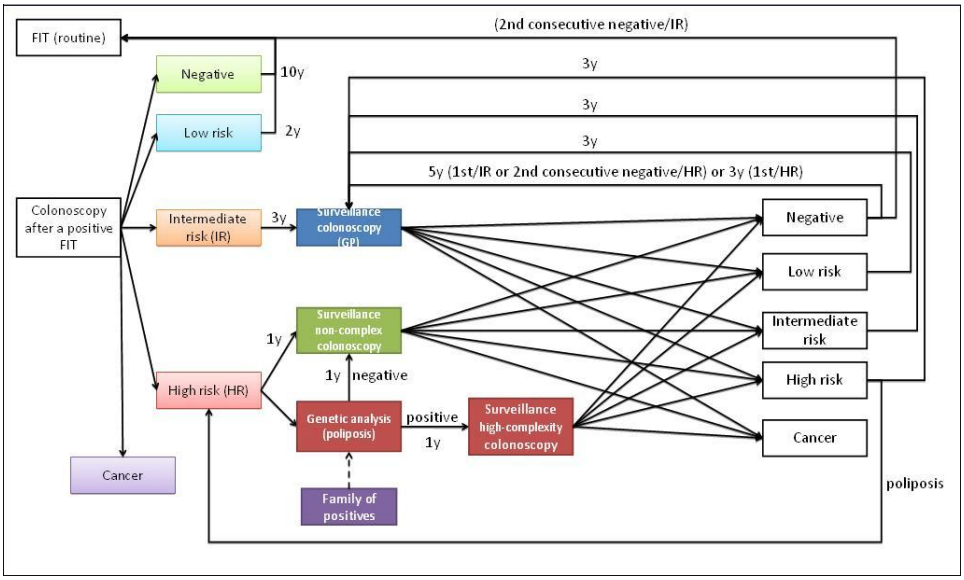
FIG. 2.1. Conceptual model of the screening process



In figure 2.1 we can see the conceptual model of the screening process. There were five different cases where individuals were invited to the program after two years: opportunistic screening through colonoscopy 5 years ago, no participation, participation with a negative FIT, a positive FIT and colonoscopy refusal or findings of low-risk adenomas.

Individuals were invited after 4 years if they had a colonoscopy (opportunistic screening) 3 years ago, while those with a negative result of the colonoscopy after a positive FIT were invited after 10 years to routine screening. Colonoscopy results of high- or intermediate-risk adenomas entered the path of surveillance through colonoscopy.

FIG. 2.2. Conceptual model of the surveillance process



The conceptual model of surveillance of adenomas is adapted from European guidelines as shown in figure 1.5, the adaptation of those criteria to our model is shown in figure 2.2. If the result of the colonoscopy after a positive FIT is LRA (Low risk adenomas), routine screening through FIT after 2 years is recommended, and begin again the screening process.

If the result of the colonoscopy after a positive FIT is IRA(Intermediate risk adenomas) or HRA (High risk adenomas) a surveillance colonoscopy is recommended after 3 and 1 years, respectively. The intensity of surveillance colonoscopies (after 1, 3 or 5 years) will depend on the result of every colonoscopy as shown in figure 2.2.

2.2. Techniques

The simulation model was implemented by using Arena (Rockwell Software) version 14.5.

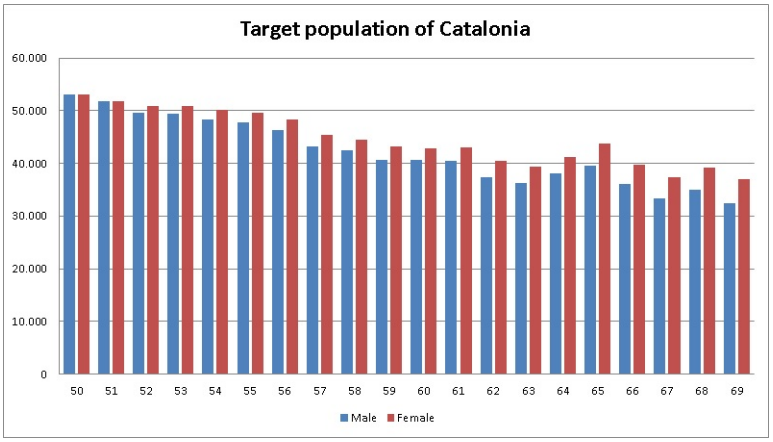
At the beginning of the simulation the target population was 100,000 men and women aged 50-69 years undergoing biennial screening. From the second year on, persons aged 50 years old entered the ¡following current predictions on Catalan population. Persons aged 68-69 were excluded after their last screening round.

A total of 1,750 replications of the model with independent streams of random numbers were run. The time units were years. A time horizon of 20 years (from 2015 to 2034) was chosen to simulate the life history of a person entering a screening program (from 50 to 69 years) and to allow long term prediction of colonoscopy demand according to the program results. Individual persons from 50 to 69 years of age were simulated. Persons aged 70 years or older were followed-up until 79 years of age only if they were having surveillance colonoscopies, if not were taken off the study. The entire population involved in the system each year was included. Thus, individual people entering and exiting the model were simulated throughout the simulation horizon.

This model takes into account the ageing of the population in the following way.

An initial population was introduced. The model starts in 2015 with the population structure of the actual target population of Catalonia which is shown in Figure 2.3, upscaling to 100,000 inhabitants.

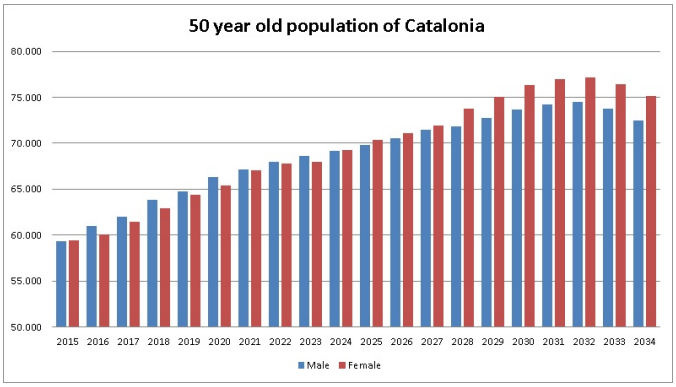
FIG. 2.3. Target population of Catalonia



Since the start of the simulation every year enters in the model the new population that turns fifty years old. From 2015-2029 these data for Catalonia can be found

in INEbase [2] , from 2029-2034 the Spanish prediction was used also of INEbase applying it to the Catalan population. This data is shown on figure 2.4.

FIG. 2.4. Prediction of 50 year old population of Catalonia



Data of the first and second rounds of a Spanish CRC Screening Program including 31 geographical areas was used to calculate the parameters related to screening. Percentage of exclusions, opportunistic screening, participation, positivity colonoscopy refusal and adherence were treated as probabilistic parameters, last one was the unique that was constant for each run. As this parameters clearly differ significantly by age groups and gender, different distributions were estimated for each parameter and for strata combining four different groups of age (50-54, 55-59, 60-64 and 65-69) and gender.

The models used to estimate those parameters were of the form:

$$A + B \cdot beta(\alpha, \beta) \; ; \; A, B \in \mathbb{R} \; ; \; \alpha, \beta > 0$$

We can see the parameters used in figure 2.5

Beta distribution was selected due to this distribution is appropriate to adjust probabilities because it take values between 0 and 1, and the A and B parameters are used to rescale it .

This values of the betas distribution used to estimate the parameters were calculated using the data of the first (2009-2011), second (2011-2013) and third (2013-2015) rounds of the Colorectal Cancer Screening Program of Barcelona and the software "Input Analyzer " of Arena.

[2] See INEbase

FIG. 2.5. Models used to estimate the parameters

	Exclusion initial screening		% of exclusions due to medical reasons	
Age group	Male	Female	Male	Female
50-54	3+17*BETA(0.855,1.77)	2+33*BETA(0.51,2.27)	11.92	12.45
55-59	3+17*BETA(0.993,1.99)	3+33*BETA(0.613,2.53)	21.04	20.37
60-64	3+17*BETA(1.09,2.34)	3+25*BETA(0.512,1.97)	32.78	28.81
65-69	3+13*BETA(1.21,1.83)	2+28*BETA(0.507,2.49)	41.27	34.65

	Opportunistic screening		% to be invited after 2 years	
Age group	Male	Female	Male	Female
50-54	0.01+3.37*BETA(2.36,3.6)	0+3.41*BETA(2.28,3.26)	40	
55-59	3.93*BETA(1.87,2.49)	0+4.95*BETA(1.46,2.06)		
60-64	0.39+3.61*BETA(1.48,1.67)	1+4.59*BETA(1.24,2.45)		
65-69	5.7*BETA(2.33,3.36)	0.47+5.16*BETA(1.97,2.58)		

	Participation initial screening		Participation successive screening	
Age group	Male	Female	Participants in the previous round	Non participants in the previous round
50-54	18+33*BETA(2.25,1.18)	31+26*BETA(2.38,1.85)	71 + 17 * BETA(1.39, 0.714)	14 + 17 * BETA(1.39, 0.714)
55-59	20+35*BETA(2.46,1.4)	36+22*BETA(1.68,1.46)		
60-64	27+31*BETA(2.07,1.27)	35+25*BETA(2.01,1.45)		
65-69	27+29*BETA(1.55,1.02)	35+21*BETA(1.72,1.36)		

	FIT positivity initial screening		FIT positivity successive screening	
Age group	Male	Female	Male	Female
50-54	3+6*BETA(1.3,1.75)	1.1+4.8*BETA(2.73,2.42)	3.31 + 2.29 * BETA(2.4, 2.26)	
55-59	3+11*BETA(2.63,3.54)	1.33+6.67*BETA(3.07,2.96)		
60-64	5+12*BETA(2.49,4.1)	2+13*BETA(2.07,4.58)		
65-69	7+11*BETA(1.39,1.86)	2+11*BETA(2.47,4)		

	Colonoscopy refusal initial screening		Colonoscopy refusal successive screening	
Age group	Male	Female	Male	Female
50-54	18+33*BETA(2.25,1.18)	31+26*BETA(2.38,1.85)	71 + 24 * BETA(2.24, 1.49)	
55-59	20+35*BETA(2.46,1.4)	36+22*BETA(1.68,1.46)		
60-64	27+31*BETA(2.07,1.27)	35+25*BETA(2.01,1.45)		
65-69	27+29*BETA(1.55,1.02)	35+21*BETA(1.72,1.36)		

TABLE 2.1. Distribution of results from colonoscopy after a positive FIT

	Initial screening (%)	Successive screening (%)
Negative	32.0	41.9
Low risk adenoma	17.1	20.4
Intermediate adenomas	28.2	25.6
High risk adenoma	16.6	9.0
Cancer	6.1	3.1

Distribution of the results of the colonoscopy after a positive FIT for initial screening were obtained from the first round of the this program. For successive screening the results were obtained from the second round, by analysing those individuals that had a negative FIT in the first round. It is shown on table 2.1

Table 2.2. Distribution of results from surveillance colonoscopy

	After high risk adenomas		After intermediate risk adenomas
	1 year	3 years	3 years
Negative	72.48	78.40	67.99
Low risk adenoma	24.95	20.41	28.27
Intermediate and high risk adenoma	2.57	0.89	3.27
Cancer	0	0.30	0.47

In table 2.2 we can see the distribution of results from surveillance colonoscopy according to follow-up time and result of the colonoscopy after the positive fecal inmunochemical test. This distribution was obtained from Winawer et al. [3], considering those labelled as "pathologically advanced adenomas"as of high or intermediate risk, the rest of adenomas as low risk,invasive cancer as is, and the rest of colonoscopies as negative. For follow-up after the first surveillance colonoscopy, results of high or intermediate risk were considered as of high risk in the scheme. For follow-up at 5 years, results of 3-years follow-up were applied according to the risk obtained in the colonoscopy after a positive FIT.

[3]See Winawer S.J., Zauber A.G., O'Brien M.J. et al (1993)

2.3. Verification and validation

For validation purposes, a subgroup of runs representing the current scenario was analysed. This group included 393 runs and was defined as mean 20-year participation between 40% and 60%, mean 20-year FIT positivity between 4.7% and 6.8% and adherence to surveillance colonoscopies between 35% and 70%. This group of 393 runs was used for the main analysis of this study.

The following results were used to validate the model: the number of invited people through time (by initial and successive screenings), the participation rate, positivity of FIT (by initial and successive screenings), the number of colonoscopies over time (by initial and successive screenings), distribution of colonoscopy findings (by initial and successive screenings), and life expectancy. Validation results were checked by the research team and the model was considered as valid, credible and useful for the purposes of the study.

2.4. Results

Table 2.3 shows that biennial screening of an initial population of 100,000 men and women led to a mean of more than 40,000 colonoscopies both after a positive FIT and for surveillance during the 20-year horizon.

TABLE 2.3. Average results of 20 years of screening

Year	Colonoscopies after positive FIT	Surveillance colonoscopies	All colonoscopies
2015	1222 [1212,1231]	0 [0,0]	1222 [1212,1231]
2016	1221 [1211,1232]	180 [179,181]	1401 [1390,1413]
2017	1119 [1108,1130]	184 [183,185]	1303 [1291,1315]
2018	1110 [1100,1120]	458 [451,461]	1568 [1551,1581]
2019	1171 [1157,1184]	626 [622,630]	1797 [1779,1814]
2020	1165 [1152,1178]	609 [604,613]	1774 [1756,1791]
2021	1178 [1166,1190]	816 [811,821]	1994 [1977,2011]
2022	1190 [1179,1201]	878 [871,885]	2068 [2050,2086]
2023	1213 [1200,1225]	869 [862,876]	2082 [2062,2101]
2024	1231 [1217,1244]	994 [987,1001]	2225 [2204,2245]
2025	1246 [1233,1260]	1007 [999,1015]	2253 [2232,2275]
2026	1274 [1259,1289]	1051 [1042,1059]	2325 [2301,2348]
2027	1303 [1288,1317]	1082 [1074,1090]	2385 [2362,2407]
2028	1319 [1304,1333]	1084 [1074,1094]	2403 [2378,2427]
2029	1332 [1317,1346]	1117 [1107,1126]	2449 [2424,2472]
2030	1327 [1313,1342]	1138 [1128,1148]	2465 [2441,2490]
2031	1366 [1351,1381]	1198 [1186,1209]	2564 [2537,1339]
2032	1373 [1358,1388]	1209 [1197,1121]	2582 [2555,2509]
2033	1385 [1369,1401]	1209 [1198,1220]	2594 [2567,2621]
2034	1376 [1361,1391]	1238 [1227,1249]	2614 [2588,2640]
Total	**25121**	**16947**	**42068**

TABLE 2.4. Average participation and positivity of 20 years of screening

Year	Participation	Positivity
2015	43.9 %	6.7 %
2016	43.8 %	6.7 %
2017	47.4 %	5.3 %
2018	47.4 %	5.3 %
2019	49.1 %	5.3 %
2020	49.0 %	5.2 %
2021	49.7 %	5.1 %
2022	49.7 %	5.1 %
2023	50.1 %	5.0%
2024	50.1 %	5.1%
2025	50.3 %	5.0%
2026	50.3 %	5.0%
2027	50.4 %	5.0%
2028	50.4 %	4.9%
2029	50.5 %	4.9%
2030	50.6 %	4.9%
2031	50.6 %	4.9%
2032	50.7 %	4.9%
2033	50.7 %	4.9%
2034	50.8 %	4.9%

Yearly results show an increase in the number of surveillance colonoscopies, although positivity, as a percentage, decreased through time. This colonoscopies had a mean adherence of 52.058%.

FIG. 2.6. Participation and positivity through time

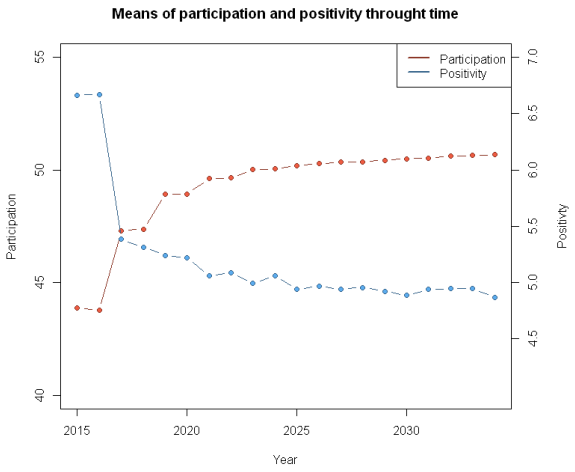
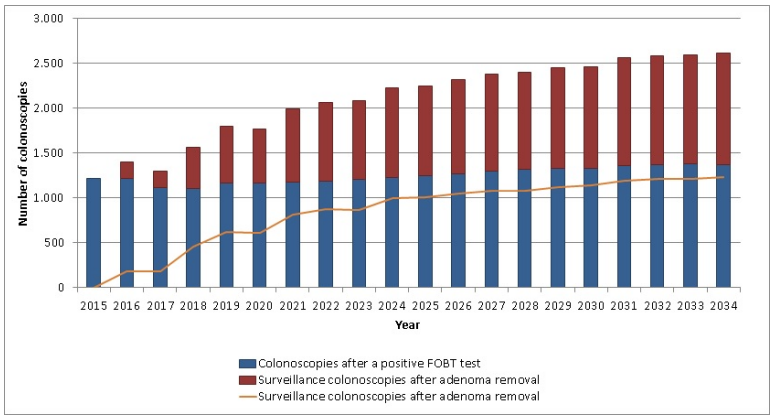


Figure 2.6 and table 2.4 shows variations in mean participation and mean positivity through time. The first two years represent the first round of the program and show higher positivity and lower participation because all participants are of initial screening. Afterwards, the increasing number of successive screenings impacts on an increasing participation rate and a decreasing positivity, both stabilizing at the long term.

FIG. 2.7. Number of colonoscopies through time



The main outcome of this study is the number of colonoscopies needed for both the screening program and the surveillance of non-cancer findings.

Figure 2.7 shows in stacked bars the number of colonoscopies: below, those after a positive FIT, above, the number of surveillance colonoscopies, which show a sharper increase beginning at the second year. Overall number of colonoscopies doubled after 14 years (2028).

In addition the number of surveillance colonoscopies through time is also shown by the orange line, facilitating comparison between the number of colonoscopies according to their type. The result is that the number of colonoscopies after a positive FIT is always higher than the number of surveillance colonoscopies, but over time this difference is smaller each year, and long-term this two numbers are really close.

Chapter 3
Sensitivity analysis

As the man said, for every complex problem
there is a simple solution,
and it's wrong.
- Umberto Eco [1]

In order to assess how the uncertainty associated with the estimation of the participation, FIT positivity and adherence to surveillance colonoscopies affects the results of this simulation model, a sensitivity analysis was performed.

The first idea was to create two mixed effects models using the number of colonoscopies after a positive FIT and the number of surveillance colonoscopies as response variables to see the effect of an increase or decrease of those three variables through time. This models were selected because they can represent the variability among simulations placing a random effect in the intercept, and the variability through time adding a random effect on the time.

Literature was consulted and any article that apply this technique for a sensitivity analysis was found, so it could be an innovation in terms of techniques to consider. However, it turned out that with these data these models were not able to be used, as any model found verified the hypothesis of the models, specifically the hypothesis of normality of the errors, neither changing the model nor applying transformations to the variables. So it was decided to adapt the sensitivity analysis using a linear model for each year and each type of colonoscopy (to a total of 40 models), which met its purpose and that verified the previous assumptions.

[1]See Eco U. (1989)

3.1. Mixed Effects Models (ME)

A mixed effects model is a statistical model containing both fixed effects and random effects. They are particularly useful in settings where repeated measurements are made on the same statistical units (longitudinal study), or where measurements are made on clusters of related statistical units.

In matrix notation a mixed model can be represented as

$$y = X\beta + Zb + \epsilon$$

where

- y is a known vector of observations, with mean $E(y) = X\beta$;
- β is an unknown vector of fixed effects;
- b is an unknown vector of random effects, with mean $E(b) = 0$ and variance-covariance matrix $var(b) = G$;
- ϵ is an unknown vector of random errors, with mean $E(\epsilon) = 0$ and variance $var(\epsilon)$=R;
- X and Z are known design matrices relating the observations y to β and b, respectively.

It has several assumptions that must be met

- **Linearity.** The regression function is linear.
- **Homocedasticity.** The errors variance is constant.
- **Normality of errors.** The errors are normally distributed.
- **Independence of errors.** The random variables representing errors "$\epsilon_1, \cdots, \epsilon_n$" are mutually independent.
- **Normality of random effects.** The random effects are normally distributed.

Usually it is used to model several individuals with several measures on a variable, in this case we will use them as if each simulation its an individual and we take 20 measures, one per year, that will be the means of the variables participation, positivity and adherence of the run. The objective is to model the number of screening or surveillance colonoscopies made that year in terms of the variables adherence, positivity and participation. We have estimated the parameters with the R software version 3.2.2.

3.1.1. ME model for the number of colonoscopies after a positive FIT.

For the number of colonoscopies after a positive FIT, the model used has random effects at the intercept and at the variable time and fixed effects at the intercept, the variable time, positivity, participation and both interactions between positivity and participation with time, to find out whether the effect of positivity and participation is different through time, because we understand that the effect of one of this variables over the response variable will depend of the value of the other one.

The results are:

Intercept: $\beta_0 = -957.24$

Coefficient of time: $\beta_1 = -5.58$

Coefficient of positivity: $\beta_2 = 172.93$

Coefficient of participation: $\beta_3 = 22.62$

Coefficient of interaction between time and positivity: $\beta_{12} = 3.35$

Coefficient of interaction between time and participation: $\beta_{13} = 0.12$

then:

$$N_t = -957.24 + b_{0i} + (-5.58 + b_{1i}) \cdot t + 172.93 \cdot PO + 22.62 \cdot PA + 3.35 \cdot PO \cdot t + 0.12 \cdot PA \cdot t$$

Being:

Number of colonoscopies after a positive FIT at time t: N_t

Years since start of simulation: $t \in (1, 20)$

Positivity: PO

Participation: PA

$$\text{Random effects: } b_i = \begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix} \sim N_2(0, D)$$

$$D = \begin{pmatrix} 3689.60 & 87.36 \\ 87.36 & 3.23 \end{pmatrix}$$

On figure 3.1 we can see the distribution of the random effects

In this sensitivity analysis we want to estimate the effect of an increase of a 1% on positivity or participation on the response variable. Thus, we will be interested on this variables:

- **Increase of the number of colonoscopies after a positive FIT** for a 1% increase in positivity:

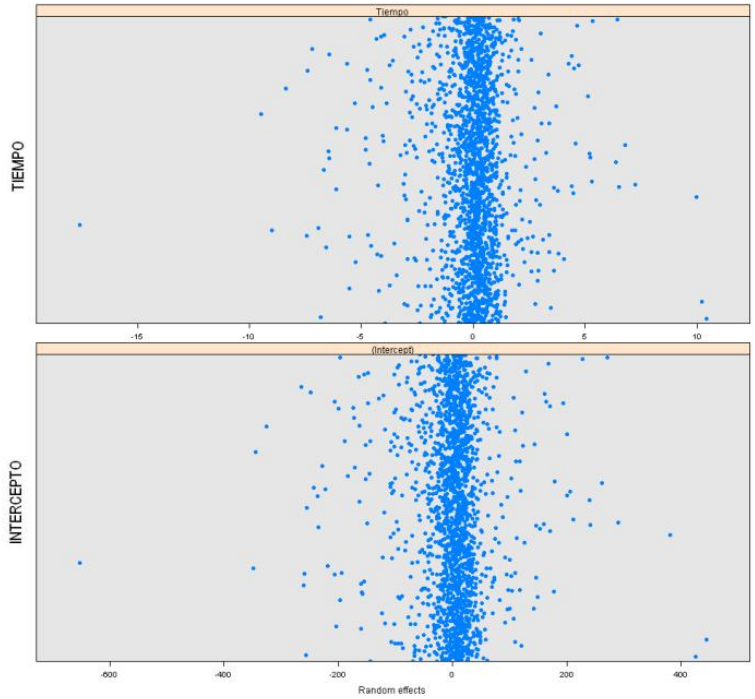
$$N = \cancel{\beta_0 + b_{0i} + (\beta_1 + b_{1i}) \cdot t} + \beta_2 \cdot (x+1) + \cancel{\beta_3 \cdot PA} + \beta_{12} \cdot (x+1) \cdot t + \cancel{\beta_{13} \cdot PA \cdot t} -$$

$$- \cancel{(\beta_0 + b_{0i} + (\beta_1 + b_{1i}) \cdot t} + \beta_2 \cdot x + \cancel{\beta_3 \cdot PA} + \beta_{12} \cdot x \cdot t + \cancel{\beta_{13} \cdot PA \cdot t)} =$$

$$= \cancel{\beta_2 \cdot x} + \beta_2 + \cancel{\beta_{12} \cdot x \cdot t} + \beta_{12} \cdot t - \cancel{\beta_2 \cdot x} - \cancel{\beta_{12} \cdot x \cdot t} =$$

$$= \beta_2 + \beta_{12} \cdot t$$

FIG. 3.1. Random effects of the mixed effects model



- **Increase of the number of colonoscopies after a positive FIT** for a 1% increase in participation:

$$N = \cancel{\beta_0 + b_{0i} + (\beta_1 + b_{1i}) \cdot t + \beta_2 \cdot PO + \beta_3 \cdot (x+1) + \beta_{12} \cdot PO \cdot t} + \beta_{13} \cdot (x+1) \cdot t -$$

$$\cancel{-(\beta_0 + b_{0i} + (\beta_1 + b_{1i}) \cdot t + \beta_2 \cdot PO + \beta_3 \cdot x + \beta_{12} \cdot PO \cdot t} + \beta_{13} \cdot x \cdot t =$$

$$= \cancel{\beta_3 \cdot x} + \beta_3 + \cancel{\beta_{13} \cdot x \cdot t} + \beta_{13} \cdot t - \cancel{\beta_3 \cdot x} - \cancel{\beta_{13} \cdot x \cdot t} =$$

$$= \beta_3 + \beta_{13} \cdot t$$

As we have the correlations we can compute the confidence intervals as follows:

$$IC_{95\%}(\beta_2 + t \cdot \beta_{12}) = \beta_2 + t \cdot \beta_{12} \pm 1.96 \cdot (\frac{Var(\beta_2 + t \cdot \beta_{12})}{n})^{\frac{1}{2}} =$$

$$= \beta_2 + t \cdot \beta_{12} \pm 1.96 \cdot (\frac{Var(\beta_2)}{n} + \frac{2 \cdot t \cdot Cov(\beta_2, \beta_{12})}{n} + t^2 \cdot \frac{Var(\beta_{12})}{n})^{\frac{1}{2}}$$

$$IC_{95\%}(\beta_3 + t \cdot \beta_{13}) = \beta_3 + t \cdot \beta_{13} \pm 1.96 \cdot (\frac{Var(\beta_3 + t \cdot \beta_{13}}{n})^{\frac{1}{2}} =$$

$$= \beta_3 + t \cdot \beta_{13} \pm 1.96 \cdot (\frac{Var(\beta_3)}{n} + \frac{2 \cdot t \cdot Cov(\beta_3, \beta_{13})}{n} + t^2 \cdot \frac{Var(\beta_{13})}{n}))^{\frac{1}{2}}$$

TABLE 3.1. Sensitivity analysis with the ME Model

Year	Increase of colonoscopies after a positive FIT for an increase of a 1% in positivity [IC95%]	Increase of colonoscopies after a positive FIT for an increase a 1% in participation [IC95%]
2015	176.3 [173.8,178.7]	22.8 [22.5,23.0]
2016	179.6 [177.1,182.2]	22.9 [22.6,23.1]
2017	183.0 [180.3,185.7]	23.0 [22.8,23.3]
2018	186.3 [183.6,189.1]	23.1 [22.9,23.4]
2019	189.7 [186.9,192.5]	23.3 [23.0,23.5]
2020	193.0 [190.2,195.9]	23.4 [23.1,23.6]
2021	196.4 [193.5,199.3]	23.5 [23.3,23.7]
2022	199.8 [196.9,202.6]	23.6 [23.4,23.9]
2023	203.1 [200.3,205.9]	23.8 [23.5,24.0]
2024	206.5 [203.7,209.2]	23.9 [23.6,24.1]
2025	209.8 [207.1,212.5]	24.0 [23.8,24.2]
2026	213.2 [210.6,215.7]	24.1 [23.9,24.4]
2027	216.5 [214.1,219.0]	24.2 [24.0,24.5]
2028	219.9 [217.6,222.2]	24.4 [24.1,24.6]
2029	223.2 [221.0,225.4]	24.5 [24.2,24.8]
2030	226.6 [224.5,228.7]	24.6 [24.3,24.9]
2031	229.9 [228.0,231.9]	24.7 [24.5,25.0]
2032	233.3 [231.4,235.2]	24.9 [24.6,25.1]
2033	236.6 [234.9,238.4]	25.0 [24.7,25.3]
2034	240.0 [238.3,241.7]	25.1 [24.8,25.4]
2035	243.3 [241.7,244.9]	25.2 [24.9,25.5]

As we can see for table 3.1, a change in 1% in positivity changes substantially the number of colonoscopies after a positive FIT per 100,000 inhabitants, and this change increases over time. It will increase in 176.3 colonoscopies in 2015, 209.8 in 2024 and 240.0 in 2034.

On the other hand an increase of a 1% of the participation will increase in almost 23 the number of colonoscopies after a positive FIT in the first year, showing a slight increase to 25 colonoscopies in 2034.

3.1.2. Validation of Mixed effects models.

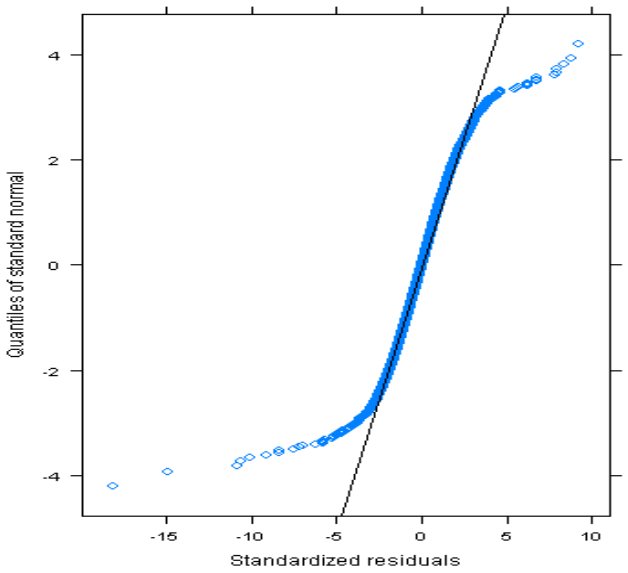
To validate if this model is correct, we will we will check the assumptions of the mixed effects model. As it is shown in the qqplot, figure 3.2, the residuals of our model are not normally distributed. Let's compute a normality test to verify it.

```
> lillie.test(lme.1$residuals[,1])

Lilliefors (Kolmogorov-Smirnov) normality test

data:  lme.1$residuals[, 1]
D = 0.054787, p-value < 2.2e-16
```

FIG. 3.2. QQplot of the mixed effects model residuals



As supposed by the qqplot, the residuals of this model are not normally distributed, hence, this model is not valid. A great number of transformations (square root, logarithm, box cox, $\cdots$) where tried to see if normalizing the data we could fix this problem, but it was impossible, so it arises the need of using another type of models, in this case the simple linear regression.

3.2. General Linear Model (LM)

In general terms the general linear model is similar to the mixed effects one, but in the LM we use fixed effects only instead of fixed and random.

In matrix notation a general linear model can be represented as

$$y = X\beta + \epsilon$$

where

- y is a known vector of observations, with mean $E(y) = X\beta$;
- β is an unknown vector of fixed effects;
- ϵ is an unknown vector of random errors, with mean $E(\epsilon) = 0$ and variance $var(\epsilon)$=R;
- X is a known design matrix relating the observations y to β.

It has 4 assumptions that must be met

- **Linearity.** The regression function is linear.
- **Homocedasticity.** The error variance is constant.
- **Normality.** The error is normally distributed.
- **Independence.** The random variables representing errors "$\epsilon_1, \cdots, \epsilon_n$" are mutually independent.

Thus, 40 models where computed, one for each year and each type of colonoscopy.

3.2.1. Advantages and disadvantages of LM vs ME.

These models have advantages and disadvantages with respect to the mixed effects models.

- Disadvantages
  - This models are simpler than the mixed effects models and do not use random effects, so they do not take into account the variability within subjects, in this case between simulations.
  - Instead of computing 2 models we have to compute 40
- Advantages
  - Doing one model for each year makes that the prediction is much more accurate.
  - Despite of computing 20 times more models, the runtime of computing this models is substantially lower

3.2.2. Linear Model for the number of colonoscopies after positive FIT.

Let:

$NScreening$: Number of colonoscopies after a positive FIT

β_0: Intercept

PA: Participation

β_1: Coefficient of participation:

PO: Positivity

β_2: Coefficient of positivity

β_{12}: Coefficient of the interaction between positivity and participation:

We will adjust models with the following structure:

$$NScreening = \beta_0 + PA \cdot \beta_1 + PO \cdot \beta_2 + PA \cdot PO \cdot \beta_{12}$$

In the same way than in mixed effects model our estimate of the increase in the number of colonoscopies after a positive FIT for a 1% increase in positivity with a participation of PA will be:

$$\beta_2 + PA \cdot \beta_{12}$$

Which variance will be:

$$Var(\beta_2 + PA \cdot \beta_{12}) = Var(\beta_2) + PA^2 \cdot Var(\beta_{12}) + 2 \cdot PA \cdot Cov(\beta_2, \beta_{12})$$

and confidence intervals:

$$IC_{95\%} = \beta_2 + PA \cdot \beta_{12} \pm 1.96 \cdot (Var(\beta_2 + PA \cdot \beta_{12})/n)^{\frac{1}{2}}$$

On the other hand our estimate of the increase in the number of colonoscopies after a positive FIT for a 1% increase in participation with a positivity of PO will be:

$$\beta_1 + PO \cdot \beta_{12}$$

Which variance will be:

$$Var(\beta_1 + PO \cdot \beta_{12}) = Var(\beta_1) + PO^2 \cdot Var(\beta_{12}) + 2 \cdot PO \cdot Cov(\beta_1, \beta_{12})$$

and confidence intervals:

$$IC_{95\%} = \beta_1 + PO \cdot \beta_{12} \pm 1.96 \cdot (Var(\beta_1 + PO \cdot \beta_{12})/n)^{\frac{1}{2}}$$

In the table 3.2 we can see the values of this coefficients for the 20 models

TABLE 3.2. Coefficients of regression

Year	β_0	β_1	β_2	β_{12}
2015	-6	0.3673	1.7904	4.0849
2016	25.5827	-0.4208	-4.5877	4.272
2017	-40.6472	0.9674	6.9964	4.2118
2018	-61.3547	1.4881	13.7298	4.0874
2019	-16.7071	0.4694	4.6533	4.4371
2020	16.5079	0.1905	-1.8505	4.5016
2021	-1.7479	0.7778	6.9669	4.387
2022	-64.5311	2.0108	18.0132	4.1977
2023	-13.7993	1.1154	9.6662	4.4718
2024	-42.0731	1.9162	14.82	4.3627
2025	65.5724	-0.4867	-4.9494	4.9344
2026	8.0693	1.0353	9.1515	4.6409
2027	28.2684	0.4076	1.4362	4.9709
2028	-14.0557	1.3078	9.7122	4.8464
2029	10.1986	1.1901	7.744	4.9061
2030	-22.0543	1.524	16.0019	4.8157
2031	5.5452	1.0709	10.4677	4.9979
2032	-44.2581	2.115	19.2061	4.8257
2033	-0.1785	1.1892	9.3565	5.0821
2034	46.219	0.3501	1.7562	5.236

In the tables 3.3 and 3.4 we can see the increase of colonoscopies after a positive FIT when we change the variables, through time and for a population of 100,000 inhabitants.

TABLE 3.3. Increase of colonoscopies after a positive FIT conditioned on positivity

YEAR	Increase of colonoscopies after positive FIT for an increase of a 1% in positivity with a participation of 40%	Increase of colonoscopies after positive FIT for an increase of a 1% in positivity with a participation of 50%	Increase of colonoscopies after positive FIT for an increase of a 1% in positivity with a participation of 60%
2015	165.2 [165.0,165.3]	206.0 [205.9,206.1]	246.9 [246.8,247.0]
2016	166.3 [166.1,166.4]	209.0 [208.9,209.1]	251.7 [251.6,251.8]
2017	175.5 [175.3,175.6]	217.6 [217.5,217.7]	259.7 [259.6,259.8]
2018	177.2 [177.1,177.4]	218.1 [218.0,218.2]	259.0 [258.9,259.1]
2019	182.1 [182.0,182.3]	226.5 [226.4,226.6]	270.9 [270.8,271.0]
2020	178.2 [178.1,178.4]	223.2 [223.1,223.4]	268.2 [268.1,268.3]
2021	182.4 [182.3,182.6]	226.3 [226.2,226.4]	270.2 [270.1,270.3]
2022	185.9 [185.8,186.1]	227.9 [227.8,228]	269.9 [269.8,270.0]
2023	188.5 [188.4,188.7]	233.3 [233.1,233.4]	278.0 [277.9,278.1]
2024	189.3 [189.2,189.5]	233.0 [232.8,233.1]	276.6 [276.5,276.7]
2025	192.4 [192.3,192.6]	241.8 [241.6,241.9]	291.1 [291.0,291.2]
2026	194.8 [194.6,194.9]	241.2 [241.1,241.3]	287.6 [287.5,287.7]
2027	200.3 [200.1,200.4]	250.0 [249.9,250.1]	299.7 [299.6,299.8]
2028	203.6 [203.4,203.7]	252.0 [251.9,252.2]	300.5 [300.4,300.6]
2029	204.0 [203.8,204.1]	253.0 [252.9,253.2]	302.1 [302.0,302.2]
2030	208.6 [208.5,208.8]	256.8 [256.7,256.9]	304.9 [304.8,305.0]
2031	210.4 [210.2,210.5]	260.4 [260.2,260.5]	310.3 [310.2,310.4]
2032	212.2 [212.1,212.4]	260.5 [260.4,260.6]	308.7 [308.6,308.8]
2033	212.6 [212.5,212.8]	263.5 [263.3,263.6]	314.3 [314.2,314.4]
2034	211.2 [211.0,211.4]	263.6 [263.4,263.7]	315.9 [315.8,316.0]

TABLE 3.4. Increase of colonoscopies after a positive FIT conditioned on participation

YEAR	Increase of colonoscopies after positive FIT for an increase of a 1% in participation with a positivity of 4%	Increase of colonoscopies after positive FIT for an increase of a 1% in participation with a positivity of 5%	Increase of colonoscopies after positive FIT for an increase of a 1% in participation with a positivity of 6%
2015	16.7 [15.4,18.0]	20.8 [19.7,21.8]	24.9 [24.0,25.8]
2016	16.7 [15.4,18.0]	20.9 [19.9,22.0]	25.2 [24.3,26.1]
2017	17.8 [16.4,19.2]	22.0 [20.9,23.2]	26.2 [25.3,27.2]
2018	17.8 [16.5,19.2]	21.9 [20.8,23.1]	26.0 [25.1,27.0]
2019	18.2 [16.9,19.6]	22.7 [21.5,23.8]	27.1 [26.2,28.0]
2020	18.2 [16.8,19.5]	22.7 [21.6,23.8]	27.2 [26.3,28.1]
2021	18.3 [16.9,19.7]	22.7 [21.5,23.9]	27.1 [26.1,28.1]
2022	18.8 [17.5,20.1]	23.0 [21.9,24.1]	27.2 [26.3,28.1]
2023	19.0 [17.6,20.4]	23.5 [22.3,24.6]	27.9 [27.0,28.9]
2024	19.4 [18.0,20.7]	23.7 [22.6,24.9]	28.1 [27.1,29.0]
2025	19.3 [17.9,20.6]	24.2 [23.1,25.3]	29.1 [28.2,30.1]
2026	19.6 [18.2,21.0]	24.2 [23.1,25.3]	28.9 [27.9,29.8]
2027	20.3 [18.9,21.6]	25.3 [24.2,26.4]	30.2 [29.3,31.2]
2028	20.7 [19.4,22.0]	25.5 [24.4,26.6]	30.4 [29.5,31.3]
2029	20.8 [19.5,22.1]	25.7 [24.6,26.8]	30.6 [29.7,31.5]
2030	20.8 [19.4,22.1]	25.6 [24.5,26.7]	30.4 [29.5,31.3]
2031	21.1 [19.7,22.4]	26.1 [25.0,27.1]	31.1 [30.2,32.0]
2032	21.4 [20.1,22.8]	26.2 [25.1,27.4]	31.1 [30.1,32.0]
2033	21.5 [20.2,22.8]	26.6 [25.5,27.7]	31.7 [30.8,32.6]
2034	21.3 [20.0,22.6]	26.5 [25.4,27.6]	31.8 [30.8,32.7]

As we can see an increase of a 1% in the positivity increases substantially the number of colonoscopies after a positive FIT, and this increase is higher through time, and higher as the participation increases

On the other hand we can see that a increase of a 1% in the participation increases the number of colonoscopies, but this changes are more or less 10 times smaller than in terms of Positivity, so we can see that more or less increasing the positivity in a 1% has the same effect that increasing participation in a 10%.

3.2.3. Linear Model for the number of surveillance colonoscopies.

As the model for the number of surveillance colonoscopies is more complex, we would use the backward method to create the models. We will begin with all the possible variables and interactions and for each time we will remove the less significant variable until we have the model with the minimum AIC and which verify all the assumptions of the linear model. Let:

$NSurvei$: Number of surveillance colonoscopies

β_0: Intercept

PA: Participation

β_1: Coefficient of participation

PA^2: Squared participation

β_{11}: Coefficient of participation to square

PO: Positivity

β_2: Coefficient of positivity

PO^2: Squared positivity

β_{22}: Coefficient of positivity to square

AD: Adherence

β_3: Coefficient of adherence

β_{12}: Coefficient of the interaction between participation and positivity

All our models have the following structure, but some will not have significance in all the covariables. The interaction between positivity and adherence and the interaction between adherence and participation have no sense (are not reflected in the model we created). Moreover, the square of the adherence is not included because it was not significant in either model.

$$NSurvei = \beta_0 + PA \cdot \beta_1 + PO \cdot \beta_2 + AD \cdot \beta_3 + PA^2 \cdot \beta_{11} + \\ + PO^2 \cdot \beta_{22} + PA \cdot PO \cdot \beta_{12}$$

Our estimate of the increase of the number of surveillance colonoscopies when the positivity increases from x to x+1 % with a participation PA will be:

$$\cancel{\beta_0} + \cancel{PA \cdot \beta_1} + (x+1) \cdot \beta_2 + \cancel{AD \cdot \beta_3} + \cancel{PA^2 \cdot \beta_{11}} + (x+1)^2 \cdot \beta_{22} + \\ + PA \cdot (x+1) \cdot \beta_{12} \\ - \cancel{\beta_0} - \cancel{PA \cdot \beta_1} - x \cdot \beta_2 - \cancel{AD \cdot \beta_3} - \cancel{PA^2 \cdot \beta_{11}} - x^2 \cdot \beta_{22} - \\ - PA \cdot x \cdot \beta_{12} =$$

$$= \beta_2 + (2x+1) \cdot \beta_{22} + PA \cdot \beta_{12}$$

which variance will be:

$$Var(\beta_2 + (2x+1) \cdot \beta_{22} + PA \cdot \beta_{12}) =$$

$$= Var(\beta_2) + Var((2x+1) \cdot \beta_{22}) + Var(PA \cdot \beta_{12}) + Cov(\beta_2, (2x+1) \cdot \beta_{22}) + Cov(\beta_2, PA \cdot \beta_{12}) + \\ + Cov((2x+1) \cdot \beta_{22}, PA \cdot \beta_{12}) =$$

$$= Var(\beta_2) + (2x+1)^2 \cdot Var(\beta_{22}) + PA^2 \cdot Var(\beta_{12}) + (2x+1) \cdot Cov(\beta_2, \beta_{22}) + PA \cdot Cov(\beta_2, \beta_{12}) \\ + PA \cdot (2x+1) \cdot Cov(\beta_{22}, \beta_{12})$$

and so the following confidence intervals:

$$IC_{95\%} = \beta_2 + (2x+1) \cdot \beta_{22} + PA \cdot \beta_{12} \pm 1.96 \cdot (Var(\beta_2 + (2x+1) \cdot \beta_{22} + PA \cdot \beta_{12})/n)^{\frac{1}{2}}$$

The values of this coefficients for the 20 models are shown on table 3.5

TABLE 3.5. Coefficients of regression

Year	β_0	β_1	β_2	β_{12}	β_3	β_{11}	β_{22}
2016	-102.95	4.10	15.31	0	-0.05	0	0
2017	74.96	0.27	-9.01	0.55	0.08	0	0
2018	195.90	0.39	-28.33	1.56	-0.01	0	0
2019	154.6	0	0	0	0.47	0.1048	5.67
2020	163.99	1.12	-31.40	1.97	0.16	0	0
2021	156.46	2.1	-31.66	2.65	1.08	0	0
2022	42.06	4.11	-16.12	2.36	1.96	0	0
2023	-454.37	15.74	87.45	0	1.37	0	0
2024	125.1	0	0	0	2.86	0.17	9.49
2025	-672.69	18.73	117.65	0	2.91	0	0
2026	99.23	0	0	0	3.27	0.17	10.95
2027	-58.38	4.78	-17.9066	3.2	4.35	0	0
2028	-44.4	4.72	-23.87	3.14	4.29	0	0
2029	141.69	0.41	-60.81	4.08	4.48	0	0
2030	-859.87	20.75	139.27	0	4.63	0	0
2031	30.66	3.06	-42.35	3.81	4.93	0	0
2032	-891.88	21.65	146.15	0	5.14	0	0
2033	55.48	2.31	-48.09	4.01	5.09	0	0
2034	-982.63	22.67	150.79	0	5.81	0	0

TABLE 3.6. Increase in the number of surveillance colonoscopies by increasing positivity

YEAR	Increase of surveillance colonoscopies for an increase from 4% to 5% in positivity with a participation of 40%	Increase of surveillance colonoscopies for an increase from 4% to 5% in positivity with a participation of 50%	Increase of surveillance colonoscopies for an increase from 4% to 5% in positivity with a participation of 60%
2016	15.3 [13.8,16.8]	15.3 [13.8,16.8]	15.3 [13.8,16.8]
2017	13.1 [12.7,13.6]	18.7 [18.3,19.1]	24.2 [23.9,24.6]
2018	34.1 [33.9,34.3]	49.7 [49.5,49.9]	65.3 [65.1,65.5]
2019	51.0 [50.3,51.8]	51.0 [50.3,51.8]	51.0 [50.3.51.8]
2020	47.5 [47.3,47.7]	67.3 [67.1,67.4]	87.0 [86.8,87.1]
2021	74.2 [74.0,74.4]	100.7 [100.5,100.9]	127.2 [127.0,127.3]
2022	78.4 [78.3,78.6]	102.1 [101.9,102.2]	125.7 [125,6.125.9]
2023	87.4 [86.9,87.9]	87.4 [86.9,87.9]	87.4 [86.9,87.9]
2024	85.4 [84.9,85.9]	85.4 [84.9,85.9]	85.4 [84.9,85.9]
2025	117.6 [117.2,118.1]	117.6 [117.2,118.1]	117.6 [117.2,118.1]
2026	98.6 [98.1,99.1]	98.6 [98.1,99.1]	98.6 [98.1,99.1]
2027	102.8 [102.7,103.0]	133.0 [132.9,133.1]	163.2 [163.1,163.3]
2028	101.9 [101.8,102.0]	133.4 [133.2,133.5]	164.8 [164.7,164.9]
2029	102.6 [102.5,102.7]	143.4 [143.3,143.6]	184.3 [184.2,184.4]
2030	139.3 [138.9,139.7]	139.3 [138.9,139.7]	139.3 [138.9,139.7]
2031	110.2 [110.1,110.4]	148.4 [148.2,148.5]	186.5 [186.4,186.6]
2032	146.1 [145.7,146.5]	146.1 [145.7,146.5]	146.1 [145.7,146.5]
2033	112.8 [112.7,112.9]	153.0 [152.9,153.1]	193.2 [193.1,193.3]
2034	150.0 [149.6,150.4]	150.0 [149.6,150.4]	150.0 [149.6,150.4]

Tables 3.6 and 3.7 show the increase of the number of surveillance colonoscopies through time for an increase in positivity of 1% and conditioned on 3 different levels of participation. This tables are almost equal because they just change in the years which model has a non-negative value in the coefficient of the participation and the positivity to the square.

As we can see this increase is, in general, higher through time and in addition is higher as long as the value of the participation increase due to the positive value of the interaction between positivity and participation.

TABLE 3.7. Increase in the number of surveillance colonoscopies by increasing positivity

YEAR	Increase of surveillance colonoscopies for an increase from 5% to 6% in positivity with a participation of 40%	Increase of surveillance colonoscopies for an increase from 5% to 6% in positivity with a participation of 50%	Increase of surveillance colonoscopies for an increase from 5% to 6% in positivity with a participation of 60%
2016	15.3 [13.8,16.8]	15.3 [13.8,16.8]	15.3 [13.8,16.8]
2017	13.1 [12.7,13.6]	18.7 [18.3,19.1]	24.2 [23.9,24.6]
2018	34.1 [33.9,34.3]	49.7 [49.5,49.9]	65.3 [65.1,65.5]
2019	62.4 [61.8,63.0]	62.4 [61.8,63.0]	62.4 [61.8,63.0]
2020	47.5 [47.3,47.7]	67.3 [67.1,67.4]	87.0 [86.8,87.1]
2021	74.2 [74.0,74.4]	100.7 [100.5,100.9]	127.2 [127.0,127.3]
2022	78.4 [78.3,78.6]	102.1 [101.9,102.2]	125.7 [125.6,125.9]
2023	87.4 [86.9,87.9]	87.4 [86.9,87.9]	87.4 [86.9,87.9]
2024	104.4 [103.9,104.9]	104.4 [103.9,104.8]	104.4 [103.9,104.8]
2025	117.6 [117.2,118.1]	117.6 [117.2,118.1]	117.6 [117.2,118.1]
2026	120.5 [120.0,120.9]	120.5 [120.0,120.9]	120.5 [120.0,120.9]
2027	102.8 [102.7,103.0]	133.0 [132.9,133.1]	163.2 [163.1,163.3]
2028	101.9 [101.8,102.0]	133.4 [133.2,133.5]	164.8 [164.7,164.9]
2029	102.6 [102.5,102.7]	143.4 [143.3,143.6]	184.3 [184.2,184.4]
2030	139.3 [138.9,139.7]	139.3 [138.9,139.7]	139.3 [138.9,139.7]
2031	110.2 [110.1,110.4]	148.4 [148.2,148.5]	186.5 [186.4,186.6]
2032	146.1 [145.7,146.5]	146.1 [145.7,146.5]	146.1 [145.7,146.5]
2033	112.8 [112.7,112.9]	153.0 [152.9,153.1]	193.2 [193.1,193.3]
2034	150.0 [149.6,150.4]	150.0 [149.6,150.4]	150.0 [149.6,150.4]

In the same way we can estimate of the increase of the number of surveillance colonoscopies when the participation increases from x to x+10 % with a positivity PO. In this case it will be by 10% and not by 1% because a 1% of PA is not relevant for the magnitude of the variable, as we seen in the models for the number of colonoscopies after a positive FIT.

$$\begin{gathered}\cancel{\beta_0} + (x+10)\cdot\beta_1 + \cancel{PO\cdot\beta_2} + \cancel{AD\cdot\beta_3} + (x+10)^2\cdot\beta_{11} + \cancel{PO^2\cdot\beta_{22}} + \\ +(x+10)\cdot PO\cdot\beta_{12} \\ -\cancel{\beta_0} - x\cdot\beta_1 - \cancel{PO\cdot\beta_2} - \cancel{AD\cdot\beta_3} - x^2\cdot\beta_{11} - \cancel{PO^2\cdot\beta_{22}} - \\ -x\cdot PO\cdot\beta_{12} =\end{gathered}$$

$$= 10 \cdot \beta_1 + (20x + 100) \cdot \beta_{11} + 10 \cdot PO \cdot \beta_{12}$$

which variance, calculated as previously, is:

$$Var(10 \cdot \beta_1 + (20x + 100) \cdot \beta_{11} + 10 \cdot PO \cdot \beta_{12}) =$$
$$= 100 \cdot Var(\beta_1) + (20x + 100)^2 \cdot Var(\beta_{11}) + 100 \cdot PO^2 \cdot Var(\beta_{12}) +$$
$$+100 \cdot (20x + 100)^2 \cdot Cov(\beta_1, \beta_{11}) + 10,000 \cdot PO^2 \cdot Cov(\beta_1, \beta_{12}) +$$
$$+(20x + 100)^2 \cdot 100 \cdot PO^2 \cdot Cov(\beta_{11}, \beta_{12})$$

and this time, the confidence intervals are:

$$IC_{95\%} = 10 \cdot \beta_1 + (20x + 100) \cdot \beta_{11} + 10 \cdot PO \cdot \beta_{12} \pm$$
$$\pm 1.96 \cdot (Var(10 \cdot \beta_1 + (20x + 100) \cdot \beta_{11} + 10 \cdot PO \cdot \beta_{12}))/n)^{\frac{1}{2}}$$

TABLE 3.8. Increase in the number of surveillance colonoscopies increasing participation

YEAR	Increase of surveillance colonoscopies for an increase from 40% to 50% in participation with a positivity of 4%	Increase of surveillance colonoscopies for an increase from 40% to 50% in participation with a positivity of 5%	Increase of surveillance colonoscopies for an increase from 40% to 50% in participation with a positivity of 6%
2016	41.0 [39.6,42.4]	41.0 [39.6,42.4]	41.0 [39.6,42.4]
2017	24.9 [24.5,25.3]	30.4 [30.0,30.8]	36.0 [35.6,36.3]
2018	66.3 [66.1,66.5]	81.9 [81.7,82.1]	97.5 [97.3,97.7]
2019	84.9 [93.4,95.3]	84.9 [93.4,95.3]	84.9 [93.4,95.3]
2020	90.1 [89.9,90.3]	109.8 [109.6,110]	129.5 [129.4,129.7]
2021	126.9 [126.7,127.1]	153.3 [153.1,153.5]	179.8 [179.6,180.0]
2022	135.6 [135.5,135.8]	159.3 [159.1,159.4]	182.9 [182.8,183.0]
2023	157.4 [156.7,158.1]	157.4 [156.7,158.1]	157.4 [156.7,158.1]
2024	137.8 [152.4,153.8]	137.8 [152.4,153.8]	137.8 [152.4,153.8]
2025	187.3 [186.7,188.0]	187.3 [186.7,188.0]	187.3 [186.7,188.0]
2026	140.6 [155.5,156.9]	140.6 [155.5,156.9]	140.6 [155.5,156.9]
2027	168.6 [168.4,168.7]	198.7 [198.6,198.9]	228.9 [228.8,229.1]
2028	173.0 [172.9,173.1]	204.4 [204.3,204.6]	235.9 [235.8,236.0]
2029	167.5 [167.4,167.6]	208.3 [208.2,208.5]	249.2 [249.1,249.3]
2030	207.5 [206.9,208.2]	207.5 [206.9,208.2]	207.5 [206.9,208.2]
2031	183.2 [183.0,183.3]	221.3 [221.2,221.5]	259.5 [259.4,259.6]
2032	216.5 [215.9,217.1]	216.5 [215.9,217.1]	216.5 [215.9,217.1]
2033	184.0 [183.8,184.1]	224.2 [224.1,224.3]	264.4 [264.3,264.5]
2034	172.9 [172.4,173.4]	172.9 [172.4,173.4]	172.9 [172.4,173.5]

In the tables 3.8 and 3.9 it is shown the increase in the number of surveillance colonoscopies when participation is increased in a 10%, through time and for a population of 100,000 inhabitants conditioned to 3 different values of positivity and for an adherence of 52.06% (The mean of the simulations).

TABLE 3.9. Increase in the number of surveillance colonoscopies increasing participation

YEAR	Increase of surveillance colonoscopies for an increase from 50% to 60% in participation with a positivity of 4%	Increase of surveillance colonoscopies for an increase from 50% to 60% in participation with a positivity of 5%	Increase of surveillance colonoscopies for an increase from 50% to 60% in participation with a positivity of 6%
2016	41 [39.6,42.4]	41 [39.6,42.4]	41.0 [39.6,42.4]
2017	24.9 [24.5,25.3]	30.4 [30.0,30.8]	36.0 [35.6,36.3]
2018	66.3 [66.1,66.5]	81.9 [81.7,82.1]	97.5 [97.3,97.7]
2019	105.8 [114.5,116.1]	115.3 [114.5,116.1]	115.3 [114.5,116.1]
2020	90.1 [89.9,90.3]	109.8 [109.6,110.0]	129.5 [129.4,129.7]
2021	126.9 [126.7,127.1]	153.3 [153.1,153.5]	179.8 [179.6,180.0]
2022	135.6 [135.5,135.8]	159.3 [159.1,159.4]	182.9 [182.8,183.0]
2023	157.4 [156.7,158.1]	157.4 [156.7,158.1]	157.4 [156.7,158.1]
2024	171.8 [186.5,187.7]	187.1 [186.5,187.7]	187.1 [186.5,187.7]
2025	187.3 [186.7,188.0]	187.3 [186.7,188.0]	187.3 [186.7,188.0]
2026	175.3 [190.4,191.5]	190.9 [190.4,191.5]	190.9 [190.4,191.5]
2027	168.6 [168.4,168.7]	198.7 [198.6,198.9]	228.9 [228.8,229.1]
2028	173.0 [172.9,173.1]	204.4 [204.3,204.6]	235.9 [235.8,236.0]
2029	167.5 [167.4,167.6]	208.3 [208.2,208.5]	249.2 [249.1,249.3]
2030	207.5 [206.9,208.2]	207.5 [206.9,208.2]	207.5 [206.9,208.2]
2031	183.2 [183.0,183.3]	221.3 [221.2,221.5]	259.5 [259.4,259.6]
2032	216.5 [215.9,217.1]	216.5 [215.9,217.1]	216.5 [215.9,217.1]
2033	184.0 [183.8,184.1]	224.2 [224.1,224.3]	264.4 [264.3,264.5]
2034	172.9 [172.4,173.4]	172.9 [172.4,173.4]	172.9 [172.4,173.5]

In this case we can see that like the previous one that, except in some cases, the increase of number of colonoscopies is higher through time, and this increase is always equal or higher when positivity rises.

3.2.4. Validation of the linear models.

As said before, to validate a linear model it has to pass 4 tests to check the 4 assumptions. Despite all have been validated, as there are 40 models, in this work we will show just the validation of three models of each type, chosen randomly.

```
> sample(1:20,3,replace=F)
[1] 16  4 12
```

So, the years randomly choose have been the 2018, 2026 and 2030, that corresponds to the models:

Modelc4:

$$NScreening = -61.35 + PA \cdot 1.49 + PO \cdot 13.73 + PA \cdot PO \cdot 4.09$$

Modelc12:

$$NScreening = 8.07 + PA \cdot 1.04 + PO \cdot 9.15 + PA \cdot PO \cdot 4.64$$

Modelc16:

$$NScreening = -22.05 + PA \cdot 1.52 + PO \cdot 16 + PA \cdot PO \cdot 4.82$$

Models4:

$$NSurvei = 195.9 + PA \cdot 0.39 - PO \cdot 28.33 - AD \cdot +0.01 + PA \cdot PO \cdot 1.56$$

Models12:

$$NSurvei = 99.22 + AD \cdot 3.27 + PA^2 \cdot 0.17 + PO^2 \cdot 10.95$$

Models16:

$$NSurvei = -859.87 + PA \cdot 20.75 + PO \cdot 139.27 + AD \cdot 4.63$$

1.- **Linearity**: We will use Ramsay's reset test to check the linearity of the model, using the R function "resettest " of the package lmtest.

```
> resettest(modelc4)

RESET test

data:  modelc4
RESET = 3.3824, df1 = 2, df2 = 1629, p-value = 0.0342

> resettest(modelc12)

RESET test

data:  modelc2
RESET = 21.054, df1 = 2, df2 = 1629, p-value = 9.391e-10

> resettest(modelc16)

RESET test

data:  modelc16
RESET = 20.857, df1 = 2, df2 = 1629, p-value = 1.137e-09

> resettest(models4)

RESET test

data:  models4
RESET = 8.5478, df1 = 2, df2 = 1628, p-value = 0.0002028

> resettest(models12)

RESET test

data:  models12
RESET = 20.103, df1 = 2, df2 = 1629, p-value = 2.373e-09

> resettest(models16)

RESET test

data:  models16
RESET = 10.304, df1 = 2, df2 = 1629, p-value = 3.575e-05
```

As p-value is lower than 0.05 there is no evidence to reject the hypothesis of linearity.

2.- **Normality**: As we can not use the Shapiro-Wilk test, because it is indicated to populations with $n < 50$, we will use the Lilliefors (Kolmogorov-Smirnov) test with the R function "lillie.test" of the package nortest.

```
> lillie.test(modelc4$residuals)

Lilliefors (Kolmogorov-Smirnov) normality test

data:  modelc4$residuals
D = 0.017033, p-value = 0.2988

> lillie.test(modelc12$residuals)

Lilliefors (Kolmogorov-Smirnov) normality test

data:  modelc12$residuals
D = 0.018326, p-value = 0.2024

> lillie.test(modelc16$residuals)

Lilliefors (Kolmogorov-Smirnov) normality test

data:  modelc16$residuals
D = 0.013059, p-value = 0.7158

> lillie.test(models4$residuals)

Lilliefors (Kolmogorov-Smirnov) normality test

data:  models4$residuals
D = 0.024263, p-value = 0.2511

> lillie.test(models12$residuals)

Lilliefors (Kolmogorov-Smirnov) normality test

data:  models12$residuals
D = 0.022282, p-value = 0.05616

> lillie.test(models16$residuals)

Lilliefors (Kolmogorov-Smirnov) normality test

data:  models16$residuals
D = 0.017478, p-value = 0.2626
```

We accept the null hypothesis of normality of residuals in all cases.

3.- **Homocedasticity.**: We will use the Harrison McCabe test, using the R function "hmctest" of the package lmtest.

```
> hmctest(modelc4)

Harrison-McCabe test

data:  modelc4
HMC = 0.51958, p-value = 0.867

> hmctest(modelc12)

Harrison-McCabe test

data:  modelc12
HMC = 0.53274, p-value = 0.969

> hmctest(modelc16)

Harrison-McCabe test

data:  modelc16
HMC = 0.50406, p-value = 0.628

> hmctest(models4)

Harrison-McCabe test
```

```
data:  models4
HMC = 0.50041, p-value = 0.513

> hmctest(models12)

Harrison-McCabe test

data:  models12
HMC = 0.48138, p-value = 0.144

> hmctest(models16)

Harrison-McCabe test

data:  models16
HMC = 0.48304, p-value = 0.172
```

We accept the null hypothesis of homocedasticity in all cases.

4.- **Independence**: We will use Ljung-Box tests , using the R function "box.test"

```
> Box.test(modelc4$residuals,lag=1,type="Ljung-Box")

Box-Ljung test

data:  modelc4$residuals
X-squared = 1.6905e-06, df = 1, p-value = 0.999

> Box.test(modelc12$residuals,lag=1,type="Ljung-Box")

Box-Ljung test

data:  modelc12$residuals
X-squared = 0.28916, df = 1, p-value = 0.5908

> Box.test(modelc16$residuals,lag=1,type="Ljung-Box")

Box-Ljung test

data:  modelc16$residuals
X-squared = 0.10712, df = 1, p-value = 0.7434

> Box.test(models4$residuals,lag=1,type="Ljung-Box")

Box-Ljung test

data:  models4$residuals
X-squared = 0.29844, df = 1, p-value = 0.5849


> Box.test(models12$residuals,lag=1,type="Ljung-Box")

Box-Ljung test

data:  models12$residuals
X-squared = 0.77395, df = 1, p-value = 0.379

> Box.test(models16$residuals,lag=1,type="Ljung-Box")

Box-Ljung test

data:  models16$residuals
X-squared = 2.0234, df = 1, p-value = 0.1549
```

As the p-value is higher than 0.05 under the level of significance of 95% we accept the null hypothesis of independence

In summary, with the linear models we have checked how participation, adherence and positivity affect to the number of colonoscopies. It has been proved that positivity is the most sensitive variable, and its effect is ten times higher than the effect of participation. An increase in adherence also provokes an increase in the number of colonoscopies, and this increase will ne higher as long as time passes.

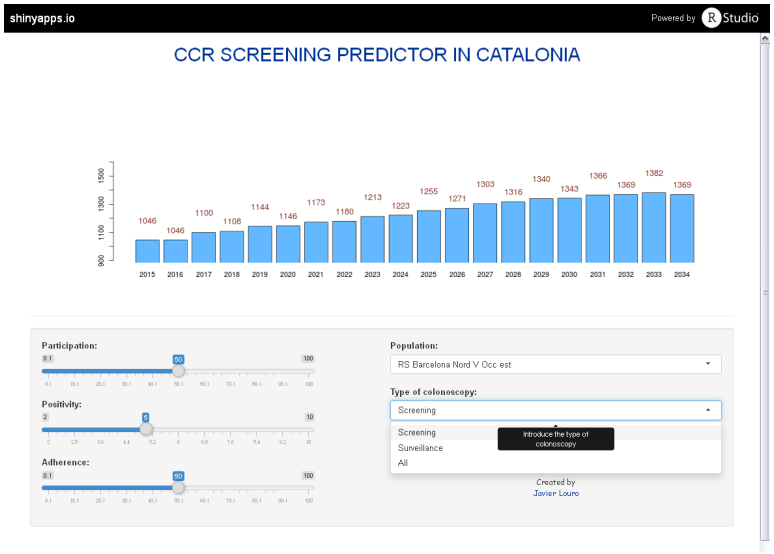
Using this models, we can predict the number of both colonoscopies after positive FIT and those of surveillance for each endoscopic unit, applying to this models the specific territorial levels of adherence, positivity and participation.

3.3. Application

This sensitivity analysis has two practical applications. It allows to asses how the variability associated with the estimation of the adherence, positivity and participation affects the results, but also to predict with a value of this three variables the number of both after a positive FIT and surveillance colonoscopies that is going to do an endoscopic unit a determined year, just using the respective model.

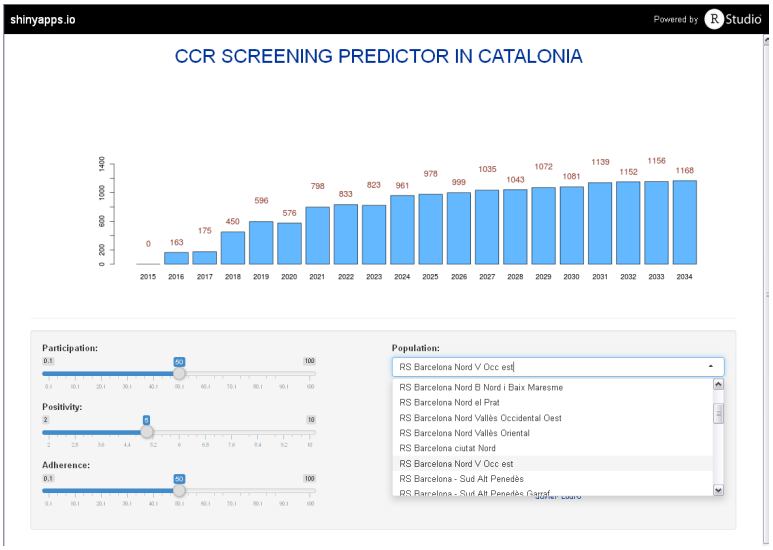
To show this results in a clean and easy-to-handle way, an application was created using the R software Shiny. This application shows the barplot of the number of colonoscopies per year of the selected type and endoscopic unit. Additional sidebar are included to change participation, positivity and adherence. We can see the interface in figures 3.3 and 3.4. The R code of the entire application its placed in the appendix with all the R code of this dissertation. The app is uploaded at "jlae.shinyapps.io/CRCpredictor " and it can be used using the username "CRC" and the password "1234 " .

FIG. 3.3. Interface of the shiny app, colonoscopies slider



In addition to this information we have 2 buttons, the first one "Download XLS" allows to download an excel file with the number of colonoscopies both after a positive FIT and surveillance, and the total, for the selected population, participation, adherence and positivity. We can see this button on figure 3.5. The second button allows you to download the plot in pdf. We can see it on figure 3.6.

FIG. 3.4. Interface of the shiny app, population slider



Briefly this is a friendly-user way to show the results of the sensitivity analysis and is useful for decision-makers to predict the future number of colonoscopies by selecting the variables and the population of interest.

FIG. 3.5. The "Download XLS button"

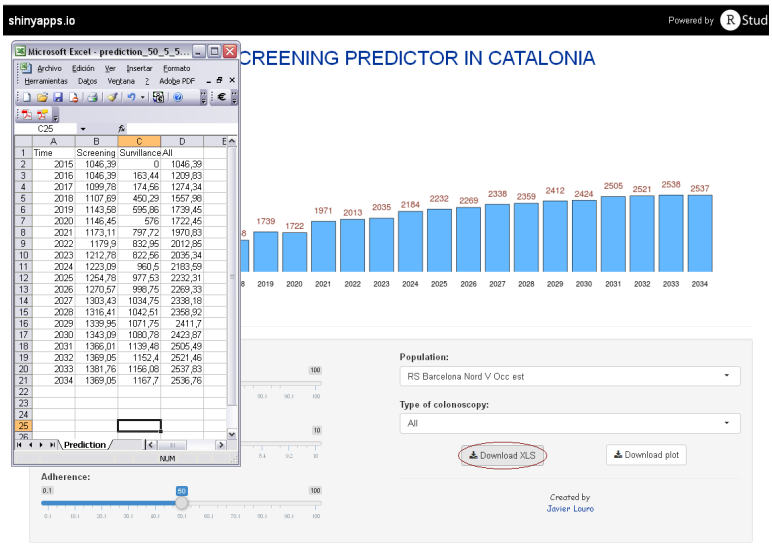
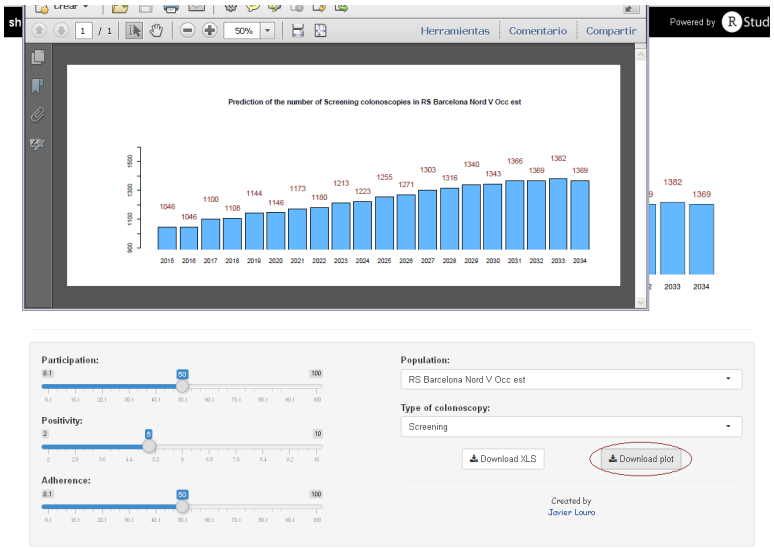


FIG. 3.6. The "Download plot button"



With this application, each manager can select an endoscopic unit, and, introducing the different values of participation, adherence and positivity that corresponds to this territory, predict the resources this endoscopic unit will need in the future twenty years to meet the predicted needs of colonoscopies.

Chapter 4
Future refinements of the model

Statisticians, like artists, have the bad habit
of falling in love with their models
- George E.P. Box

The aim of this chapter is to explain how this model will be improved in the future to provide more refined territorial predictions.

The specific territorial predictions shown in the previous chapter were calculated by re-scaling the results of the model (for a population of 100,000 inhabitants) to the size of each population.

The main restriction of this model is that it considers that the populations have the same age and sex structure and differ by population number only. Age and sex structure is similar among Healthcare Regions (RS), but at the level of endoscopic units relevant differences appear. As inputs depend both on age and sex, and outputs will be affected by difference in the population.

Two solutions were considered, to create a model for each endoscopic unit, which was discarded by the great time consumption of executing the model 38 times, or modifying inputs to represent equally all combinations of age and gender and collect outputs to re-scale them according to the specific structure of each population. Thus, the model will give the following information:

- 38 Factors representing the number of colonoscopies after positive FIT per year for each "n" habitants, by sex (male or female) and age (yearly).

- 38 Factors representing the number of surveillance colonoscopies per year for each "n " inhabitants, by sex (male or female) and age (yearly).

- 40 factors representing the number of colonoscopies after positive FIT per year for each "n" habitants, each one for a sex (male or female) and for those individuals who turn 50 years in a year between 2016 and 2035.

- 40 factors representing the number of surveillance colonoscopies for each "n" habitants per year, each one for a sex (male or female) and for those individuals who turn 50 years in a year between 2016 and 2035.

That is a total of 156 factors that make up two matrix (we call those matrix C and S) with 20 rows and 78 columns each as follows:

$$C = \begin{pmatrix} CC_{M,2016} & CC_{F,2016} & CCY_{M,2016} & CCY_{F,2016} \\ CC_{M,2017} & CC_{F,2017} & CCY_{M,2017} & CCY_{F,2017} \\ \vdots & \vdots & \vdots & \vdots \\ CC_{M,2035} & CC_{F,2035} & CCY_{M,2035} & CCY_{F,2035} \end{pmatrix}$$

and

$$S = \begin{pmatrix} CS_{M,2016} & CS_{F,2016} & CSY_{M,2016} & CSY_{F,2016} \\ CS_{M,2017} & CS_{F,2017} & CSY_{M,2017} & CSY_{F,2017} \\ \vdots & \vdots & \vdots & \vdots \\ CS_{M,2035} & CS_{F,2035} & CSY_{M,2035} & CSY_{F,2035} \end{pmatrix}$$

where:

- $CC_{M,y} = \{CC_{M,y,i}\}_{i\in\{51,\cdots,69\}}$ a vector where each $CC_{M,y,i}$ represents the number of colonoscopies after positive FIT made to males of age "i" in the year "y" per "n" habitants (with a previously fixed "n").
- $CC_{F,y} = \{CC_{F,y,i}\}_{i\in\{51,\cdots,69\}}$ a vector where each $CC_{F,y,i}$ represents the number of colonoscopies after positive FIT made to females of age "i" in the year "y" per "n" habitants (with a previously fixed "n").
- $CCY_{M,y} = \{CCY_{M,y,j}\}_{j\in\{2016,\cdots,2035\}}$ a vector where each $CCY_{M,y,j}$ represents the number of colonoscopies after positive FIT made in the year "y" to males that turned 50 years old the year "j" per "n" inhabitants (with a previously fixed "n").
- $CCY_{F,y} = \{CCY_{F,y,j}\}_{j\in\{2016,\cdots,2035\}}$ a vector where each $CCY_{F,y,j}$ represents the number of colonoscopies after positive FIT made in the year "y" to females that turned 50 years old the year "j" per "n" habitants (with a previously fixed "n").
- $CS_{M,y} = \{CS_{M,y,i}\}_{i\in\{51,\cdots,69\}}$ a vector where each $CS_{M,y,i}$ represents the number of surveillance colonoscopies made to males of age "i" in the year "y" per "n" habitants (with a previously fixed "n") .
- $CS_{F,y} = \{CS_{F,y,i}\}_{i\in\{51,\cdots,69\}}$ a vector where each $CS_{F,y,i}$ represents the number of surveillance colonoscopies made to females of age "i" in the year "y" per "n" habitants (with a previously fixed "n").
- $CSY_{M,y} = \{CSY_{M,y,j}\}_{j\in\{2016,\cdots,2035\}}$ a vector where each $CSY_{M,y,j}$ represents the number of surveillance colonoscopies made in the year "y" to males that turned 50 years old the year "j" per "n" inhabitants (with a previously fixed "n").
- $CSY_{F,y} = \{CSY_{F,y,j}\}_{j\in\{2016,\cdots,2035\}}$ a vector where each $CSY_{F,y,j}$ represents the number of surveillance colonoscopies in the year "y" made to females that

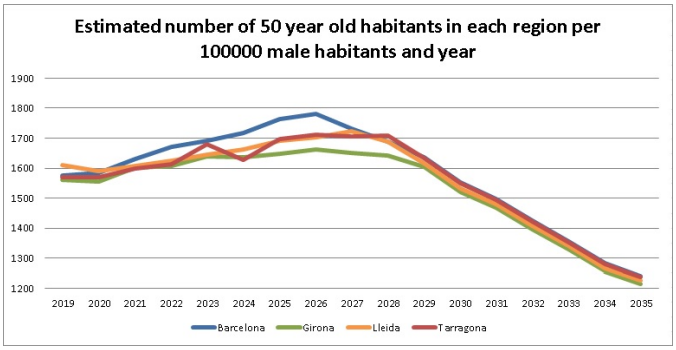
had turned 50 years old the year “j” per “n” habitants (with a previously fixed “n”).

This new model is rescheduled to begin in 2016 instead of 2015 as the one explained on chapter 2.

The Director Plan of Oncology of Catalonia provided the number of inhabitants aged 50 to 69 years assigned to every endoscopic unit, as well as the number of inhabitants who will turn 50 in the next two years, all calculated based on data from the Central Registry of Insured Persons (RCA). Hence the population of each endoscopic unit who turns 50 in 2016, 2017 and 2018 was obtained. For the remaining years until 2035 INEbase data was used.

Since this time the prediction will be made for each endoscopic unit, instead of using the prediction of Catalonia the prediction of each province was used. So for each endoscopic unit we assume that the number of inhabitants aged 50 years old from 2019-2029 will follow the same patter as the one of the province where the endoscopic unit is located. Between 2029 and 2035 we use data at the Spanish level as INEbase does not provide estimations beyond 2029 neither by province nor by Autonomous Region.

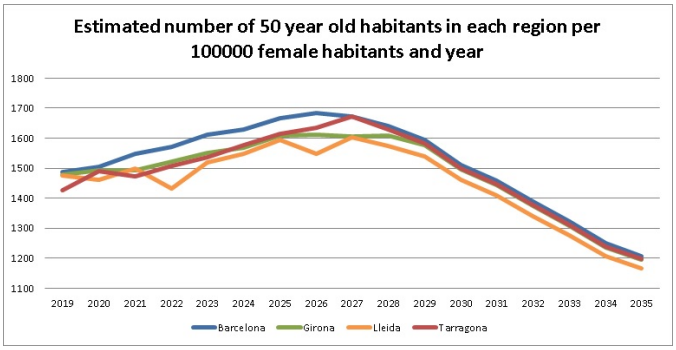
FIG. 4.1. Males who will turn 50 per year in a population of 100,000



In figure 4.1 it is shown the number of individuals that turn 50 for each province per 100,000 men and in figure 4.2 per 100,000 women.

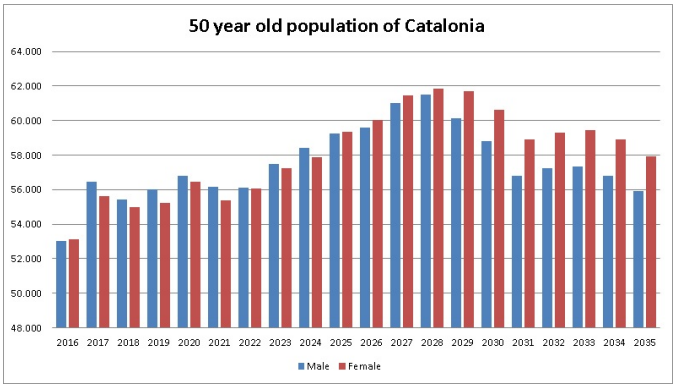
Both in males and females plots the estimated number of 50 year old inhabitants grow until 2027 in all regions and decrease since then. It has to be remembered that these estimates are new and this decrease was not observed in earlier predictions used in the previous model, as we can see comparing the new barplot of population (figure 4.3) with the one used before (figure 2.4). This makes even more necessary to readjust the model results.

FIG. 4.2. Females who will turn 50 per year in a population of 100,000



With these new model that we created restructuring the previous one we can predict the number of both after positive FIT and surveillance colonoscopies for each endoscopic unit taking in to account how the population is distributed in terms of age or sex.

FIG. 4.3. New prediction of 50 year old population of Catalonia



For each endoscopic unit we can build the column vector V of length 78:

$$V = (NM_{51}, \cdots, NM_{69}, NF_{51}, \cdots, NF_{69}, NMY_{2016}, \cdots, NMY_{2035}, NFY_{2016}, \cdots, NFY_{2035})^t$$

with:

- NM_i: Number of males of age "i".
- NF_i: Number of females of age "i".
- NMY_j: Number of males who turn 50 on year "j".

- NFY_j: Number of males who turn 50 on year "j".

So finally we will compute the vectors CC and CS where:

$$(CC)^t = C \cdot V$$

$$(CS)^t = S \cdot V$$

the fact that $C, S \in \mathcal{M}_{20,78}$ and $V \in \mathcal{M}_{78,1}$ implies that $(CC)^t \in \mathcal{M}_{20,1}$ and $(CS)^t \in \mathcal{M}_{20,1}$ and so CC and CS will be two row vectors of length 20; $CC \in \mathcal{M}_{1,20}$, $CC = (CC_1, \cdots, CC_{20})$ and $CS \in \mathcal{M}_{1,20}$, $CS = (CS_1, \cdots, CS_{20})$ where each CC_y will be the number of colonoscopies after positive FIT on year "y" and each CS_y will be the number of surveillance colonoscopies on year "y".

As an example, we will show preliminary results for one endoscopic unit.

In the table 4.1 we can see the future prediction of colonoscopies for the endoscopic unit of "Hospital del Mar" predicted with this new model. This area has a target population of 69,147 inhabitants.

As it is shown this endoscopic unit will have to do 14,371 colonoscopies over the 20 year-horizon. From the differences between the number of colonoscopies that are predicted for men and women, we can see the importance of including an accurate gender distribution.

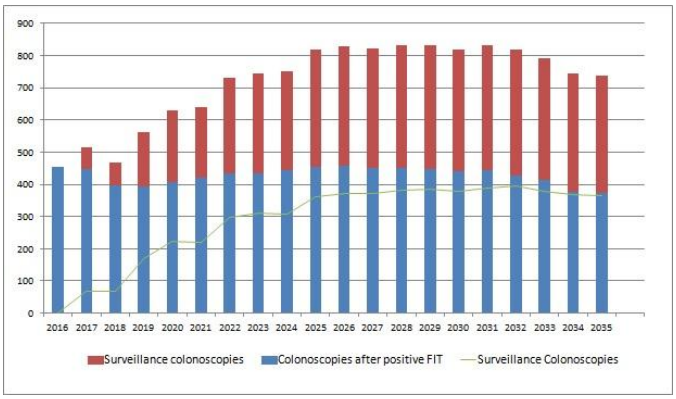
TABLE 4.1. Results on number of colonoscopies of the new model in "Hospital del Mar"

YEAR	After positive FIT of Men	After positive FIT of women	Surveillance of Men	Surveillance of women	Total after positive FIT	Total Surveillance	Total
2016	177.8	274.8	0.0	0.0	452.5	0.0	452.5
2017	176.0	270.7	26.7	40.2	446.6	67.0	513.6
2018	188.0	209.2	27.2	41.6	397.1	68.8	466.0
2019	183.0	209.6	73.1	97.0	392.6	170.1	562.7
2020	197.8	208.0	94.2	130.1	405.9	224.3	630.1
2021	203.6	215.9	102.1	118.6	419.5	220.8	640.3
2022	215.4	217.7	137.8	160.9	433.1	298.7	731.8
2023	217.3	215.0	145.0	167.0	432.3	312.0	744.3
2024	223.2	219.6	150.2	158.3	442.8	308.5	751.3
2025	233.4	221.5	176.2	186.1	454.9	362.4	817.3
2026	237.5	220.4	184.6	187.1	457.9	371.7	829.6
2027	230.3	221.2	186.3	184.0	451.5	370.3	821.8
2028	233.8	216.9	192.7	189.4	450.7	382.1	832.8
2029	230.8	216.2	200.3	184.1	447.0	384.3	831.4
2030	228.8	212.2	194.1	183.8	441.1	377.9	819.0
2031	231.8	213.3	200.7	187.6	445.0	388.4	833.4
2032	219.6	206.4	207.9	185.5	426.0	393.4	819.4
2033	215.7	198.5	196.1	182.5	414.1	378.5	792.7
2034	200.1	175.5	194.6	173.6	375.6	368.2	743.8
2035	196.2	178.0	196.7	166.3	374.2	363.1	737.3
TOTAL							14361

It is also important to see how the impact of the new predictions on population ageing, as a decrease in, both after positive FIT and surveillance from the year 2031 is observed, coinciding with the decrease in the population of 50-year olds predicted by INEbase data.

In the figure 4.4 it is shown a bar plot with this results.

FIG. 4.4. Average number of colonoscopies predicted for the "Hospital del Mar"



The next step will be to rebuild the sensitivity analysis to estimate new models to predict the number of colonoscopies to each endoscopic unit varying the positivity, participation and adherence, and then reprogram the application of the section 3.3 using this data.

Chapter 5
Discussion and conclusions

Contrariwise, if it was so, it might be;
and if it were so, it would be;
but as it isn't, it ain't.
That's logic.
- Lewis Carroll [1]

Despite colorectal cancer has a relatively low mortality rate if detected early, is still the second leading cause of cancer death in Spain and, more specifically, in Catalonia. The fact that CRC screening is currently being extended in Catalonia makes the results of the present dissertation relevant for planning purposes.

As previously was stated, this expansion of the screening program to the whole territory of Catalonia arises the need of, by statistical means, estimate how the demand of colonoscopies will be affected by the needs generated by the screening program. Our study presents a discrete event simulation model of a population-based colorectal screening program following the European Guidelines that provides a valuable source of information for health care services planning.

In short, we managed to achieve the main objective of this project, as we have defined a mathematical model capable of analysing through discrete event simulation the number of colonoscopies that each endoscopic unit must provide to the early detection screening program of colorectal cancer in Catalonia, due to the early detection Colorectal Cancer Screening Program of Catalonia.

The conclusions in terms of results we have drawn is that for the population of all Catalonia the total number of colonoscopies increase from 21,286 in the first year, passing on 39.060 passed 10 years, to 45,319 after 20 years, making a total of almost 730,000 colonoscopies during these 20 years according to the model predictions. Of these 730,000 colonoscopies the 60% will be colonoscopies after a positive FIT, and 40% will be those for the surveillance of the detected pre-malignant lesions.

This model also predicted a strong increase in participation from 2017 (up to 10%) and a clear decrease in positivity (in a bit more than 1.5%) also from 2017 due to the increase of successive screening.

[1]See Carrol L. (1872)

Besides, using linear models a sensitivity analysis has been performed, which allowed us to see the impact on the number of colonoscopies of the uncertainty associated with the estimation of the adherence, positivity and participation parameters. Moreover, it allowed predicting the number of colonoscopies for some value of these three parameters, and translating these findings into a simple web application.

Despite trying mixed-effects models, a full sensitivity analysis was done by creating general linear models by year and we could conclude that an increase of adherence, participation and positivity rates increase the number of colonoscopies, but with different magnitude.

Adherence affects to the number of surveillance colonoscopies only, and it is the variable with the lowest effect in magnitude. An increase of a 10% in this variable make an increase of between 0 colonoscopies in the short-term to over 50 in the long-term. In general this increment does not depend on participation and positivity rates.

The participation also increment more the number of colonoscopies in the long-term than in the short-term. This increase is also higher as higher is the positivity due to the interaction between this variables. In average, an increase of a 10% in the participation will increase the number of colonoscopies after positive FIT in around 200 in the first year, 280 after ten years (in 2024) and in 320 at the end of the time horizon of the simulation. On the other hand, in average, an increase of a 10% in the participation will increase the number of surveillance colonoscopies on less than 50 in the second and third year (the first year the number of surveillance colonoscopies is always null because surveillance has not started yet), in 137 in 2024 and in more than 170 in last year.

Finally, the FIT positivity is far the most sensitive variable. Changing a 1% of this variable is similar on changing a 10% on participation. This increment is affected also by time and participation. In average, increasing a 1% the positivity will lead to an increase of 170 colonoscopies after a positive FIT in first year, of 233 on 2024 and of 264 on 2034. In terms of surveillance colonoscopies this increase will be of less than 20 both in years 2016 and 2017, of 60 on 2024, and around 160 in 2034.

All this previous data is referred to the results of this model, so this increments will be always per 100,000 inhabitants of the population.

In order to deliver these results for every endoscopic unit in a friendly user way, an application was created using R software to present customized data for each unit. This application can be useful for planning the necessary resources in a 20-year horizon and will be presented to the Technical Screening Office of Catalonia.

One of the limitations of this study is that the simulation of the tests results are based on empirical distributions according to the results obtained from a program rather than applying sensitivity and specificity of tests according to the natural history stage of the disease. Modelling the natural history of the disease was out of the scope of this study. Another limitation is that surveillance colonoscopies of cancers detected under the screening program were not taken into account, as they depend on several individual factors subject to clinical decision.

Finally, the foresight of the population by province from 2029 on was not available and we had to adapt the Spanish tendency to all provinces and units, knowing that the differences in age structure are relevant.

One key strength of this work is that we used data from a CRC screening program covering around 200.000 inhabitants, compliant with the European Guidelines for Quality Assurance. Age and sex-specific parameters were estimated from the areas corresponding to the CRC screening program of Barcelona between 2009 and 2015. Moreover, the present dissertation develops a sensitivity analysis methodology using regression models to assess the effect of crucial parameters on the outputs of the model and through time and achieves, through the development of an application, a friendly-user and interactive way to present results to decision-makers. However, future work, as introduced in this dissertation, is needed to improve the accuracy of predictions by endoscopic unit taking into account future population predictions.

In conclusion, this simulation model and its analysis have shown to be powerful tools for health services planning and to inform decision-making. Beyond the modelling/technical matters, this piece of research should facilitate reflections on the capacity of the health system to meet the demand of colonoscopies induced by the CRC screening program, and how endoscopy workforce should be subject to a conscious planning.

Appendix A
R CODE: Read data

```
library(nlme)
library(foreign)
library(rsconnect)
library(shiny)
library(lmtest)
library(nortest)
library(shinyBS)
library(shinyjs)
library(graphics)
library(WriteXLS)
library(foreign)

### ~LECTURA DE DATOS~ ###

data<-read.spss("ResultadosPerASENS.sav", to.data.frame=T)
summary(data)
data$Participacio<-data$Participacio*100
data$Positivitat<-data$Positivitat*100
data$Tiempo<-data$TNOW+0.0001
```

Appendix B
R CODE: Sensitivity analysis, mixed effects models

```
### ~MIXED EFFECTS~ ###

lme.1<-lme(NColonoCribado ~ Tiempo+ Tiempo:Participacio+ Tiempo:Positivitat +
Positivitat + Participacio,
random=~1+Tiempo|NREP,data,method="ML",na.action=na.omit)
summary(lme.1)

lme.2<-lme(vNColonoSeguimiento~ Tiempo+ Tiempo:Participacio
+Tiempo:Positivitat +Adherencia:Tiempo +
Positivitat + Adherencia + Participacio,
random=~1+Tiempo|NREP,data,method="ML",na.action=na.omit)
summary(lme.2)

pdf("ranef1.pdf",width=8,height=5)
par(mfrow=c(1,2))
boxplot(ranef(lme.1)[,1],main="Random effects on the intercept"
,ylim=c(-250,250))
abline(h=1)
boxplot(ranef(lme.1)[,2],main="Random effects on the time", ylim=c(-8,8))
abline(h=1)
dev.off()

hist(lme.1$residuals[,1])
hist(lme.2$residuals[,1])
lillie.test(lme.1$residuals[,1])
lillie.test(lme.2$residuals[,1])

lillie.test(ranef(lme.1)[,2])
ad.test(ranef(lme.1)[,2])
cvm.test(ranef(lme.1)[,2])
pearson.test(ranef(lme.1)[,2])
lillie.test(ranef(lme.1)[,2])

windows()
par(mfrow=c(2,1))
qqnorm(lme.1,abline=c(0,1),main="Screening model")
qqnorm(lme.1,abline=c(0,1),main= "Surveillance model")
```

Appendix C
R CODE: Sensitivity analysis, linear models

```
### ~LINEAL MODELS~ ###

modelc1<-with(subset(data,Tiempo==1),lm(NColonoCribado~
 Participacio+Positivitat+Positivitat:Participacio))
modelc2<-with(subset(data,Tiempo==2),lm(NColonoCribado~
 Participacio+Positivitat+Positivitat:Participacio))
modelc3<-with(subset(data,Tiempo==3),lm(NColonoCribado~
 Participacio+Positivitat+Positivitat:Participacio))
modelc4<-with(subset(data,Tiempo==4),lm(NColonoCribado~
 Participacio+Positivitat+Positivitat:Participacio))
modelc5<-with(subset(data,Tiempo==5),lm(NColonoCribado~
 Participacio+Positivitat+Positivitat:Participacio))
modelc6<-with(subset(data,Tiempo==6),lm(NColonoCribado~
 Participacio+Positivitat+Positivitat:Participacio))
modelc7<-with(subset(data,Tiempo==7),lm(NColonoCribado~
 Participacio+Positivitat+Positivitat:Participacio))
modelc8<-with(subset(data,Tiempo==8),lm(NColonoCribado~
 Participacio+Positivitat+Positivitat:Participacio))
modelc9<-with(subset(data,Tiempo==9),lm(NColonoCribado~
 Participacio+Positivitat+Positivitat:Participacio))
modelc10<-with(subset(data,Tiempo==10),lm(NColonoCribado~
 Participacio+Positivitat+Positivitat:Participacio))
modelc11<-with(subset(data,Tiempo==11),lm(NColonoCribado~
 Participacio+Positivitat+Positivitat:Participacio))
modelc12<-with(subset(data,Tiempo==12),lm(NColonoCribado~
 Participacio+Positivitat+Positivitat:Participacio))
modelc13<-with(subset(data,Tiempo==13),lm(NColonoCribado~
 Participacio+Positivitat+Positivitat:Participacio))
modelc14<-with(subset(data,Tiempo==14),lm(NColonoCribado~
 Participacio+Positivitat+Positivitat:Participacio))
modelc15<-with(subset(data,Tiempo==15),lm(NColonoCribado~
 Participacio+Positivitat+Positivitat:Participacio))
modelc16<-with(subset(data,Tiempo==16),lm(NColonoCribado~
 Participacio+Positivitat+Positivitat:Participacio))
modelc17<-with(subset(data,Tiempo==17),lm(NColonoCribado~
 Participacio+Positivitat+Positivitat:Participacio))
modelc18<-with(subset(data,Tiempo==18),lm(NColonoCribado~
 Participacio+Positivitat+Positivitat:Participacio))
modelc19<-with(subset(data,Tiempo==19),lm(NColonoCribado~
 Participacio+Positivitat+Positivitat:Participacio))
modelc20<-with(subset(data,Tiempo==20),lm(NColonoCribado~
 Participacio+Positivitat+Positivitat:Participacio))

models1<-with(subset(data,Tiempo==Tiempo[1]),lm(vNColonoSeguimiento~
 Participacio+Positivitat+Adherencia+Participacio:Positivitat))
models2<-with(subset(data,Tiempo==Tiempo[2]),lm(vNColonoSeguimiento~
 Participacio+Positivitat+Adherencia))
```

```
models3<-with(subset(data,Tiempo==Tiempo[3]),lm(vNColonoSeguimiento~
 Participacio+Positivitat+Adherencia+Participacio:Positivitat))
models4<-with(subset(data,Tiempo==Tiempo[4]),lm(vNColonoSeguimiento~
 Participacio+Positivitat+Adherencia+Participacio:Positivitat))
models5<-with(subset(data,Tiempo==Tiempo[5]),lm(vNColonoSeguimiento~
I(Participacio^2)+Adherencia+I(Positivitat^2)))
models6<-with(subset(data,Tiempo==Tiempo[6]),lm(vNColonoSeguimiento~
 Participacio+Positivitat+Adherencia+Participacio:Positivitat))
models7<-with(subset(data,Tiempo==Tiempo[7]),lm(vNColonoSeguimiento~
 Participacio+Positivitat+Adherencia+Participacio:Positivitat))
models8<-with(subset(data,Tiempo==Tiempo[8]),lm(vNColonoSeguimiento~
 Participacio+Positivitat+Adherencia+Participacio:Positivitat))
models9<-with(subset(data,Tiempo==Tiempo[9]),lm(vNColonoSeguimiento~
 Participacio+Positivitat+Adherencia))
models10<-with(subset(data,Tiempo==Tiempo[10]),lm(vNColonoSeguimiento~
  I(Participacio^2)+Adherencia+I(Positivitat^2)))
models11<-with(subset(data,Tiempo==Tiempo[11]),lm(vNColonoSeguimiento~
 Participacio+Positivitat+Adherencia))
models12<-with(subset(data,Tiempo==Tiempo[12]),lm(vNColonoSeguimiento~
  I(Participacio^2)+Adherencia+I(Positivitat^2)))
models13<-with(subset(data,Tiempo==Tiempo[13]),lm(vNColonoSeguimiento~
 Participacio+Positivitat+Adherencia+Participacio:Positivitat))
models14<-with(subset(data,Tiempo==Tiempo[14]),lm(vNColonoSeguimiento~
 Participacio+Positivitat+Adherencia+Participacio:Positivitat))
models15<-with(subset(data,Tiempo==Tiempo[15]),lm(vNColonoSeguimiento~
 Participacio+Positivitat+Adherencia+Participacio:Positivitat))
models16<-with(subset(data,Tiempo==Tiempo[16]),lm(vNColonoSeguimiento~
 Participacio+Positivitat+Adherencia))
models17<-with(subset(data,Tiempo==Tiempo[17]),lm(vNColonoSeguimiento~
 Participacio+Positivitat+Adherencia+Participacio:Positivitat))
models18<-with(subset(data,Tiempo==Tiempo[18]),lm(vNColonoSeguimiento~
 Participacio+Positivitat+Adherencia))
models19<-with(subset(data,Tiempo==Tiempo[19]),lm(vNColonoSeguimiento~
 Participacio+Positivitat+Adherencia+Positivitat:Participacio))
models20<-with(subset(data,Tiempo==Tiempo[20]),lm(vNColonoSeguimiento~
 Participacio+Positivitat+Adherencia))

summary(modelc1); summary(modelc2); summary(modelc3)
summary(modelc4); summary(modelc5); summary(modelc6)
summary(modelc7); summary(modelc8); summary(modelc9)
summary(modelc10); summary(modelc11); summary(modelc12)
summary(modelc13); summary(modelc14); summary(modelc15)
summary(modelc16); summary(modelc17); summary(modelc18)
summary(modelc19); summary(modelc20)
vcov(modelc1); vcov(modelc2); vcov(modelc3)
vcov(modelc4); vcov(modelc5); vcov(modelc6)
vcov(modelc7); vcov(modelc8); vcov(modelc9)
vcov(modelc10); vcov(modelc11); vcov(modelc12)
vcov(modelc13); vcov(modelc14); vcov(modelc15)
vcov(modelc16); vcov(modelc17); vcov(modelc18)
vcov(modelc19); vcov(modelc20)

summary(models1); summary(models2); summary(models3)
summary(models4); summary(models5); summary(models6)
summary(models7); summary(models8); summary(models9)
summary(models10); summary(models11); summary(models12)
summary(models13); summary(models14); summary(models15)
summary(models16); summary(models17); summary(models18)
summary(models19); summary(models20)
vcov(models1); vcov(models2); vcov(models3)
vcov(models4); vcov(models5); vcov(models6)
vcov(models7); vcov(models8); vcov(models9)
vcov(models10); vcov(models11); vcov(models12)
vcov(models13); vcov(models14); vcov(models15)
vcov(models16); vcov(models17); vcov(models18)
vcov(models19); vcov(models20)

### ~VALIDATION~ ###


lillie.test(modelc1$residuals); lillie.test(modelc2$residuals)
lillie.test(modelc3$residuals); lillie.test(modelc4$residuals)
```

```
lillie.test(modelc5$residuals); lillie.test(modelc6$residuals)
lillie.test(modelc7$residuals); lillie.test(modelc8$residuals)
lillie.test(modelc9$residuals); lillie.test(modelc10$residuals)
lillie.test(modelc11$residuals); lillie.test(modelc12$residuals)
lillie.test(modelc13$residuals); lillie.test(modelc14$residuals)
lillie.test(modelc15$residuals); lillie.test(modelc16$residuals)
lillie.test(modelc17$residuals); lillie.test(modelc18$residuals)
lillie.test(modelc19$residuals); lillie.test(modelc20$residuals)

resettest(modelc1); resettest(modelc2)
resettest(modelc3); resettest(modelc4)
resettest(modelc5); resettest(modelc6)
resettest(modelc7); resettest(modelc8)
resettest(modelc9); resettest(modelc10)
resettest(modelc11); resettest(modelc12)
resettest(modelc13); resettest(modelc14)
resettest(modelc15); resettest(modelc16)
resettest(modelc17); resettest(modelc18)
resettest(modelc19); resettest(modelc20)

hmctest(modelc1); hmctest(modelc2)
hmctest(modelc3); hmctest(modelc4)
hmctest(modelc5); hmctest(modelc6)
hmctest(modelc7); hmctest(modelc8)
hmctest(modelc9); hmctest(modelc10)
hmctest(modelc11); hmctest(modelc12)
hmctest(modelc13); hmctest(modelc14)
hmctest(modelc15); hmctest(modelc16)
hmctest(modelc17); hmctest(modelc18)
hmctest(modelc19); hmctest(modelc20)

Box.test(modelc1$residuals,lag=1,type="Ljung-Box")
Box.test(modelc2$residuals,lag=1,type="Ljung-Box")
Box.test(modelc3$residuals,lag=1,type="Ljung-Box")
Box.test(modelc4$residuals,lag=1,type="Ljung-Box")
Box.test(modelc5$residuals,lag=1,type="Ljung-Box")
Box.test(modelc6$residuals,lag=1,type="Ljung-Box")
Box.test(modelc7$residuals,lag=1,type="Ljung-Box")
Box.test(modelc8$residuals,lag=1,type="Ljung-Box")
Box.test(modelc9$residuals,lag=1,type="Ljung-Box")
Box.test(modelc10$residuals,lag=1,type="Ljung-Box")
Box.test(modelc11$residuals,lag=1,type="Ljung-Box")
Box.test(modelc12$residuals,lag=1,type="Ljung-Box")
Box.test(modelc13$residuals,lag=1,type="Ljung-Box")
Box.test(modelc14$residuals,lag=1,type="Ljung-Box")
Box.test(modelc15$residuals,lag=1,type="Ljung-Box")
Box.test(modelc16$residuals,lag=1,type="Ljung-Box")
Box.test(modelc17$residuals,lag=1,type="Ljung-Box")
Box.test(modelc18$residuals,lag=1,type="Ljung-Box")
Box.test(modelc19$residuals,lag=1,type="Ljung-Box")
Box.test(modelc20$residuals,lag=1,type="Ljung-Box")

lillie.test(models1$residuals); lillie.test(models2$residuals)
lillie.test(models3$residuals); lillie.test(models4$residuals)
lillie.test(models5$residuals); lillie.test(models6$residuals)
lillie.test(models7$residuals); lillie.test(models8$residuals)
lillie.test(models9$residuals); lillie.test(models10$residuals)
lillie.test(models11$residuals); lillie.test(models12$residuals)
lillie.test(models13$residuals); lillie.test(models14$residuals)
lillie.test(models15$residuals); lillie.test(models16$residuals)
lillie.test(models17$residuals); lillie.test(models18$residuals)
lillie.test(models19$residuals); lillie.test(models20$residuals)

resettest(models1); resettest(models2)
resettest(models3); resettest(models4)
resettest(models5); resettest(models6)
resettest(models7); resettest(models8)
resettest(models9); resettest(models10)
resettest(models11); resettest(models12)
resettest(models13); resettest(models14)
resettest(models15); resettest(models16)
resettest(models17); resettest(models18)
```

```
resettest(models19); resettest(models20)

hmctest(models1); hmctest(models2)
hmctest(models3); hmctest(models4)
hmctest(models5); hmctest(models6)
hmctest(models7); hmctest(models8)
hmctest(models9); hmctest(models10)
hmctest(models11); hmctest(models12)
hmctest(models13); hmctest(models14)
hmctest(models15); hmctest(models16)
hmctest(models17); hmctest(models18)
hmctest(models19); hmctest(models20)

Box.test(models1$residuals,lag=1,type="Ljung-Box")
Box.test(models2$residuals,lag=1,type="Ljung-Box")
Box.test(models3$residuals,lag=1,type="Ljung-Box")
Box.test(models4$residuals,lag=1,type="Ljung-Box")
Box.test(models5$residuals,lag=1,type="Ljung-Box")
Box.test(models6$residuals,lag=1,type="Ljung-Box")
Box.test(models7$residuals,lag=1,type="Ljung-Box")
Box.test(models8$residuals,lag=1,type="Ljung-Box")
Box.test(models9$residuals,lag=1,type="Ljung-Box")
Box.test(models10$residuals,lag=1,type="Ljung-Box")
Box.test(models11$residuals,lag=1,type="Ljung-Box")
Box.test(models12$residuals,lag=1,type="Ljung-Box")
Box.test(models13$residuals,lag=1,type="Ljung-Box")
Box.test(models14$residuals,lag=1,type="Ljung-Box")
Box.test(models15$residuals,lag=1,type="Ljung-Box")
Box.test(models16$residuals,lag=1,type="Ljung-Box")
Box.test(models17$residuals,lag=1,type="Ljung-Box")
Box.test(models18$residuals,lag=1,type="Ljung-Box")
Box.test(models19$residuals,lag=1,type="Ljung-Box")
Box.test(models20$residuals,lag=1,type="Ljung-Box")

sample(1:20,3,replace=F)
```

Appendix D
R CODE: Shiny app

D.1. Server: Beginning and charging interface

```
### ~APP SERVER~ ###

shinyServer(function(input, output, session) {

Logged = FALSE;
my_username <- "CRC"
my_password <- "1234"

output$plot <- renderPlot({
withProgress(message = 'Loading', value = 0, {
n <- 100
for (i in 1:n) {
incProgress(1/n, detail = paste(i))
Sys.sleep(0.03)
}})
})

ui1 <- function(){
tagList(
div(id = "login",
wellPanel(textInput("userName", "Username"),
passwordInput("passwd", "Password"),
br(),actionButton("Login", "Log in"))),
tags$style(type="text/css", "#login {font-size:10px;   text-align: left;position:absolute
;top: 40%;left:
50%;margin-top: -100px;margin-left: -150px;}")
)}
```

D.2. Server: Barplot

```
ui2 <- function(){tagList(tabPanel("Test"))}

output$regPlot <- renderPlot({

if (input$pob==" Barcelona - Sud l'Hospitalet LL") {factor<-60996}
if (input$pob==" Barcelona - Sud Alt Penedès") {factor<-21974}
if (input$pob==" Barcelona - Ciutat Esquerra i Litoral") {factor<-198911}
if (input$pob==" Lleida tot excepte LLeida") {factor<-84210}
if (input$pob==" Terres Ebre  Baix Ebre") {factor<-19103}
if (input$pob==" Girona Baix Empordà, Garrotxa, Ripollès") {factor<-50328}
if (input$pob==" Catalunya Central Osona") {factor<-37254}
if (input$pob==" Camp de Tarragona Alt Camp-Conca") {factor<-14554 }
if (input$pob==" Alt Pirineu i Aran Cerdanya, Pallars, Urgell") {factor<-13091}
```

```
if (input$pob==" Barcelona Nord V Occ est") {factor<- 100375}
if (input$pob==" Barcelona Nord el Prat") {factor<-15996}
if (input$pob==" Barcelona - Sud Maresem Central (o BN Baix maresme, doble)") {factor<-62915}
if (input$pob==" Catalunya Central Solsonès, Bages, Berguedà") {factor<-60238}
if (input$pob==" Alt Pirineu i Aran La Cerdanya") {factor<-3616}
if (input$pob==" Terres Ebre  Montsià + Altebrat (resta TE)") {factor<-24908}
if (input$pob==" Barcelona - Barcelona Dreta") {factor<-108564}
if (input$pob==" Girona Alt Empordà") {factor<-31643}
if (input$pob==" Girona Gironès") {factor<-50112}
if (input$pob==" Camp Tarragona Tarragonès-Baix Penedès") {factor<-75480}
if (input$pob==" Barcelona - Sud Alt Penedès Garraf") {factor<-36077}
if (input$pob==" Barcelona - Sud Baix Llobregat Centre-Litoral i LH.LL.") {factor<-143167}
if (input$pob==" Barcelona - Sud Baix Llobregat Nord") {factor<-32985}
if (input$pob==" Barcelona Nord Vallès Occidental Oest") {factor<-96611}
if (input$pob==" Barcelona Nord Vallès Oriental") {factor<-101513}
if (input$pob==" Barcelona ciutat Nord") {factor<-92487}
if (input$pob==" Barcelona Nord B Nord i Baix Maresme") {factor<-104105}
if (input$pob==" Camp de Tarragona Baix Camp-Priorat") {factor<-46625}
if (input$pob==" Catalunya Central Anoia") {factor<-25922}
if (input$pob==" Girona Alt Maresme") {factor<-26738}
if (input$pob==" Girona Selva") {factor<-35984}

newVal1<- data.frame(Participacio=as.numeric(input$par),Positivitat=as.numeric(input$pos))
newVal2<- data.frame(Participacio=as.numeric(input$par),Adherencia=as.numeric(input$adh)
,Positivitat=as.numeric(input$pos))

pred1<-vector()

pred1[1]<-predict(modelc1, newVal1, level=0)
pred1[2]<-predict(modelc2, newVal1, level=0)
pred1[3]<-predict(modelc3, newVal1, level=0)
pred1[4]<-predict(modelc4, newVal1, level=0)
pred1[5]<-predict(modelc5, newVal1, level=0)
pred1[6]<-predict(modelc6, newVal1, level=0)
pred1[7]<-predict(modelc7, newVal1, level=0)
pred1[8]<-predict(modelc8, newVal1, level=0)
pred1[9]<-predict(modelc9, newVal1, level=0)
pred1[10]<-predict(modelc10, newVal1, level=0)
pred1[11]<-predict(modelc11, newVal1, level=0)
pred1[12]<-predict(modelc12, newVal1, level=0)
pred1[13]<-predict(modelc13, newVal1, level=0)
pred1[14]<-predict(modelc14, newVal1, level=0)
pred1[15]<-predict(modelc15, newVal1, level=0)
pred1[16]<-predict(modelc16, newVal1, level=0)
pred1[17]<-predict(modelc17, newVal1, level=0)
pred1[18]<-predict(modelc18, newVal1, level=0)
pred1[19]<-predict(modelc19, newVal1, level=0)
pred1[20]<-predict(modelc20, newVal1, level=0)
pred2<-vector()

pred2[1]<-predict(models1, newVal2, level=0)
pred2[2]<-predict(models2, newVal2, level=0)
pred2[3]<-predict(models3, newVal2, level=0)
pred2[4]<-predict(models4, newVal2, level=0)
pred2[5]<-predict(models5, newVal2, level=0)
pred2[6]<-predict(models6, newVal2, level=0)
pred2[7]<-predict(models7, newVal2, level=0)
pred2[8]<-predict(models8, newVal2, level=0)
pred2[9]<-predict(models9, newVal2, level=0)
pred2[10]<-predict(models10, newVal2, level=0)
pred2[11]<-predict(models11, newVal2, level=0)
pred2[12]<-predict(models12, newVal2, level=0)
pred2[13]<-predict(models13, newVal2, level=0)
pred2[14]<-predict(models14, newVal2, level=0)
pred2[15]<-predict(models15, newVal2, level=0)
pred2[16]<-predict(models16, newVal2, level=0)
pred2[17]<-predict(models17, newVal2, level=0)
pred2[18]<-predict(models18, newVal2, level=0)
pred2[19]<-predict(models19, newVal2, level=0)
pred2[20]<-predict(models20, newVal2, level=0)

pred1<-(pred1*factor)/100000
```

```
pred2<-(pred2*factor)/100000

for (i in 1:20){
if(pred1[i]<0){pred1[i]<-0}
}
for (i in 1:20){
if(pred2[i]<0){pred2[i]<-0}
}
pred=pred1+pred2
ti<-2015:2034
bp1<-rbind(pred1,ti)
bp2<-rbind(pred2,ti)
bp<-rbind(pred,ti)
colnames(bp)<-2015:2034
colnames(bp1)<-2015:2034
colnames(bp2)<-2015:2034

if (input$col=="Screening") {

par(mar=c(5,10,10,2))
x<-barplot(bp1[1,],col="steelblue1",ylim=c(min(pred1)*0.85,1.2*max(pred1)),xpd=F)
graphics::text(x,bp1[1,]+rep(c(max(pred1)/10,max(pred1)/20),10),
labels=round(bp1[1,],0),cex=1.2,col="tomato4")
}

if (input$col=="Surveillance") {
par(mar=c(5,10,10,2))
x<-barplot(bp2[1,],col="steelblue1",ylim=c(min(pred2)*0.85,1.3*max(pred2)),xpd=F)
graphics::text(x,bp2[1,]+rep(c(max(pred2)/4,max(pred2)/8),10),
labels=round(bp2[1,],0),cex=1.2,col="tomato4")
}

if (input$col=="All") {
par(mar=c(5,10,10,2))
x<-barplot(bp[1,],col="steelblue1",ylim=c(min(pred)*0.85,1.3*max(pred)),xpd=F)
graphics::text(x,bp[1,]+rep(c(max(pred2)/5,max(pred2)/8),10),
labels=round(bp[1,],0),cex=1.2,col="tomato4")
}
})
```

D.3. Server: Download XLS button

```
output$downloadData <- downloadHandler(

filename = function() {
paste("prediction_", input$par, "_",input$pos, "_", input$adh, '.xls', sep='')
},
content = function(file) {

if (input$pob==" Barcelona - Sud l'Hospitalet LL") {factor<-60996}
if (input$pob==" Barcelona - Sud Alt Penedès") {factor<-21974}
if (input$pob==" Barcelona - Ciutat Esquerra i Litoral") {factor<-198911}
if (input$pob==" Lleida tot excepte LLeida") {factor<-84210}
if (input$pob==" Terres Ebre  Baix Ebre") {factor<-19103}
if (input$pob==" Girona Baix Empordà, Garrotxa, Ripollès") {factor<-50328}
if (input$pob==" Catalunya Central Osona") {factor<-37254}
if (input$pob==" Camp de Tarragona Alt Camp-Conca") {factor<-14554 }
if (input$pob==" Alt Pirineu i Aran Cerdanya, Pallars, Urgell") {factor<-13091}
if (input$pob==" Barcelona Nord V Occ est") {factor<- 100375}
if (input$pob==" Barcelona Nord el Prat") {factor<-15996}
if (input$pob==" Barcelona - Sud Maresem Central (o BN Baix maresme, doble)") {factor<-62915}
if (input$pob==" Catalunya Central Solsonès, Bages, Berguedà") {factor<-60238}
if (input$pob==" Alt Pirineu i Aran La Cerdanya") {factor<-3616}
if (input$pob==" Terres Ebre  Montsià + Altebrat (resta TE)") {factor<-24908}
if (input$pob==" Barcelona - Barcelona Dreta") {factor<-108564}
if (input$pob==" Girona Alt Empordà") {factor<-31643}
if (input$pob==" Girona Gironès") {factor<-50112}
```

```
if (input$pob==" Camp Tarragona Tarragonès-Baix Penedès") {factor<-75480}
if (input$pob==" Barcelona - Sud Alt Penedès Garraf") {factor<-36077}
if (input$pob==" Barcelona - Sud Baix Llobregat Centre-Litoral i LH.LL.") {factor<-143167}
if (input$pob==" Barcelona - Sud Baix Llobregat Nord") {factor<-32985}
if (input$pob==" Barcelona Nord Vallès Occidental Oest") {factor<-96611}
if (input$pob==" Barcelona Nord Vallès Oriental") {factor<-101513}
if (input$pob==" Barcelona ciutat Nord") {factor<-92487}
if (input$pob==" Barcelona Nord B Nord i Baix Maresme") {factor<-104105}
if (input$pob==" Camp de Tarragona Baix Camp-Priorat") {factor<-46625}
if (input$pob==" Catalunya Central Anoia") {factor<-25922}
if (input$pob==" Girona Alt Maresme") {factor<-26738}
if (input$pob==" Girona Selva") {factor<-35984}

newVal1<- data.frame(Participacio=as.numeric(input$par),Positivitat=as.numeric(input$pos))
newVal2<- data.frame(Participacio=as.numeric(input$par),Adherencia=as.numeric(input$adh),
Positivitat=as.numeric(input$pos))

pred1<-vector()

pred1[1]<-predict(modelc1, newVal1, level=0)
pred1[2]<-predict(modelc2, newVal1, level=0)
pred1[3]<-predict(modelc3, newVal1, level=0)
pred1[4]<-predict(modelc4, newVal1, level=0)
pred1[5]<-predict(modelc5, newVal1, level=0)
pred1[6]<-predict(modelc6, newVal1, level=0)
pred1[7]<-predict(modelc7, newVal1, level=0)
pred1[8]<-predict(modelc8, newVal1, level=0)
pred1[9]<-predict(modelc9, newVal1, level=0)
pred1[10]<-predict(modelc10, newVal1, level=0)
pred1[11]<-predict(modelc11, newVal1, level=0)
pred1[12]<-predict(modelc12, newVal1, level=0)
pred1[13]<-predict(modelc13, newVal1, level=0)
pred1[14]<-predict(modelc14, newVal1, level=0)
pred1[15]<-predict(modelc15, newVal1, level=0)
pred1[16]<-predict(modelc16, newVal1, level=0)
pred1[17]<-predict(modelc17, newVal1, level=0)
pred1[18]<-predict(modelc18, newVal1, level=0)
pred1[19]<-predict(modelc19, newVal1, level=0)
pred1[20]<-predict(modelc20, newVal1, level=0)

pred2<-vector()
pred2[1]<-predict(models1, newVal2, level=0)
pred2[2]<-predict(models2, newVal2, level=0)
pred2[3]<-predict(models3, newVal2, level=0)
pred2[4]<-predict(models4, newVal2, level=0)
pred2[5]<-predict(models5, newVal2, level=0)
pred2[6]<-predict(models6, newVal2, level=0)
pred2[7]<-predict(models7, newVal2, level=0)
pred2[8]<-predict(models8, newVal2, level=0)
pred2[9]<-predict(models9, newVal2, level=0)
pred2[10]<-predict(models10, newVal2, level=0)
pred2[11]<-predict(models11, newVal2, level=0)
pred2[12]<-predict(models12, newVal2, level=0)
pred2[13]<-predict(models13, newVal2, level=0)
pred2[14]<-predict(models14, newVal2, level=0)
pred2[15]<-predict(models15, newVal2, level=0)
pred2[16]<-predict(models16, newVal2, level=0)
pred2[17]<-predict(models17, newVal2, level=0)
pred2[18]<-predict(models18, newVal2, level=0)
pred2[19]<-predict(models19, newVal2, level=0)
pred2[20]<-predict(models20, newVal2, level=0)

pred1<-(pred1*factor)/100000
pred2<-(pred2*factor)/100000

for (i in 1:20){
if(pred1[i]<0){pred1[i]<-0}
}
for (i in 1:20){
if(pred2[i]<0){pred2[i]<-0}
}
```

```
pred=pred1+pred2
ti<-2015:2034
bp1<-rbind(pred1,ti)
bp2<-rbind(pred2,ti)
bp<-rbind(pred,ti)
colnames(bp)<-2015:2034
colnames(bp1)<-2015:2034
colnames(bp2)<-2015:2034

d<-data.frame(Time=ti,Screening=round(pred1,2), Survillance=round(pred2,2),all=round(pred,2))
WriteXLS("d",ExcelFileName = file, SheetNames = "Prediction")
})
```

D.4. Server: Download plot button

```
output$downloadpdf <- downloadHandler(

filename = function() {
paste("prediction_", input$par, "_",input$pos, "_", input$adh, '.pdf', sep='')
},
content = function(file) {
if (input$pob==" Barcelona - Sud l'Hospitalet LL") {factor<-60996}
if (input$pob==" Barcelona - Sud Alt Penedès") {factor<-21974}
if (input$pob==" Barcelona - Ciutat Esquerra i Litoral") {factor<-198911}
if (input$pob==" Lleida tot excepte LLeida") {factor<-84210}
if (input$pob==" Terres Ebre  Baix Ebre") {factor<-19103}
if (input$pob==" Girona Baix Empordà, Garrotxa, Ripollès") {factor<-50328}
if (input$pob==" Catalunya Central Osona") {factor<-37254}
if (input$pob==" Camp de Tarragona Alt Camp-Conca") {factor<-14554 }
if (input$pob==" Alt Pirineu i Aran Cerdanya, Pallars, Urgell") {factor<-13091}
if (input$pob==" Barcelona Nord V Occ est") {factor<- 100375}
if (input$pob==" Barcelona Nord el Prat") {factor<-15996}
if (input$pob==" Barcelona - Sud Maresem Central (o BN Baix maresme, doble)") {factor<-62915}
if (input$pob==" Catalunya Central Solsonès, Bages, Berguedà") {factor<-60238}
if (input$pob==" Alt Pirineu i Aran La Cerdanya") {factor<-3616}
if (input$pob==" Terres Ebre  Montsià + Altebrat (resta TE)") {factor<-24908}
if (input$pob==" Barcelona - Barcelona Dreta") {factor<-108564}
if (input$pob==" Girona Alt Empordà") {factor<-31643}
if (input$pob==" Girona Gironès") {factor<-50112}
if (input$pob==" Camp Tarragona Tarragonès-Baix Penedès") {factor<-75480}
if (input$pob==" Barcelona - Sud Alt Penedès Garraf") {factor<-36077}
if (input$pob==" Barcelona - Sud Baix Llobregat Centre-Litoral i LH.LL.") {factor<-143167}
if (input$pob==" Barcelona - Sud Baix Llobregat Nord") {factor<-32985}
if (input$pob==" Barcelona Nord Vallès Occidental Oest") {factor<-96611}
if (input$pob==" Barcelona Nord Vallès Oriental") {factor<-101513}
if (input$pob==" Barcelona ciutat Nord") {factor<-92487}
if (input$pob==" Barcelona Nord B Nord i Baix Maresme") {factor<-104105}
if (input$pob==" Camp de Tarragona Baix Camp-Priorat") {factor<-46625}
if (input$pob==" Catalunya Central Anoia") {factor<-25922}
if (input$pob==" Girona Alt Maresme") {factor<-26738}
if (input$pob==" Girona Selva") {factor<-35984}

newVal1<- data.frame(Participacio=as.numeric(input$par),Positivitat=as.numeric(input$pos))
newVal2<- data.frame(Participacio=as.numeric(input$par),Adherencia=as.numeric(input$adh)
,Positivitat=as.numeric(input$pos))

pred1<-vector()

pred1[1]<-predict(modelc1, newVal1, level=0)
pred1[2]<-predict(modelc2, newVal1, level=0)
pred1[3]<-predict(modelc3, newVal1, level=0)
pred1[4]<-predict(modelc4, newVal1, level=0)
pred1[5]<-predict(modelc5, newVal1, level=0)
pred1[6]<-predict(modelc6, newVal1, level=0)
pred1[7]<-predict(modelc7, newVal1, level=0)
pred1[8]<-predict(modelc8, newVal1, level=0)
pred1[9]<-predict(modelc9, newVal1, level=0)
```

```
pred1[10]<-predict(modelc10, newVal1, level=0)
pred1[11]<-predict(modelc11, newVal1, level=0)
pred1[12]<-predict(modelc12, newVal1, level=0)
pred1[13]<-predict(modelc13, newVal1, level=0)
pred1[14]<-predict(modelc14, newVal1, level=0)
pred1[15]<-predict(modelc15, newVal1, level=0)
pred1[16]<-predict(modelc16, newVal1, level=0)
pred1[17]<-predict(modelc17, newVal1, level=0)
pred1[18]<-predict(modelc18, newVal1, level=0)
pred1[19]<-predict(modelc19, newVal1, level=0)
pred1[20]<-predict(modelc20, newVal1, level=0)

pred2<-vector()
pred2[1]<-predict(models1, newVal2, level=0)
pred2[2]<-predict(models2, newVal2, level=0)
pred2[3]<-predict(models3, newVal2, level=0)
pred2[4]<-predict(models4, newVal2, level=0)
pred2[5]<-predict(models5, newVal2, level=0)
pred2[6]<-predict(models6, newVal2, level=0)
pred2[7]<-predict(models7, newVal2, level=0)
pred2[8]<-predict(models8, newVal2, level=0)
pred2[9]<-predict(models9, newVal2, level=0)
pred2[10]<-predict(models10, newVal2, level=0)
pred2[11]<-predict(models11, newVal2, level=0)
pred2[12]<-predict(models12, newVal2, level=0)
pred2[13]<-predict(models13, newVal2, level=0)
pred2[14]<-predict(models14, newVal2, level=0)
pred2[15]<-predict(models15, newVal2, level=0)
pred2[16]<-predict(models16, newVal2, level=0)
pred2[17]<-predict(models17, newVal2, level=0)
pred2[18]<-predict(models18, newVal2, level=0)
pred2[19]<-predict(models19, newVal2, level=0)
pred2[20]<-predict(models20, newVal2, level=0)

pred1<-(pred1*factor)/100000
pred2<-(pred2*factor)/100000

for (i in 1:20){
if(pred1[i]<0){pred1[i]<-0}
}

for (i in 1:20){
if(pred2[i]<0){pred2[i]<-0}
}

pred=pred1+pred2
ti<-2015:2034
bp1<-rbind(pred1,ti)
bp2<-rbind(pred2,ti)
bp<-rbind(pred,ti)
colnames(bp)<-2015:2034
colnames(bp1)<-2015:2034
colnames(bp2)<-2015:2034

if (input$col=="Screening") {
pdf(file,width=15, height=6)
par(mar=c(5,10,10,2))
x<-barplot(bp1[1,],col="steelblue1",ylim=c(min(pred1)*0.85,1.2*max(pred1)),xpd=F,
main=paste("Prediction of the number of", input$col, "colonoscopies in", input$pob))
graphics::text(x,bp1[1,]+rep(c(max(pred1)/10,max(pred1)/20),10),
labels=round(bp1[1,],0),cex=1.2,col="tomato4")
dev.off()
}

if (input$col=="Surveillance") {
pdf(file,width=15, height=6)
par(mar=c(5,10,10,2))
x<-barplot(bp2[1,],col="steelblue1",ylim=c(min(pred2)*0.85,1.3*max(pred2)),xpd=F,

main=paste("Prediction of the number of", input$col, "colonoscopies in", input$pob))
graphics::text(x,bp2[1,]+rep(c(max(pred2)/4,max(pred2)/8),10),
```

```
labels=round(bp2[1,],0),cex=1.2,col="tomato4")
dev.off()
}
if (input$col=="All") {
pdf(file,width=15, height=6)
par(mar=c(5,10,10,2))
x<-barplot(bp[1,],col="steelblue1",ylim=c(min(pred)*0.85,1.3*max(pred)),xpd=F,
main=paste("Prediction of the number of", input$col, "colonoscopies in", input$pob))
graphics::text(x,bp[1,]+rep(c(max(pred)/5,max(pred)/8),10),

labels=round(bp[1,],0),cex=1.2,col="tomato4")
dev.off()
}})
```

D.5. Server: Introduce Password

```
USER <<- reactiveValues(Logged = Logged)

observe({
if (USER$Logged == FALSE) {
if (!is.null(input$Login)) {
if (input$Login > 0) {
Username <- isolate(input$userName)
Password <- isolate(input$passwd)
Id.username <- which(my_username == Username)
Id.password <- which(my_password == Password)
if (length(Id.username) > 0 & length(Id.password) > 0) {
if (Id.username == Id.password) {
USER$Logged <<- TRUE
}}}}}})
observe({

if (USER$Logged == FALSE) {
output$page <- renderUI({
div(class="outer",do.call(bootstrapPage,c("",ui1())))
})
}

if (USER$Logged == TRUE)

{
```

D.6. Server: Interface code (UI)

```
output$page <- renderUI({fluidPage(
div(id="start",titlePanel(HTML('<table style="width:100%">
<tr>
<td style="width:5%;padding-right:3%">
</td>
<td style="width:95%;padding-left:16.2%">
<font face="arial" color=003399 size=6>CRC SCREENING PREDICTOR IN CATALONIA</font>
</td>
<td style="width:-100%;padding-right:3%">
</td>
</tr>
</table>'
))),
div(fluidPage(
plotOutput('regPlot'),

hr(),
```

```
wellPanel(
fluidRow(
column(5,

sliderInput(inputId = "par",
label = "Participation:", min = 0.1, max = 100, value = 50, step= 0.1),

sliderInput(inputId = "pos",
label = "Positivity:", min = 2, max = 10, value = 5, step= 0.01),
sliderInput(inputId = "adh",label = "Adherence:",  min = 0.1, max = 100, value = 50, step= 0.1)

),
column(6, offset = 1,
selectInput(inputId = "pob",
label = "Population:",
choices = c(" Alt Pirineu i Aran La Cerdanya",
" Alt Pirineu i Aran Cerdanya, Pallars, Urgell",
" Barcelona - Barcelona Dreta",
" Barcelona - Ciutat Esquerra i Litoral",
" Barcelona Nord B Nord i Baix Maresme",
" Barcelona Nord el Prat",
" Barcelona Nord Vallès Occidental Oest",
" Barcelona Nord Vallès Oriental"," Barcelona ciutat Nord",
" Barcelona Nord V Occ est",
" Barcelona - Sud Alt Penedès",
" Barcelona - Sud Alt Penedès Garraf",
" Barcelona - Sud Baix Llobregat Centre-Litoral i LH.LL.",
" Barcelona - Sud Baix Llobregat Nord",
" Barcelona - Sud l'Hospitalet LL",
" Barcelona - Sud Maresem Central (o BN Baix maresme, doble)",
" Camp de Tarragona Alt Camp-Conca",
" Camp de Tarragona Baix Camp-Priorat",
" Camp Tarragona Tarragonès-Baix Penedès",
" Catalunya Central Anoia",
" Catalunya Central Osona",
" Catalunya Central Solsonès, Bages, Berguedà",
" Girona Alt Empordà",
" Girona Alt Maresme",
" Girona Baix Empordà, Garrotxa, Ripollès",
" Girona Gironès",
" Girona Selva",
" Lleida tot excepte LLeida",
" Terres Ebre  Baix Ebre",
" Terres Ebre  Montsià + Altebrat (resta TE)"
),
selected = " Barcelona-Sud l'Hospitalet LL"),

selectInput(inputId = "col",
label = "Type of colonoscopy:",
choices = c("Screening","Surveillance","All"),
selected = "Screening"),

column(5, offset = 1,downloadButton('downloadData', 'Download XLS')),

column(1,offset = 1, downloadButton('downloadpdf', 'Download plot')),

HTML('<table style="width:100%">
<tr><td style="width:5%;padding-right:3%"></td></tr>
</table>'),

hr(),

HTML('<center><font face="Comic sans MS,arial,verdana" size=2> Created by </font></center><center>
<font face="Comic sans MS,arial,verdana" size=2 color=003399>Javier Louro</font></center><center>')

),

bsTooltip("par", "Introduce the percentage of participation"),
bsTooltip("pob", "Introduce the population to predict"),
bsTooltip("col", "Introduce the type of colonoscopy"),
bsTooltip("pos", "Introduce the percentage of positivity"),
bsTooltip("adh", "Introduce the percentage of adherence"),
```

```
hr()
)))))
})
}
})
```

D.7. Server: Read the data and models

```
read <- function(){

progress <- shiny::Progress$new(session, min=1, max=3)
progress$set(message = "Reading data\cdots",value=1)
on.exit(progress$close())

data<-read.spss("res.sav", to.data.frame=T)
data$Participacio<-data$Participacio*100
data$Positivitat<-data$Positivitat*100

data$Tiempo<-data$TNOW+0.0001
return(data)

}

data <- read()
readcrib1 <- function(){
progress <- shiny::Progress$new(session, min=1, max=3)
progress$set(message = "Reading data\cdots", value=4)

on.exit(progress$close())
modelc1<-with(subset(data,Tiempo==Tiempo[1]),lm(NColonoCribado~ Participacio+Positivitat
+Positivitat:Participacio))
return(modelc1)

}
modelc1 <- readcrib1()

readcrib2 <- function(){
progress <- shiny::Progress$new(session, min=1, max=3)
progress$set(message = "Reading data\cdots", value=4)

on.exit(progress$close())
modelc2<-with(subset(data,Tiempo==Tiempo[2]),lm(NColonoCribado~ Participacio+Positivitat
+Positivitat:Participacio))
return(modelc1)

}
modelc2 <- readcrib2()
readcrib3 <- function(){
progress <- shiny::Progress$new(session, min=1, max=3)

progress$set(message = "Reading data\cdots", value=4)

on.exit(progress$close())
modelc3<-with(subset(data,Tiempo==Tiempo[3]),lm(NColonoCribado~ Participacio+Positivitat
+Positivitat:Participacio))
return(modelc3)

}
modelc3 <- readcrib3()
readcrib4 <- function(){
progress <- shiny::Progress$new(session, min=1, max=3)
progress$set(message = "Reading data\cdots", value=4)

on.exit(progress$close())
modelc4<-with(subset(data,Tiempo==Tiempo[4]),lm(NColonoCribado~ Participacio+Positivitat
+Positivitat:Participacio))
```

```
return(modelc4)

}
modelc4 <- readcrib4()
readcrib5 <- function(){
progress <- shiny::Progress$new(session, min=1, max=3)
progress$set(message = "Reading data\cdots", value=4)

on.exit(progress$close())
modelc5<-with(subset(data,Tiempo==Tiempo[5]),lm(NColonoCribado~ Participacio+Positivitat
+Positivitat:Participacio))
return(modelc5)

}
modelc5 <- readcrib5()
readcrib6 <- function(){
progress <- shiny::Progress$new(session, min=1, max=3)
progress$set(message = "Reading data\cdots", value=4)

on.exit(progress$close())
modelc6<-with(subset(data,Tiempo==Tiempo[6]),lm(NColonoCribado~ Participacio+Positivitat
+Positivitat:Participacio))
return(modelc6)

}
modelc6 <- readcrib6()
readcrib7 <- function(){
progress <- shiny::Progress$new(session, min=1, max=3)
progress$set(message = "Reading data\cdots", value=4)

on.exit(progress$close())
modelc7<-with(subset(data,Tiempo==Tiempo[7]),lm(NColonoCribado~ Participacio+Positivitat
+Positivitat:Participacio))
return(modelc7)

}
modelc7 <- readcrib7()
readcrib8 <- function(){
progress <- shiny::Progress$new(session, min=1, max=3)
progress$set(message = "Reading data\cdots", value=4)

on.exit(progress$close())
modelc8<-with(subset(data,Tiempo==Tiempo[8]),lm(NColonoCribado~ Participacio+Positivitat
+Positivitat:Participacio))
return(modelc8)

}
modelc8 <- readcrib8()
readcrib9 <- function(){
progress <- shiny::Progress$new(session, min=1, max=3)
progress$set(message = "Reading data\cdots", value=4)

on.exit(progress$close())
modelc9<-with(subset(data,Tiempo==Tiempo[9]),lm(NColonoCribado~ Participacio+Positivitat
+Positivitat:Participacio))
return(modelc9)

}
modelc9 <- readcrib9()
readcrib10 <- function(){
progress <- shiny::Progress$new(session, min=1, max=3)
progress$set(message = "Reading data\cdots", value=4)

on.exit(progress$close())
modelc10<-with(subset(data,Tiempo==Tiempo[10]),lm(NColonoCribado~ Participacio+Positivitat
+Positivitat:Participacio))
return(modelc10)

}
modelc10 <- readcrib10()
readcrib11 <- function(){
progress <- shiny::Progress$new(session, min=1, max=3)
```

```
progress$set(message = "Reading data\cdots", value=4)

on.exit(progress$close())
modelc11<-with(subset(data,Tiempo==Tiempo[11]),lm(NColonoCribado~ Participacio+Positivitat
+Positivitat:Participacio))
return(modelc11)

}
modelc11 <- readcrib11()
readcrib12 <- function(){
progress <- shiny::Progress$new(session, min=1, max=3)
progress$set(message = "Reading data\cdots", value=4)

on.exit(progress$close())
modelc12<-with(subset(data,Tiempo==Tiempo[12]),lm(NColonoCribado~ Participacio+Positivitat
+Positivitat:Participacio))
return(modelc12)

}
modelc12 <- readcrib12()
readcrib13 <- function(){
progress <- shiny::Progress$new(session, min=1, max=3)
progress$set(message = "Reading data\cdots", value=4)

on.exit(progress$close())
modelc13<-with(subset(data,Tiempo==Tiempo[13]),lm(NColonoCribado~ Participacio+Positivitat
+Positivitat:Participacio))
return(modelc13)

}

readcrib14 <- function(){
progress <- shiny::Progress$new(session, min=1, max=3)
progress$set(message = "Reading data\cdots", value=4)

on.exit(progress$close())
modelc14<-with(subset(data,Tiempo==Tiempo[14]),lm(NColonoCribado~ Participacio+Positivitat
+Positivitat:Participacio))
return(modelc14)

}
modelc14 <- readcrib14()
readcrib15 <- function(){
progress <- shiny::Progress$new(session, min=1, max=3)
progress$set(message = "Reading data\cdots", value=4)

on.exit(progress$close())
modelc15<-with(subset(data,Tiempo==Tiempo[15]),lm(NColonoCribado~ Participacio+Positivitat
+Positivitat:Participacio))
return(modelc15)

}
modelc15 <- readcrib15()
readcrib16<- function(){
progress <- shiny::Progress$new(session, min=1, max=3)
progress$set(message = "Reading data\cdots", value=4)

on.exit(progress$close())
modelc16<-with(subset(data,Tiempo==Tiempo[16]),lm(NColonoCribado~ Participacio+Positivitat
+Positivitat:Participacio))
return(modelc16)

}
modelc16 <- readcrib16()
readcrib17 <- function(){
progress <- shiny::Progress$new(session, min=1, max=3)
progress$set(message = "Reading data\cdots", value=4)

on.exit(progress$close())
modelc17<-with(subset(data,Tiempo==Tiempo[17]),lm(NColonoCribado~ Participacio+Positivitat
+Positivitat:Participacio))
return(modelc17)
```

```
}
modelc17 <- readcrib17()
readcrib18 <- function(){
progress <- shiny::Progress$new(session, min=1, max=3)
progress$set(message = "Reading data\cdots", value=4)

on.exit(progress$close())
modelc18<-with(subset(data,Tiempo==Tiempo[18]),lm(NColonoCribado~ Participacio+Positivitat
+Positivitat:Participacio))
return(modelc18)

}
modelc18 <- readcrib18()
readcrib19 <- function(){
progress <- shiny::Progress$new(session, min=1, max=3)
progress$set(message = "Reading data\cdots", value=4)

on.exit(progress$close())
modelc19<-with(subset(data,Tiempo==Tiempo[19]),lm(NColonoCribado~ Participacio+Positivitat
+Positivitat:Participacio))
return(modelc19)

}
modelc19 <- readcrib19()
readcrib20 <- function(){
progress <- shiny::Progress$new(session, min=1, max=3)
progress$set(message = "Reading data\cdots", value=4)

on.exit(progress$close())
modelc20<-with(subset(data,Tiempo==Tiempo[18]),lm(NColonoCribado~ Participacio+Positivitat
+Positivitat:Participacio))
return(modelc20)

}
modelc20 <- readcrib20()

readmode1 <- function(){
progress <- shiny::Progress$new(session, min=1, max=3)
progress$set(message = "Loading model\cdots",value=4)

on.exit(progress$close())
models1<-with(subset(data,Tiempo==Tiempo[1]),lm(vNColonoSeguimiento~ Participacio+Positivitat
+Adherencia+Participacio:Positivitat))
return(models1)

}


models1 <- readmodels1()
readmodels2 <- function(){
progress <- shiny::Progress$new(session, min=1, max=3)
progress$set(message = "Loading modelsl\cdots",value=4)

on.exit(progress$close())
models2<-with(subset(data,Tiempo==Tiempo[2]),lm(vNColonoSeguimiento~ Participacio+Positivitat
+Adherencia))
return(models2)

}


models2 <- readmodels2()
readmodels3 <- function(){
progress <- shiny::Progress$new(session, min=1, max=3)
progress$set(message = "Loading modelsl\cdots",value=4)

on.exit(progress$close())
models3<-with(subset(data,Tiempo==Tiempo[3]),lm(vNColonoSeguimiento ~ Participacio+Positivitat
+Adherencia+Participacio:Positivitat))
return(models3)
```

```
}


models3 <- readmodels3()
readmodels4 <- function(){
progress <- shiny::Progress$new(session, min=1, max=3)
progress$set(message = "Loading modelsl\cdots",value=4)

on.exit(progress$close())
models4<-with(subset(data,Tiempo==Tiempo[4]),lm(vNColonoSeguimiento~ Participacio+Positivitat+
Adherencia+Participacio:Positivitat))
return(models4)

}


models4 <- readmodels4()
readmodels5 <- function(){
progress <- shiny::Progress$new(session, min=1, max=3)
progress$set(message = "Loading modelsl\cdots",value=4)

on.exit(progress$close())
models5<-with(subset(data,Tiempo==Tiempo[5]),lm(vNColonoSeguimiento~  I(Participacio^2)+Adherencia
+I(Positivitat^2)))
return(models5)

}


models5 <- readmodels5()
readmodels6 <- function(){
progress <- shiny::Progress$new(session, min=1, max=3)
progress$set(message = "Loading modelsl\cdots",value=4)

on.exit(progress$close())
models6<-with(subset(data,Tiempo==Tiempo[6]),lm(vNColonoSeguimiento~ Participacio+Positivitat
+Adherencia+Participacio:Positivitat))
return(models6)

}


models6 <- readmodels6()
readmodels7 <- function(){
progress <- shiny::Progress$new(session, min=1, max=3)
progress$set(message = "Loading modelsl\cdots",value=4)

on.exit(progress$close())
models7<-with(subset(data,Tiempo==Tiempo[7]),lm(vNColonoSeguimiento~ Participacio+Positivitat
+Adherencia+Participacio:Positivitat))
return(models7)

}


models7 <- readmodels7()
readmodels8 <- function(){
progress <- shiny::Progress$new(session, min=1, max=3)
progress$set(message = "Loading modelsl\cdots",value=4)

on.exit(progress$close())
models8<-with(subset(data,Tiempo==Tiempo[8]),lm(vNColonoSeguimiento~ Participacio+Positivitat
+Adherencia+Participacio:Positivitat))
return(models8)

}


models8 <- readmodels8()
readmodels9 <- function(){
progress <- shiny::Progress$new(session, min=1, max=3)
progress$set(message = "Loading modelsl\cdots",value=4)
```

```
on.exit(progress$close())
models9<-with(subset(data,Tiempo==Tiempo[9]),lm(vNColonoSeguimiento~ Participacio+Positivitat
+Adherencia))
return(models9)

}


models9 <- readmodels9()
readmodels10 <- function(){
progress <- shiny::Progress$new(session, min=1, max=3)
progress$set(message = "Loading modelsl\cdots",value=4)

on.exit(progress$close())
models10<-with(subset(data,Tiempo==Tiempo[10]),lm(vNColonoSeguimiento~  I(Participacio^2)
+Adherencia+I(Positivitat^2)))
return(models10)

}


models10 <- readmodels10()
readmodels11 <- function(){
progress <- shiny::Progress$new(session, min=1, max=3)
progress$set(message = "Loading modelsl\cdots",value=4)

on.exit(progress$close())
models11<-with(subset(data,Tiempo==Tiempo[11]),lm(vNColonoSeguimiento~ Participacio+Positivitat+
Adherencia))
return(models11)

}


models11 <- readmodels11()
readmodels12 <- function(){
progress <- shiny::Progress$new(session, min=1, max=3)
progress$set(message = "Loading modelsl\cdots",value=4)

on.exit(progress$close())
models12<-with(subset(data,Tiempo==Tiempo[12]),lm(vNColonoSeguimiento~  I(Participacio^2)
+Adherencia+I(Positivitat^2)))
return(models12)

}


models12 <- readmodels12()
readmodels13 <- function(){
progress <- shiny::Progress$new(session, min=1, max=3)
progress$set(message = "Loading modelsl\cdots",value=4)

on.exit(progress$close())
models13<-with(subset(data,Tiempo==Tiempo[13]),lm(vNColonoSeguimiento~ Participacio+Positivitat+
Adherencia+Participacio:Positivitat))
return(models13)

}


models13 <- readmodels13()
readmodels14 <- function(){
progress <- shiny::Progress$new(session, min=1, max=3)
progress$set(message = "Loading modelsl\cdots",value=4)

on.exit(progress$close())
models14<-with(subset(data,Tiempo==Tiempo[14]),lm(vNColonoSeguimiento~ Participacio+Positivitat+
Adherencia+Participacio:Positivitat))
return(models14)

}
```

```
models14 <- readmodels14()
readmodels15 <- function(){
progress <- shiny::Progress$new(session, min=1, max=3)
progress$set(message = "Loading modelsl\cdots",value=4)

on.exit(progress$close())
models15<-with(subset(data,Tiempo==Tiempo[15]),lm(vNColonoSeguimiento~ Participacio+Positivitat+
Adherencia+Participacio:Positivitat))
return(models15)

}


models15 <- readmodels15()
readmodels16 <- function(){
progress <- shiny::Progress$new(session, min=1, max=3)
progress$set(message = "Loading modelsl\cdots",value=4)

on.exit(progress$close())
models16<-with(subset(data,Tiempo==Tiempo[16]),lm(vNColonoSeguimiento~ Participacio+Positivitat
+Adherencia))
return(models16)

}


models16 <- readmodels16()
readmodels17 <- function(){
progress <- shiny::Progress$new(session, min=1, max=3)
progress$set(message = "Loading modelsl\cdots",value=4)

on.exit(progress$close())
models17<-with(subset(data,Tiempo==Tiempo[17]),lm(vNColonoSeguimiento~ Participacio+Positivitat+
Adherencia+Participacio:Positivitat))
return(models17)

}


models17 <- readmodels17()
readmodels18 <- function(){
progress <- shiny::Progress$new(session, min=1, max=3)
progress$set(message = "Loading modelsl\cdots",value=4)

on.exit(progress$close())
models18<-with(subset(data,Tiempo==Tiempo[18]),lm(vNColonoSeguimiento~ Participacio+Positivitat
+Adherencia))
return(models18)

}


models18 <- readmodels18()
readmodels19 <- function(){
progress <- shiny::Progress$new(session, min=1, max=3)
progress$set(message = "Loading modelsl\cdots",value=4)

on.exit(progress$close())
models19<-with(subset(data,Tiempo==Tiempo[19]),lm(vNColonoSeguimiento~ Participacio+Positivitat
+Adherencia+Participacio:Positivitat))
return(models19)

}


models19 <- readmodels19()
readmodels20 <- function(){
progress <- shiny::Progress$new(session, min=1, max=3)
progress$set(message = "Loading modelsl\cdots",value=4)
```

```
on.exit(progress$close())
models20<-with(subset(data,Tiempo==Tiempo[20]),lm(vNColonoSeguimiento~ Participacio
+Positivitat+Adherencia:Participacio))
return(models20)

}


mode20 <- readmode20()

})
```

D.8. UI: Connection to Server

```
### ~APP UI~ ###

library(shiny)

library(shiny)
library(shinyBS)
library(shinyjs)
library(nlme)

library(foreign)
shinyUI(fluidPage(

(htmlOutput("page")), plotOutput('plot', width = "100px", height = "100px")

))
```

References

[1] Banks J. (1998). *Handbook of Simulation*. Engineering & Management Press.

[2] Botteri E., Iodice S., Bagnardi V., Raimondi S., Lowenfels A.B., Maisonneuve P. (2008). *Smoking and colorectal cancer: a meta-analysis*. JAMA; 300(3):2765-78.

[3] Bouvard V., Loomis D., Guyton K.Z., Grosse Y., Ghissassi F.., Benbrahim-Tallaa L., Guha N., Mattock H., Straif K. (2015). *International Agency for Research on Cancer Monograph Working Group. Carcinogenicity of consumption of red and processed meat.* Lancet Oncol, Oct 23.

[4] Box, G.E.P., Draper N. R., (1987). *Empirical Model-Building and Response Surfaces*. John Wiley & Sons, New York, NY.

[5] Bray F., Ren J.S., Masuyer E., Ferlay J. (2012). *Estimates of global cancer prevalence for 27 sites in the adult population in 2008*. Int J Cancer, 2013 Mar 1; 132(5):1133-45.

[6] Carroll L. (1872). *Through the looking-glass and what Alice found there*. Macmillan&Co, London.

[7] Comas M., Castells X., Hoffmeister L., Román R., Cots F., Mar J., Gutiérrez-Moreno S., Espallargues,E. (2008). *Discrete-Event Simulation Applied to Analysis of Waiting Lists. Evaluation of a Prioritization System for Cataract Surgery*. Value in health 11(7).

[8] Comas M. (2008). *Application of discrete-event simulation to health services research: analysis of needs and demand for elective*. Universitat Politècnica de Catalunya. Departament d'Estadística i Investigació Operativa.

[9] Cunningham D., Atkin W., Lenz H.J., Lynch H.T., Minsky B., Nordlinger B., Starling N. (2010). *Colorectal cancer*. Lancet. Mar 20; 375(9719):1030-47.

[10] Eco U. (1989). *Il pendolo di Foucalt*. Lumen pag 215.

[11] Ferlay J., Soerjomataram I., Ervik M., Dikshit R., Eser S., Mathers C., Rebelo M., Parkin D.M., Forman D., Bray F. (2013). *GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11*. Lyon, France. International Agency for Research on Cancer; 2013. Available from: http://globocan.iarc.fr, accessed on 07/11/2015.

[12] Giovannucci E., Ascherio A., Rimm E.B., Colditz G.A., Stampfer M.J., Willett W.C. (1995). *Physical activity, obesity, and risk for colon cancer and adenoma in men.* Ann Intern Med. Mar 1; 122(5):327-34.

[13] Law A. M., Kelton W.D. (2000). *Simulation modeling and analysis, third edition*. Mc Graw Hill.

[14] Lutgens M.W., van Oijen M.G., van der Heijden G.J., Vleggaar F.P., Siersema P.D., Oldenburg B. (2013). *Declining risk of colorectal cancer in inflammatory bowel disease: an updated meta-analysis of population-based cohort studies.* Inflamm Bowel Dis. Mar-Apr; 19(4):789-99.

[15] Segnan N., Patnick J., von Karsa L. (2010). *European guidelines for quality assurance in colorectal cancer screening and diagnosis*. Office for official publications of the European

Communities. Luxembourg, Publications Office of the European Union. First edition 2010. Available from: http://bookshop.europa.eu/en/european-guidelines-for-quality-assurance-in-colorectal-cancer-screening-and-diagnosis-pbND3210390, accessed on 07/11/2015.

[16] Tukey J. (1962). *The future of data analysis.* Annals of Mathematical Statistics; 33:13.

[17] Winawer S.J., Zauber A.G., O'Brien M.J., Ho M.N., Gottlieb L., Sternberg S.S., Waye J.D., Bond J., Schapiro M., Stewart E.T. (1993). *Randomized comparison of surveillance intervals after colonoscopic removal of newly diagnosed adenomatous polyps.* The National Polyp Study Workgroup.N Engl J Med; 328:901-6.

Online resources:

[I] Globocan:
http://globocan.iarc.fr/
(accessed on 07/11/2015).

[II] Las cifras del cancer 2014:
http://www.seom.org/seomcms/images/stories/recursos/Las_cifras_del_cancer_2014.pdf
(accessed on 09/11/2015).

[III] U. S. Preventive Services Task Force. Final Recommendation Statement Colorectal Cancer:
http://www.uspreventiveservicestaskforce.org/Page/Document/RecommendationStatementFinal/colorectal-cancer-screening
(accessed on 07/12/2015).

[IV] Generalitat de Catalunya:
http://premsa.gencat.cat/pres_fsvp/AppJava/salut/notapremsavw/278439/ca/comenca-la-implantacio-del-programa-de-deteccio-precoc-cancer-colon-recte-les-comarques-dosona-garrotxa-el-ripolles.do
(accessed on 10/12/2015)

[V] INEbase:
http://www.ine.es/inebmenu/indice.htm.
(accessed on 11/12/2015).