MASTER IN ACTUARIAL SCIENCE

MASTER'S FINAL WORK

INTERNSHIP REPORT

INFERENCE OF THE RISK ASSOCIATED WITH A FUTURE PANDEMIC FOR A HEALTH INSURANCE COMPANY

MARIA RODRIGUES DE CARVALHO GARCIA DE CASTRO

SUPERVISION:

ALFREDO DUARTE EGÍDIO DOS REIS
MARLI ANDREIA MONTEIRO AMORIM FERREIRA

OCTOBER - 2021

GLOSSARY

AD Anderson-Darling. i, 20

C Estimated total cost in the reference scenario. i, 25, 27, 30

COVID-19 Coronavirus disease - in 2019. i–iii, 1–3, 13, 14, 17, 19, 23, 26, 29, 30, 34

CS Chi-Squared. i, 20

CVM Cramer Von Mises. i, 20

DGS Direção Geral da Saúde. i, 13, 14

H Nº of estimated hospitalizations in the reference scenario. i, 25, 27, 30

$H^{w\ ICU}$ Nº of estimated hospitalizations with ICU in the reference scenario. i, 25

$H^{w/o\ ICU}$ Nº of estimated hospitalizations w/o ICU in the reference scenario. i, 25

I Nº of estimated infections in the reference scenario. i, 25, 27, 30

ICU Intensive Care Unit. i, 13–17, 19–23

IPEs Individual Protection Equipments. i, ii, 3, 4, 17, 25

K-S Kolmogorov-Smirnov. i, 20

LoS Length of Stay. i, 19–24

MERS Middle East respiratory syndrome. i, 28–30

P Estimated aggregated risk premium in the reference scenario. i, 25

PCR Polymerase Chain Reaction. i, 3–5, 17, 25, 26

SARS Severe Acute respiratory syndrome. i, 28–30

WHO World Health Organization. i, 1, 2

ABSTRACT

A pandemic is characterized by the development of severe unknown disease in many people, spreading over several regions. The current Coronavirus disease - in 2019 (COVID-19) pandemic has brought new challenges for the entire health sector. Health insurance companies did not have any cover expenses in relative to pandemic diseases. They use protective clauses since the risk is too complex as well as high. The current pandemic forced insurers to study future coverages for that hand of risks. The goal of this study is to study and estimate associated costs of medical treatments of future pandemics, taking this one as a good example. At the beginning of the pandemic, the own company decided to cover hospital admissions for COVID-19, tests performed to track the disease, and Individual Protection Equipments (IPEs), which began to be used more frequently due to pandemic disease. Therefore, to estimate the cost of a pandemic, we considered these three variables (hospitalizations, tests, and IPEs) using the data related to the COVID-19 pandemic. The main challenge of the project was to estimate the number of hospitalizations that the insurer would have to support, taking into account the incidences, the hospitalization rate, and the limit of available beds. In this sense, we used a simulation procedure. It was possible to know how many beds were occupied each day by patients, and thus know how many of the hospitalizations that were estimated to have is that the company would pay them. It is impossible to predict the characteristics of a future pandemic, so in addition to the characteristics of COVID-19, we have also carried out stress tests to evaluate more extreme situations.

KEYWORDS: Health insurance; Pandemic; COVID-19; Simulation; Bootstrap.

RESUMO

Uma pandemia tem como característica o desenvolvimento de uma doença grave não conhecida, num grande número de pessoas, que se encontram espalhadas por várias regiões. A atual pandemia COVID-19 trouxe novos desafios para todo o sector da saúde. As companhias de seguros de saúde não cobriam despesas relacionadas com doenças pandémicas. Estão protegidas por uma cláusula, dada a complexidade do risco envolvido. A atual pandemia tornou mais pertinente a discussão da eventual cobertura destes produtos. O objetivo do presente trabalho é inferir quanto ao risco associado aos custos com tratamento médico de uma futura pandemia. A seguradora de acolhimento decidiu, no início da pandemia, cobrir os internamentos hospitalares por COVID-19, os testes realizados para rastrear a doença e os EPIs, que passaram a ser utilizados com maior frequência devido à doença pandémica. Portanto para estimar o custo relacionado com uma pandemia, consideramos estas três variáveis usando os dados relativos à pandemia COVID-19. O principal desafio do projeto foi estimar o número de internamentos que a seguradora teria de pagar tendo em conta as incidências, a taxa de hospitalização e o limite de camas disponíveis. Neste sentido, utilizamos um problema de simulação no qual era possível saber quantas camas estavam ocupadas em cada dia, por clientes, e assim saber quantos dos internamentos que se estimou ter é que seriam realmente pagos pela companhia. Não se consegue prever as características que terá uma futura pandemia pelo que para além das características da COVID-19 , também realizámos stress tests para avalizar realidades mais extremas.

PALAVRAS-CHAVE: Seguro de saúde; Pandemia; COVID-19; Simulação; Bootstrap.

TABLE OF CONTENTS

List of Figures

List of Tables

ACKNOWLEDGEMENTS

I carried out this project during challenging times due to the COVID-19 pandemic crisis. I was motivated by the relevance of the theme for the situation we are living and it turned out to be an interesting and gratifying experience. I am grateful to many people for their help and support during this work. I would like to pay my special regards to Professor Alfredo Egídio dos Reis for supervising this project. I would like to express my gratefulness to my supervisor at Multicare, Marli Ferreira, for the theoretical knowledge she shared, her tireless support, and her availability during the entire process. In addition, I would like to thank my colleagues at Multicare for making me feel welcome right from the start and for all the laughter. I would like to thank my friends for their willingness to listen and for all their help. Finally, I am grateful for the support of my family and all the opportunities they have given me.

INFERENCE OF THE RISK ASSOCIATED WITH A FUTURE PANDEMIC FOR A HEALTH INSURANCE COMPANY

Maria Garcia de Castro

1 INTRODUCTION

The existing internship report results from my work developed at the Actuarial Office from the health insurance company Multicare. Its main objective is to infer the risk associated with the costs of medical treatment of a future pandemic.

1.1 Context

According to the World Health Organization (WHO), an epidemic is a new disease that affects a region, and a pandemic is an epidemic that spreads around the world through human contact.

Since the Black Death in 1347, there have been quite a few pandemics and epidemics reported. The Black Death was the deathliest after the Spanish Flu, in 1918, when 5% of the world's population died. In Portugal, there were estimated from 50 to 70 thousand casualties.

The 2019 new coronavirus, SARS-CoV-2, is a human coronavirus that spread worldwide (Petrosillo et al., 2020). The first case of COVID-19, in Portugal, was reported on the 2^{nd} of March 2020. By the 30^{th} April 2021, Portugal counted with 834 442 cases and 16 965 deaths. On this same day, 146 841 882 people had been infected and 1 104 743 had died with COVID-19 in the world ("WHO Coronavirus (COVID-19) Dashboard", 2019).

When the disease first appeared in Portugal, it was already known that some infected people might need moderate and/or severe hospital care. At the time and until today, almost all the health insurers had a clause in their policies that liberates them from covering any health costs related to pandemic diseases (no. 30 from the 4^{th} clause - general conditions). Despite an increase in outbreaks, there is a small range of pandemic-related products, in the world, given the complexity in pricing them because the risk is not well understood. The COVID-19 pandemic could help change that because it provided an important data source to help insurers better handle its risk and develop a vast range of covers in response like the ones we talked about in the following paragraphs.

Sirago is an underwriting manager that offers products for the South African insurance market. The company looks for market opportunities in particular products. Sirago has already created an original insurance solution to provide a safety net for hospitalized

people due to COVID-19. This pandemic product pays a lump sum stated benefit to infected policyholders hospitalized for more than 48 hours due to pandemic disease. It is only available to people 60 years old or less ("New Pandemic Insurance solution protects consumers against the financial ramifications of Covid-19 hospitalisation – Sirago Underwriting Managers", 2021).

The health system in India is essentially the responsibility of every state, which means that the federal government has no direct influence on the conditions in their health care system. In India, some insurers like Aditya Birla Health Insurance Co. Ltd., Bharti AXA General Insurance Co. Ltd., Cholamandalam MS General Insurance Co. Ltd., Edelweiss General Insurance Co.Ltd., Future Generali India Insurance Co.Ltd. and others have already developed products for pandemics. Bharti AXA General Insurance, for example, has a benefit-based individual product for an insured person who has to stay in a hospital continuously for more than 72 hours to treat COVID-19. The product has a policy period of 105 days, 195 days, and 285 days, respectively, as specified in the policy schedule.The policy is available to people between 18 and 65 years old. The lump sum benefit is equal to 100% of the Sum Insured and is paid with a positive diagnosis of COVID-19 ("IRDAI Welcomes You", 2020).

Discovery ZA has a World Health Organization (WHO) global outbreak benefit. This product covers risk assessment for each individual to understand his risk status at any point in time. In addition, the benefit ensures clients have access to vaccination, online consultations, testing, and cover for out-of-hospital management and appropriate supportive treatment. Related to in-hospital treatment due to COVID-19, the admission is subject to approval and preauthorization and is covered from the Hospital Benefit based on the client's chosen health plan ("WHO Outbreak Benefit", 2020).

Vitality insurance, a company focusing in private medical insurance in UK, also offers some covers to its clients. Vitality health members who have personal health insurance, business health insurance, corporate healthcare, and vitality at work can buy a discounted antigen test. Even though UK health insurance does not cover emergencies, vitality health insurance clients who tested positive for COVID-19 and need hospitalization can get a cash benefit. If the patients stay one to eight days at the hospital, they receive £250 per day. If they stay nine or more days, the benefit is £500 per day, subject to a total of £5000 ("Coronavirus FAQs", 2020).

1.2 Motivation

During the COVID-19 crisis, Multicare, the largest health insurance company in Portugal, came forward and offered to cover its clients' costs. Therefore, the company paid

for all the Polymerase Chain Reaction (PCR) tests carried out by its customers, IPEs and also for the admissions of customers who develop the most severe disease.

While in the beginning not much was known about the disease or the dimension it could take, as of the day of this report there is a greater awareness of the danger pandemics and epidemics represent and of the increasing frequency in which they occur. Therefore, the objective of this paper is to infer the risk associated with coverage of a future pandemic for the health insurance.

1.3 Project and Report Organization

The project was organized in a forward manner where three main goals were defined:

- Collect epidemiological information about COVID-19;
- Risk assessment in case of complete coverage of COVID-19 and exclusively using the private sector;
- Impact analysis for more extreme scenarios.

To infer the risk associated with future pandemics, we start by estimating the total cost of COVID-19 to the company in the first year of the pandemic. We assumed that the first year of the pandemic occurred from 1^{st} March 2020 to 28^{th} February 2021.

This report is divided into seven chapters. In Chapter 2, there is an introduction to the methodological approach. Then, in Chapter 3, the data collected and the first analysis will be showned. Subsequently, Chapter 4 presents the process until calculating the total cost of medical treatments of COVID-19 disease. In Chapter 5, more extreme situations were evaluated. Finally, in Chapter 6, we analyzed some indicators to understand if pandemic risks the company's solvency.

2 THEORETICAL INTRODUCTION

An insurance contract is made between an insurance company and one or more persons called the policyholders. The policyholder agrees to pay an amount to the insurer called premium to transfer the risk from the policyholder to the insurance company.

The amount of the damages that the insurer undertakes to pay in case of a claim, is covered by a premium paid by the insured. Usually, the premium is paid annually.

The annual pure premium, denoted by P, can be considered the average value of the annual cost of the disease, i.e., the average value of C, by $E[C]$. The actuarial premium, denoted by P^*, is composed by the pure premium plus a heading, i.e., $P^* = P + heading$.

$$P = E[C] = E[C|S] \cdot p_S + E[C|\overline{S}] \cdot p_{\overline{S}}. \tag{1}$$

Where C is the cost of the disease for an individual, p_S is the probability of an individual being sick and $p_{\overline{S}} = 1 - p_S$.

The purpose of this insurance is to cover the expenses related to pandemic disease. Usually, $E[C|\overline{S}] = 0$ because if the policyholder does not have the disease, there are no costs for the insurance company. However, in the case of a pandemic product, there are costs, from PCR tests and IPEs, even if the insured is not sick.

The annual cost of the disease is the sum of hospitalizations, PCR tests, and IPEs.

$$E[C] = E[\text{Cost of hospitalizations}] + E[\text{Cost of PCR tests}] + E[\text{Cost of IPEs}], \tag{2}$$

where $E[.]$ is the Expectation operator.

Cost of hospitalizations

The cost of hospitalizations is a combination of admissions with or without ICU. For the insurance company, the cost of hospitalization without ICU and with ICU are, respectively, C_1 and C_2. These values, C_1 and C_2, were previously agreed with private sector hospitals.

In order to evaluate the risk of contracting a disease, i.e., determining the number of infected individuals, there is the incidence rate, I, given by the following formula:

$$I = \frac{\text{no. of cases}}{\text{no. of susceptible individuals}}. \tag{3}$$

The incidence rate differs depending on the age, sex, and region of residence of each person. So it makes sense to calculate an incidence rate for each group segmented by age, sex, and region (Equation 4):

$$I_{\text{a,s,r}} = \frac{\text{no. of cases}_{\text{a,s,r}}}{\text{no. of susceptible individuals}_{\text{a,s,r}}}. \tag{4}$$

Where $a = 0, 1, ..., 8$ if age in years is in the intervals: [0;9], [10;19], ... , [80;∞[; s= M, F and r is the region (Lisbon, North, Center, Alentejo, Algarve, Regional Administrations of Azores, Regional Administrations of Madeira).

The number of hospitalizations, $H_{\text{a,s}}$, is given by the number of infected individuals and the probability of developing severe disease, $p_{\text{a,s}}$, i.e., being hospitalized. In order to determine the probability that an infected individual needs to go to the hospital, it was necessary to use theoretical topics that will be spoken in the following sections.

$$H_{a,s}^{ICU} = \text{no. of estimated cases}_{\text{a,s}} \cdot p_{a,s}^{ICU}. \tag{5}$$

$$H_{a,s}^{w/o\ ICU} = \text{no. of estimated cases}_{\text{a,s}} \cdot p_{a,s}^{w/o\ ICU}. \tag{6}$$

$$E[\text{Cost of hospitalizations}] = H_{a,s}^{w/o\ ICU} \cdot C_1 + H_{a,s}^{ICU} \cdot C_2. \tag{7}$$

Cost of PCR tests

We assumed that the number of PCR tests performed by the insurance company is directly proportional to the number of estimated cases in the portfolio because the data we used do not reject it. The constant of proportionality is α. It was determined considering the number of estimated cases and the number of tests paid by the company. In order to obtain the total cost, it is necessary to multiply by the average amount paid for each test, C_3.

$$E[\text{Cost of PCR tests}] = \alpha \cdot (\text{No. of estimated cases}) \cdot C_3. \tag{8}$$

Cost of IPEs

The cost of EPIs was considered constant in all cases. The amount spent by the company in the first year of the pandemic was used.

$$E[\text{Cost of IPEs}] = \text{Total cost of observed IPEs.} \tag{9}$$

Therefore the total cost can be given by the following expression:

$$E[C] = H_{a,s}^{w/o\ ICU} \cdot C_1 + H_{a,s}^{ICU} \cdot C_2 + \alpha \cdot (\text{No. of estimated cases}) \cdot C_3 + \text{Total cost of observed IPEs.} \tag{10}$$

2.1 Wilcoxon Mann-Whitney Test

The Wilcoxon Mann-Whitney test is a non-parametric test, which can be used to test whether two samples are likely to derive from the same population, i.e., that the two populations have the same shape. It is used when the data are not normally distributed.

Suppose we have two randomly selected values from two populations. One sample of X_1 dimension n_1 and another of X_2 of dimension n_2. X_1 and X_2 are two continuous random variables. It can be assumed that n_2<n_1.

The Wilcoxon-Mann-Whitney test was developed to compare the medians between the two populations, denoted by M_{x_1} and M_{x_2}

$$H_0 : M_{\mathrm{x}_1} = M_{\mathrm{x}_2} \quad vs \quad H_1 : M_{\mathrm{x}_1} \neq M_{\mathrm{x}_2}. \tag{11}$$

Test procedure

First, we joined the two samples in a single sample of $n_1 + n_2$ elements and ordered. The test statistic (T) is the sum of the ranks of the elements of the random sample of X_2, because $n_2 < n_1$. When the values of n_1 and n_2 are high, the random variable T has an approximately Normal distribution with mean (μ_{T}) and standard deviation (σ_{T}).

$$\mu_{\mathrm{T}} = \frac{n_2(n_1 + n_2 + 1)}{2}. \tag{12}$$

$$\sigma_{\mathrm{T}} = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}. \tag{13}$$

$$Z_{\mathrm{T}} = \frac{T - \mu_{\mathrm{T}}}{\sqrt{\sigma_{\mathrm{T}}}} \sim N(0,1). \tag{14}$$

As this is a bilateral test, H_0 is rejected when $|Z_{\mathrm{T}}| > Z_{\alpha/2}$.

α	0.01	0.05	0.1
$Z_{\alpha/2}$	-2.575829	-1.959964	-1.644854

TABLE I: STANDARD NORMAL DISTRIBUTIONL

2.2 Tests of goodness of fit

Chi-Square test

The Chi-Square test is an asymptotic test used to study the conformity of the distribution underlying a given large random sample to a given distribution of a discrete random variable.

The Chi-Square adjustment test checks the hypothesis that observations follow a specific distribution (Equation. 15), discrete or continuous, with or without knowing the parameters. This test consists of comparing the density with the probabilities distribution function. It will be admissible to state that the distribution tested fits, or not, the data with a certain level of confidence. The Chi-Square test is a simple test to use, so it is the most commonly used (Abd-Elfattah, 2011; Stigler, 2012).

$$H_0 : X \sim f_{\mathrm{X}}(x) \quad vs \quad H_1 : X \nsim f_{\mathrm{X}}(x). \tag{15}$$

The n observations of a population sample can be modeled by a sequence of independent random variables $X_1, xX_2, ..., X_{\mathrm{n}}$. Variables can only take integer values $1, ..., K$, which have the same probability mass function, $f_{\mathrm{X}}(x)$. K is the number of categories of the random range. Given the random sample $X_1, X_2, ..., X_{\mathrm{n}}$, denote by O_{i}, with $i = 1, ..., K$ the $X_{\mathrm{j}}^{'}s$ numbers, with $j = 1, ..., n$ having the value equal to i. The O_{i} values are the observed frequency and E_{i} represents the estimated frequency for the $i-th$

category.

$$\sum_{i=1}^{K} O_{\mathrm{i}} = \sum_{i=1}^{K} E_{\mathrm{i}} = n. \tag{16}$$

The test statistic is:

$$X^2 = \sum_{i=1}^{K} \frac{(O_{\mathrm{i}} - E_{\mathrm{i}})^2}{E_{\mathrm{i}}}. \tag{17}$$

For significant level α, H0 is rejected when $\mathrm{X}^2 > \chi^{\,2}{}_{\mathrm{K}(1-\alpha)}$.

Kolmogorov-Smirnov test

The Kolmogorov–Smirnov test is a nonparametric test used to compare a sample with a reference probability distribution.

The test statistic quantifies a distance between the empirical distribution function of the sample and the cumulative distribution function of the reference distribution. The null distribution of this statistic is calculated under the null hypothesis. H_0 states that the sample follows the reference distribution, which can be continuous, discrete, or mixed. Subsequently, it will be compared with a theoretical value. Then it is possible to state whether or not the distribution being tested fits (Stigler, 2012).

$(X_1, X_2, ..., X_{\mathrm{n}})$ is a random sample of a population X. X_{i} for $i = 1, 2, \ldots, n$ are n independent and identically distributed (i.i.d.) observations. F_{n}, the empirical distribution function, is defined as follows:

$$\hat{F}_{\mathrm{n}}(x) = \frac{1}{n}\#\{x_{\mathrm{i}} : x_{\mathrm{i}} \leqslant x\}. \tag{18}$$

$(x_1, x_2, ..., x_{\mathrm{n}})$ is a realization of $(X_1, X_2, ..., X_{\mathrm{n}})$.

The sample distribution function is:

$$F_{\mathrm{n}}(x) = \frac{1}{n}\#\{X_{\mathrm{i}} : X_{\mathrm{i}} \leqslant x\}. \tag{19}$$

Let F's be the distribution function of the population and $F0$ the function of the proposal distribution, which is continuous and specified.

$$H_0 : F(x) = F_0(x) \quad vs \quad H_1 : F(x) \neq F_0(x). \tag{20}$$

In this test the test statistic, D_{n}, is:

$$D_{\mathrm{n}} = sup_{-\infty < \mathrm{x} < +\infty} |F_{\mathrm{n}}(x) - F_0(x)|. \tag{21}$$

The expected value of the test statistic is obtained when replacing the sample distribution function, $F_{\mathrm{n}}(x)$ with the empirical distribution function, $\hat{F}_{\mathrm{n}}(x)$.

$$d_{\mathrm{n}} = sup_{-\infty < \mathrm{x} < +\infty} |\hat{F}_{\mathrm{n}}(x) - F_0(x)|. \tag{22}$$

So if H_0 is true, D_{n} is expected to take a small value. H_0 is rejected, for a significance level α, if the observed value d_{n} from the test statistic D_{n} is greater or equal than the critical point $D_{\mathrm{n},\alpha}$.

The Kolmogorov-Smirnov test is better than the Chi Square test. While the Chi-Square test is the total sum of the square of the distance between the two curves weighted to the expected frequency, the Kolmogorov-Smirnov test analyzes the maximum distance between the empirical and theoretical distributions. In addition, the Chi-Square test is used in large samples, while the K-S test can be applied to small ones.

There are more adjustment tests based on the same principle of comparison between expected values and observed values. They are, for example, the Anderson-Darling (AD) and CramerVon Mises (CVM) adjustment tests. The difference between these two tests lies in the expression of the test statistic and consequently in their respective theoretical values.

The Anderson-Darling and CramerVon Mises tests have the same principle, as both integrate the square of the distance between the two functions. The difference is that the Anderson-Darling test differs in the weighting of this quantity.

In general, Kolmogorov-Smirnov and CramerVon Mises tests are more effective at detecting variations in the middle of the distribution. At the same time, Anderson-Darling is more powerful at highlighting variations in the tail of the distribution (Stigler, 2012).

Anderson-Darling test

Test hypothesis:

$$H_0 : F(x) = F_0(x) \quad vs \quad H_1 : F(x) \neq F_0(x). \tag{23}$$

Test statistic:

$$A_n^2 := n \int \frac{(F_{\mathrm{n}}(x) - F_0(x))^2}{F_0(x)(1 - F_0(x))} dF_0(x). \tag{24}$$

The test statistic uses a quadratic distance between F_{n} and F_0. The distance is weighted by $w(x)$.

$$w(x) = \frac{1}{F_0(x)(1 - F_0(x))}. \tag{25}$$

If H_0 is true, then A_n^2 is supposed to be small. Therefore H_0 is rejected for large values of A_n^2.

Cramer Von Mises test

Test hypothesis:

$$H_0 : F(x) = F_0(x) \quad vs \quad H_1 : F(x) \neq F_0(x). \tag{26}$$

Test statistic:

$$W_n^2 := n \int (F_{\mathrm{n}}(x) - F_0(x))^2 dF_0(x). \tag{27}$$

If H_0 is true, then W_n^2 is supposed to be small. Therefore H_0 is rejected for large values of W_n^2.

2.3 Simulation

Simulation uses specific mathematical techniques, which allow imitating almost real world situations. Simulation models can be considered as a description of real systems. Therefore, running computer simulation models can provide more effective results without needing to interfere with the real system. When analyzed statistically, such results produce information that can significantly contribute to making decisions or solving problems.

To create a random sample from a population that has a known distribution, pseu-

dorandom numbers and observations should be generated. Pseudorandom numbers are values of a random variable with a Uniform(0,1) distribution, while pseudorandom observations are generated using the pseudorandom numbers and the required distribution.

One of the methods to generate pseudorandom numbers is the Lehmer random number generator. The method is inspired by the multiplicative linear congruential generator, introduced by Lehmer in 1949.

The formula is:

$$X_{\text{n+1}} = a \cdot X_{\text{n}}(mod \quad m) \qquad n = 1, 2, ... \tag{28}$$

$$U_{\text{n}} = X_{\text{n}}/m \tag{29}$$

Where U_{n} is the generated random number of order n, a is a high multiplicative order module m element, m is a prime number and, X_0 is called the seed and is coprime to m.

Pseudorandom observations are generated using the pseudorandom numbers. Once the sequence of pseudorandom numbers is created, it is intended to define functions of these variables that follow a required distribution. A general method for generating such a random variable is called the inverse transformation method, which is based on theorem 1.

Theorem 1: Let U be a uniform (0,1) random variable. For any continuous distribution function F the random variable X defined by

$$X = F^{-1}(U) \tag{30}$$

has distribution F. (Ross, 2013)

2.4 Bootstrap confidence intervals

The bootstrap method is a resampling technique having as a basis a computationally intensive methodology. It establishes a new framework for simulation-based statistical analysis. It replicates small size samples using simulation.

One of the applications of this method is the estimation of the confidence interval. Only the bootstrap percentile confidence interval was used in this project. This is the favorite method because it can be used whatever the sampling distribution is and it is the simplest.

Confidence intervals are grounded on the sampling distribution of a bootstrap statistic. The 95% confidence interval can be calculated with the 95% bootstrap statistics. The 95% confidence interval starts by marking off the 95% central elements of the sampling distribution. The critical values in any confidence interval are the marked percentiles that give us the 95% central elements. Therefore, the 2.5^{th} and 97.5^{th} percentiles are the critical values. That is, the interval between the two percentiles of the bootstrap distribution is a confidence interval. Hence, it is known as a bootstrap percentile confidence interval.

In summary, the bootstrap percentile confidence interval for 95% confidence can be either the interval comprised between the 2.5^{th} percentile and the 97.5^{th} percentile of the bootstrap distribution or the interval up to the 95^{th} percentile.

3 EPIDEMIOLOGICAL INFORMATION ABOUT COVID-19

This chapter will describe the type of data collected about COVID-19 in Portugal and how it was analysed.

3.1 Collected Data

We collected the data from COVID-19 daily reports released by Direção Geral da Saúde (DGS). These reports have the number of accumulated cases, deaths and recovered people, and the number of active cases, inpatients and Intensive Care Unit (ICU) inpatients on the day of the report. The number of accumulated cases and deaths is also given by age group, sex, and region.

Age groups are defined in the following ten year intervals: 00-09, 10-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79 and 80+. They are also separated by sex: Male (M) and Female (F). In addiction, we consider seven regions: *Lisbon, North, Center, Alentejo, Algarve and the Regional Administrations of Azores and Madeira*.

At this point, we calculated the new monthly cases by age group and sex. It was also necessary to calculate the new cases segmented by region.

At PORDATA, we collected the distribution of the population and calculated the incidences per month by region, age group and sex.

3.2 Incidence Rates

In the first year of the pandemic, Portugal had an incidence of 7.83%. The most affected regions were the urban centers, such as the *Lisbon* and the *North*, with 10.67% and 9.14%, respectively, as it can be seen in Figure 1. As expected, the Regional Administrations, Azores and Madeira, were the least affected regions.

The incidence in women is slightly higher than men. 7.51% of men and 8.12% of women had been infected with COVID-19. As it can be seen in the table II, the incidence is higher for people aged between 20 to 29 years and those older than 80. This last group can be justified by the several outbreaks in nursing homes.

Sex/Age	00-09	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80+	Total
Male	4.99%	6.96%	9.75%	8.76%	7.93%	7.52%	6.55%	5.90%	9.11%	7.51%
Female	5.01%	7.33%	11.20%	9.90%	9.09%	8.52%	6.07%	5.09%	10.34%	8.12%

TABLE II: INCIDENCE RATE PER GROUP AGE AND SEX IN PORTUGAL

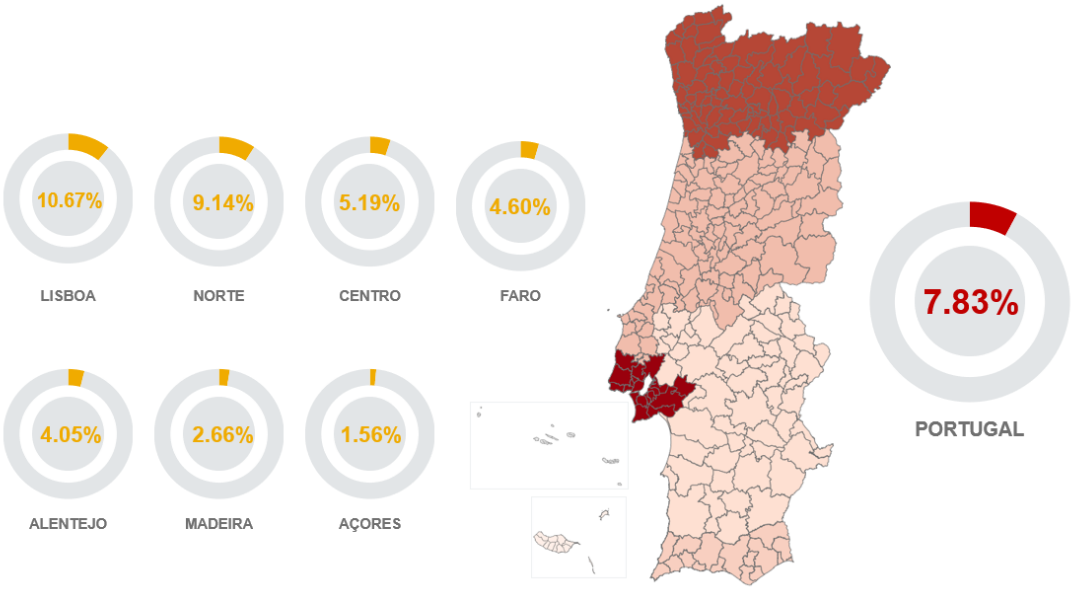

FIGURE 1: Incidence rates per region in Portugal

3.3 Hospitalizations

Regarding the hospitalizations, the DGS's daily reports provide no information besides the number of inpatients each day. Therefore, we could only calculate the occupancy rate of hospitals. In addition, it was necessary to carry out some research in order to have enough information to estimate the admissions in the company portfolio.

Some metrics from other sources will be used in this subsection. It is important to note that COVID-19 is a pandemic that has not ended yet and, therefore, all information disclosed can be constantly updated. That is, the metrics used might not be those most up-to-date.

The hospitalization rate, in Portugal, for people infected with COVID-19, is 12% (Lemos-Paião et al., 2020). It was considered that 25% of the hospitalized individuals need ICU treatments. That means that 3% of infected people need treatment in ICU and the other 9% only need hospital care. In November, during a press conference, the distribution of hospitalization was announced by age group: 4% would be under 40 years of age, 33% between 40 and 69 years and the remaining 63% over 70 years (Neves, 2021).

It was essential to verify if sex is a significant variable for hospitalizations. As Gebhard et al., 2020 refers, despite the similar incidence rate in men and women, the fatality rate and the probability of developing severe disease are higher in men. The data about deaths in Portugal is coherent with this. Therefore, the same assumption will be made for

hospitalizations. As the disease has a more severe impact and is deathlier in men, it was necessary to understand how much more severe it is. With the information provided in the article above, we produced a graphic (Figure 2) showing the male/female ratios in France, Italy, Spain, China, Switzerland and Germany. This graphic shows that France, Spain and Switerzland have similar ratios for deaths and hospitalizations. With this information, we assumed that Portugal would have the same behaviour and also have similar ratios for deaths and hospitalizations. Therefore, the hospitalizations and deaths ratios are assumed to be 1.09.

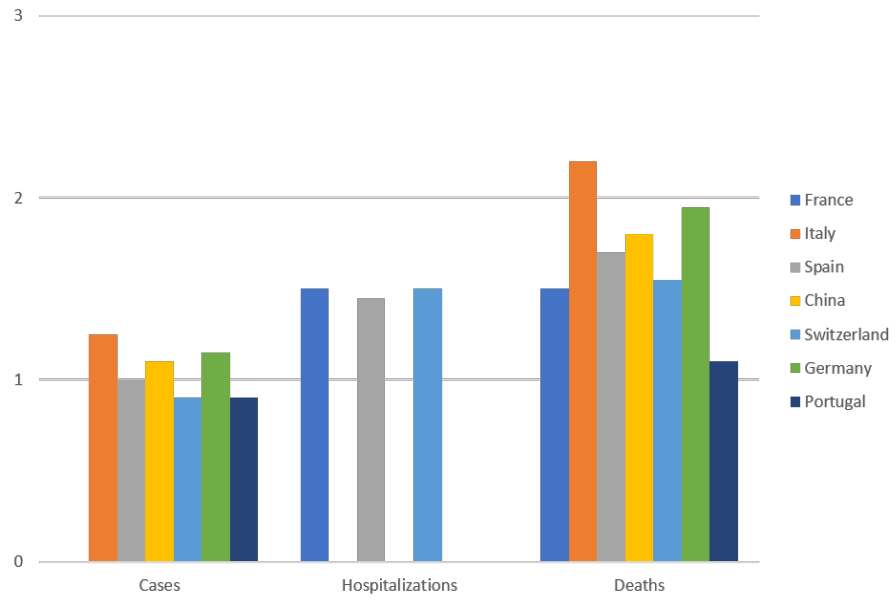

FIGURE 2: Male/Female ratios

From the beginning of the pandemic incidence in Portugal until the end of January 2021, 88.10% of deaths have been recorded in hospitals, 6.9% elsewhere (e.g.: nursing homes), 4.9% at home and 0.06% in unknown locations (Lusa, 2021).

With all of this information above, it was possible to build the following Table III. This table describes the probabilities of an infected individual needing hospital care or ICU care by sex and age.

Sex	Age Group	Hospitalization (With ICU)	Hospitalization (Without ICU)
Female	00-39	0.02%	0.99%
	40-69	0.50%	4.82%
	70+	14.46%	27.26%
Male	00-39	0.03%	1.18%
	40-69	1.37%	13.27%
	70+	22.33%	42.11%

TABLE III: ESTIMATED PROBABILITIES OF AN INFECTED INDIVIDUAL NEEDING HOSPITAL CARE

During the project, we received data from a Portuguese hospital in an urban area.

For each patient, there were:

- Sex
- Group age
- Length of Stay at the hospital excluding ICU
- Length of Stay at ICU

With this data, we calculated the probabilities of hospitalizations that we see in Table IV.

Sex	Age Group	Hospitalization (With ICU)	Hospitalization (Without ICU)
Female	00-39	0.25%	3.20%
	40-69	1.31%	6.22%
	70+	3.02%	34.11%
Male	00-39	0.53%	1.38%
	40-69	3.59%	10.57%
	70+	4.78%	47.68%

TABLE IV: PROBABILITIES OF AN INFECTED INDIVIDUAL NEEDING HOSPITAL CARE USING THE DATA OF A PORTUGUESE HOSPITAL

Comparing the two tables, we can see that we would be overestimating older people in need of ICU treatments and underestimating almost all other probabilities in our approach. Nevertheless, the absolute values of hospitalizations were not significantly different in both approaches.

From now on we will use the probabilities from Table IV.

4 RISK ASSESSMENTS IN CASE OF COMPLETE COVERAGE OF THE DISEASE

The purpose of this chapter is to estimate the annual cost and the aggregate risk premium of the current Pandemic in the Multicare portfolio, in the first year of the Pandemic. The data used for this chapter was the results from the previous chapter and the data provided by the company that contains the personal information about the policyholders and the amount paid for IPEs, PCR Tests and hospitalizations in the period between March 2020 and February 2021.

It was necessary to estimate infections and hospitalizations in the portfolio. Concerning infections, we had no data on the number of infected policyholders. It happens because we do not know if a client is infected unless he has an internment authorization application associated with a COVID-19 diagnosis. As mentioned, most COVID-19 patients do not require hospitalization and, even within those who required and were insured by Multicare, resorted to the public service. For this reason, the number of hospitalizations we observed was much lower than what could have happened.

4.1 Infections and hospitalizations

It was considered that all the policyholders are susceptible to testing positive for COVID-19. Therefore using the incidences calculated for Portugal and the probabilities of getting the most severe disease, we estimated the cases and hospitalizations per month.

In the first year of the Pandemic, we had I estimated cases of COVID-19 in the portfolio (Figure 3), which means that the incidence was 8.88%. It makes sense that it is higher than the incidence for Portugal because most policyholders live in urban centers and have between 20 to 49 years.

We had H estimated hospitalizations in the portfolio (Figure 4): $H^{w/o\ ICU}$ of which do not need any treatment at ICU while $H^{w/\ ICU}$ were at ICU. This means that 6.49% of cases in the portfolio needed to be hospitalized. This number is much lower than the hospitalization rate in Portugal because the portfolio population is much younger than the Portuguese population.

4.2 Bed capacity

Up to this point, there was no limit on the number of hospitalizations paid by the company. However, it is known that possibly the percentage of beds allocated to Multicare in the private sector, or even the entire private sector, could not respond to all patients in

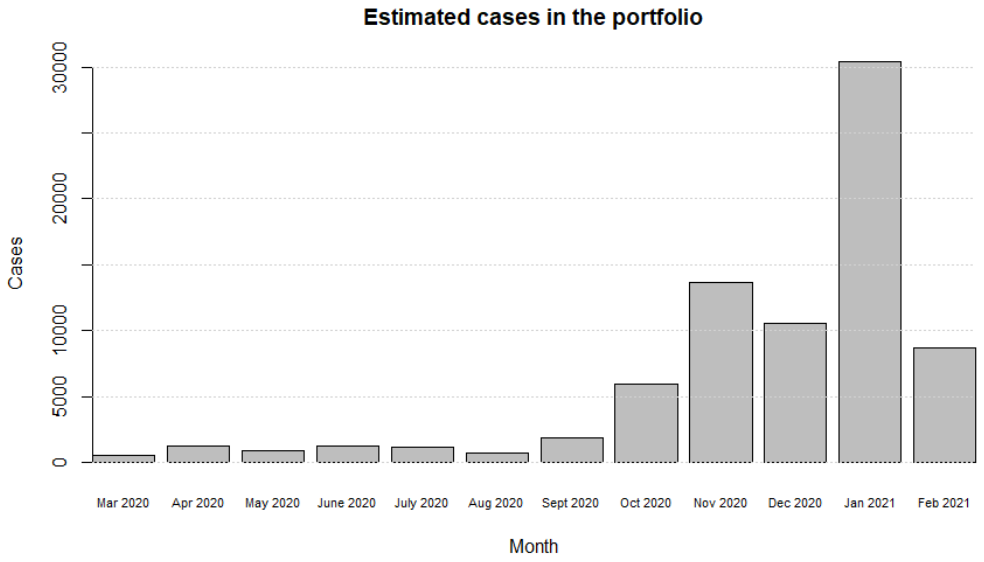

FIGURE 3: Estimated cases in the portfolio per month

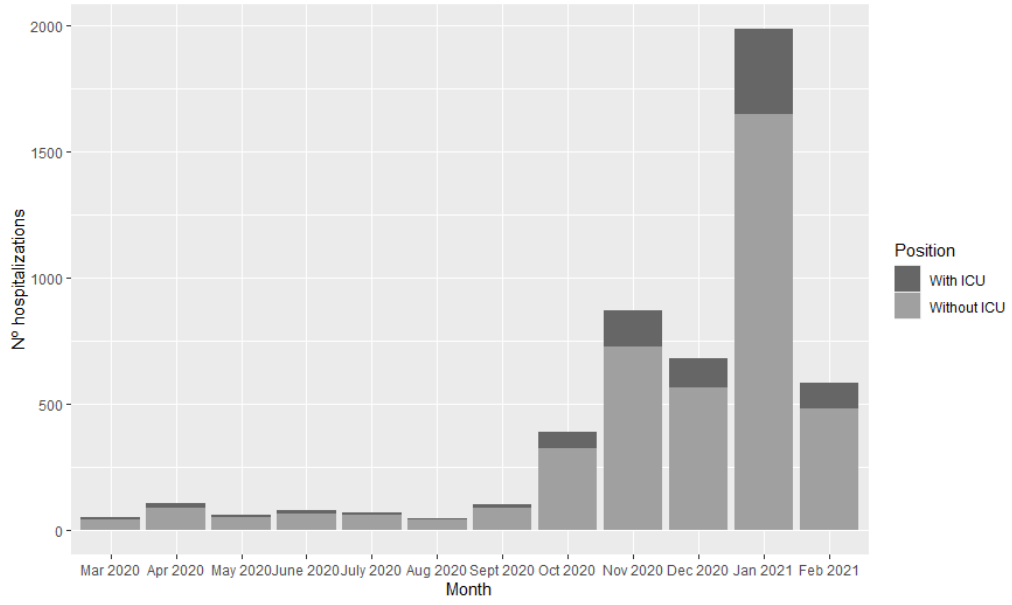

FIGURE 4: Estimated hospitalizations in the portfolio per month

need of hospital admission. So the costs of hospitalizations would be overestimated.

In order to know Multicare's responsiveness when its customers are admitted to the private sector, it was necessary to know the percentage of beds allocated to Multicare and the occupancy rate of these over the days. If a patient arrives at the hospital and there are no beds available, it was considered that he will be treated in another hospital, possibly public, without the company's share.

So if the number of customers in need of hospitalization due to COVID-19 was similar to this reference scenario, it is very likely that Multicare would not respond to all of these patients. Therefore, the costs related to hospitalizations would be lower than those admitting that the insurance company pays all the H hospitalizations.

This subsection will study the maximum amount that the insurer would pay for COVID-19 hospitalizations. According to some assumptions, a simulation was made, using the distributions of the Length of Stay (LoS) at the hospital and the number of beds occupied at each moment. When the beds allocated to the insurer are all occupied, it is considered that the insurer does not bear the costs of the next patients.

In 2019 the private sector had 11 812 beds and 212 ICU beds ("Camas dos hospitais", 2020). It was considered that a maximum of 30% of those beds are allocated to Multicare.

Length of Stay distributions

To model the distribution of hospitalization LoS, the sample of 2311 patients, who were admitted with COVID-19 in a Portuguese hospital, was used.

With this information, it was possible to create three groups of patients:

- Those who just stay at ICU
- Those who do not stay at ICU
- Those who also need ICU treatments

Of these three patient groups, four samples for LoS were created:

- LoS at ICU for patients who just stay at ICU
- LoS at the hospital for patients who do not need ICU treatments
- LoS at the hospital excluding ICU for patients who also need ICU treatments
- LoS at ICU for patients who do not just stay at ICU

We tested, with the Wilcoxon-Mann-Whitney test, if the LoS at ICU (LoS at ICU for patients who just stay at ICU and LoS at ICU for patients who do not just stay at ICU) and LoS at hospital excluding ICU (LoS at the hospital for patients who do not need ICU treatments and LoS at the hospital excluding ICU for patients who also need ICU treatments) can be grouped into two samples, i.e., H_{0_1}: All the LoS at ICU derive from the same population and H_{0_2}: All the LoS at sick bay derive from the same population.

There are two independent random samples, one corresponding to LoS at ICU for patients who just stay at ICU (from X_2), and the other corresponds to LoS at ICU for patients who do not just stay at ICU (from X_1).

The test statistic, Z_T, is 5.4583. As this is a bilateral test, H_0 is rejected when $|Z_T|>Z_{\alpha/2}$. It can be concluded, by observing Table I, that we reject H_{0_1} with a significance level of 1%,5%, and 10%. The two samples could not be aggregated.

The same test was used to check whether to join the LoS at hospital excluding ICU. The test statistic is, Z_T, is -216.84. As $|Z_T|>Z_{\alpha/2}$, H_{0_2} is rejected, and it can be concluded the same as for the ICU LoS.

As said in López-Cheda et al., 2021; Rees et al., 2020, LoS distributions are positively skewed because only a few patients have long LoS. Therefore, the distributions are often adjusted to a gamma, log-normal, or Weibull family of distributions.

Using the R package MASS, we adjusted the distributions to each LoS sample described above. For each sample, a Weibull, Gamma, Log-Normal, and Exponential distribution ware adjusted so that it was possible to conclude which distribution best suited. The Erlang distribution, which is a gamma with integer α parameter, has great applicability, mainly due to its relationship with exponential and gamma distributions. In some cases, when a gamma fitted, we adjusted an Erlang distribution. In order to conclude which distribution is more adequate for each LoS we used four adjustment tests: Chi-Squared (CS), Kolmogorov-Smirnov (K-S), Anderson-Darling (AD) and Cramer Von Mises (CVM) tests. The CS test was only used when the tests performed did not allow us to make a final decision about LoS distributions.

		KS Test	AD Test	CVM Test	CS Test
LOS at ICU (ICU patients)	Weibull (shape=1.314058, scale=18.119103)	0.39522	0.412042	0.72593	0.095891511
	Log-Normal	0.039214	0.556785	0.267172	
	Gamma	0.249917	0.901505	0.533042	0.000623007
	Exponential	0.038427	0.624578	0.428928	
LOS for patients who don't need ICU	Weibull	3.45E-08	0.523184	0.205147	9.93875E-91
	Log-Normal (meanlog=2.156640, sdlog=1.035127)	0.001681	0.947863	0.913177	0.026832011
	Gamma	2.6E-09	0.340087	0.034399	2.89023E-24
	Exponential	1.1E-07	0.508873	0.538139	1.03388E-21
LOS at Hospital excluding ICU for patients who need ICU	Weibull (shape=1.185651, scale=16.291779)	0.093977	0.893323	0.890724	0.734737884
	Log-Normal	0.012735	0.335595	0.006342	
	Gamma (shape=1.42298095, rate=0.09290733)	0.165414	0.651666	0.152991	0.997979386
	Erlang	0.015355	0.007041	0.657118	
	Exponential	0.000349	0.49628	0.910375	
LOS at ICU for patients who were not only in the ICU	Weibull (shape=1.023288, scale=7.908105)	0.330425	0.616656	0.568664	0.012988362
	Log-Normal	0.074131	0.266312	0.144593	0.000741101
	Gamma (shape=1.0929632, rate=0.1396173)	0.309	0.377838	0.802735	0.025227549
	Erlang	3.05E-06	0.000909	0.104788	
	Exponential	0.208647	0.639239	0.252509	0.01032459

TABLE V: P-VALUES OF KS, AD, CVM AND CS TESTS

Table V shows the p-values of the four tests mentioned above for each of the time-adjusted distributions. In order to decide which distribution best fits with the four LoS, we used the following criteria:

- The distribution must not be rejected in any of the tests;
- Observing the histograms and curves.

Thus, it is possible to state, with a significance level of 5%, that the LoS at ICU for ICU patients follows a Weibull distribution with shape and scale parameters equal to 1.314058 and 18.119103, respectively. LoS for patients who do not need ICU follow a Log-Normal distribution with a mean equal to 2.156640 and standard deviation of 1.035127, with 95% confidence.

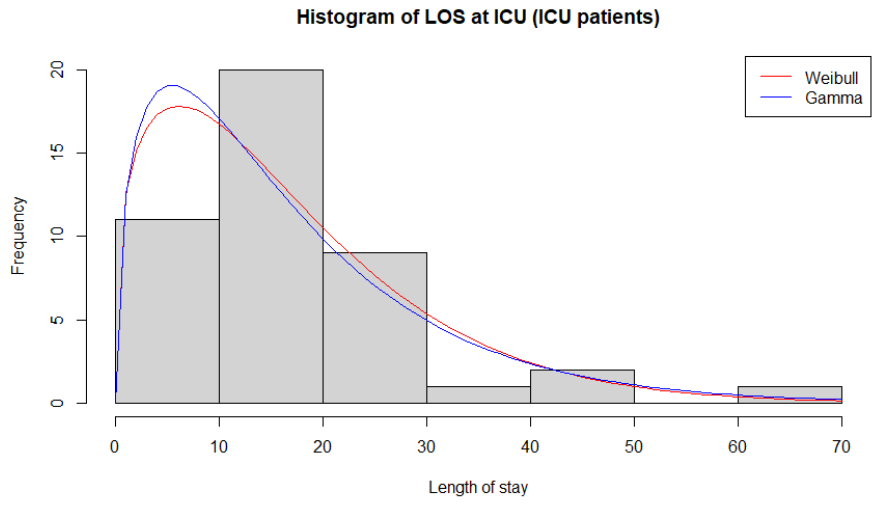

FIGURE 5: Histogram of LoS at ICU for ICU patients

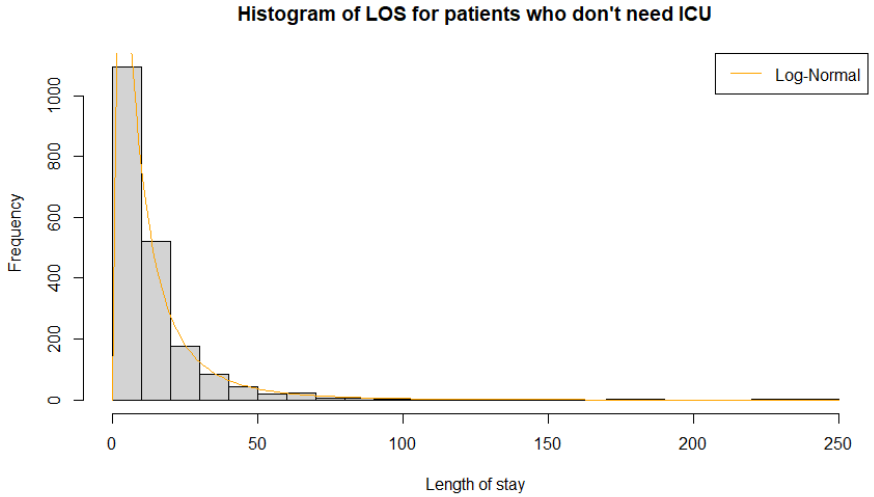

FIGURE 6: Histogram of LoS for patients who do not need ICU

In the last two times (LoS at hospital excluding ICU for patients who also need ICU treatments and LoS at ICU for patients who do not just stay at ICU), the best adjustment would not be so obvious. Figures 7 and 8 show the histograms and the distribution curve of two possible distributions for each sample of LoS. Although with a significance level of 5%, it is possible to state that the LoS at hospital excluding ICU for patients who also need ICU treatments follows a Weibull and an Gamma distributions. By observing the histograms and curves, we can conclude that the Weibull distribution fits better to the data. Similarly, we can assume that LoS at ICU for patients who do not just stay at ICU also follows a Weibull distribution.

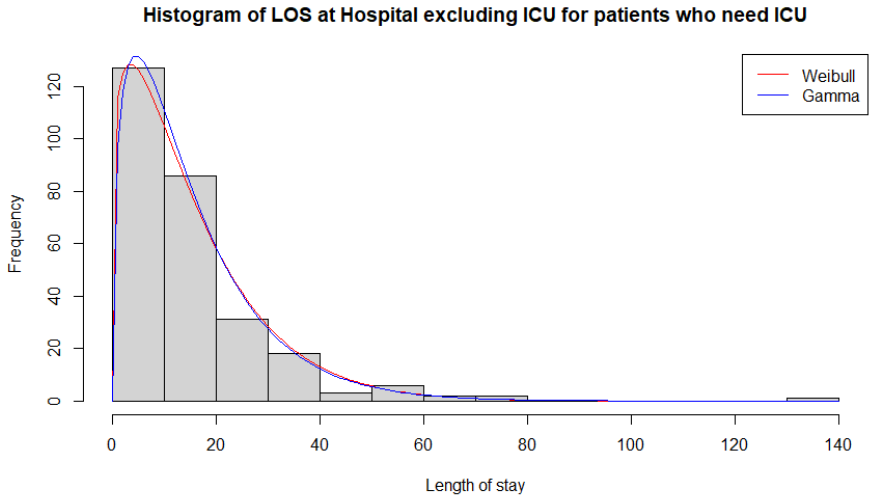

FIGURE 7: Histogram of LoS at Hospital excluding ICU for patients who need ICU

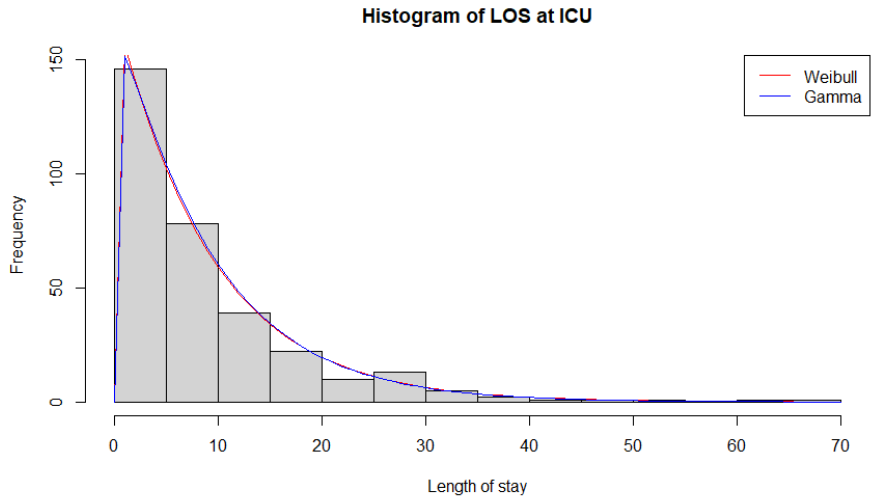

FIGURE 8: Histogram of LoS at ICU

Simulation

In this project, we made a simulation process to determine the cost of policyholders who tested positive to COVID-19, hospitalizations to which the insurer could respond considering the limit of the number of beds available for it.

The number of hospitalizations is known and it is possible to know the admission day for each patient, how many of them were in ICU and the LoS, by the adjusted distribution. In order to know how many beds are available each day, it is necessary to simulate the LoS for each patient using the distributions adjusted above.

For each patient, 1000 pseudo uniforms, u_i, were generated, and depending on the type of hospitalization, the following variables were obtained and organized as it can be seen in Table V:

- patients who need ICU:
  - 1000 pseudo Weibull, for LoS at ICU:

$$X_i = [-7.908105 \cdot \ln(1 - u_i)]^{1/1.023288} \tag{31}$$

  - 1000 pseudo Weibull, for LoS at hospital excluding ICU:

$$X_i = [-16.29178 \cdot \ln(1 - u_i)]^{1/1.185651} \tag{32}$$

- patients who do not need ICU:

– 1000 pseudo Log-Normal, for the LoS at hospital:

$$X_i = exp(1.035127 \cdot \Phi^{-1}(u_i) + 2.15664) \tag{33}$$

Policyholder Code	Entry day	ICU	LoS at ICU	LoS at hospital (w/o ICU)
...	...	...	...	...
10A	10	No	-	7
11B	11	No	-	10
...	...	...	...	...
23S	23	Yes	14	10

TABLE VI: MATRIX OF GENERATED LENGTH OF STAY FOR EACH POLICYHOLDER

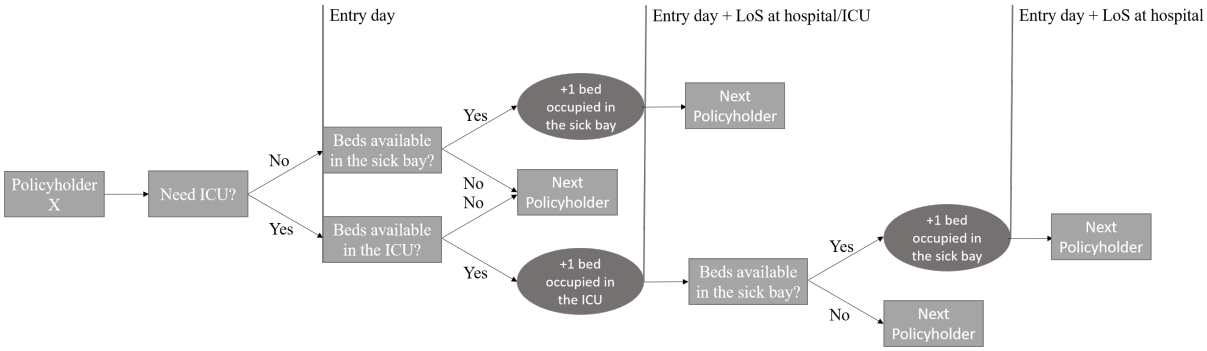

FIGURE 9: Illustrative scheme of the way the beds were occupied

Figure 9 illustrates the process used to occupy the available beds and count how many policyholders in need of hospitalization will be covered by the insurer. In the first phase, it is verified whether the patient needs ICU or not. In both cases, we see if there are available beds. If there is, the patient occupies +1 bed until the day of his discharge; otherwise, go home. In case of need for ICU treatment, the patient needs to be transferred to the sickbay. At this stage, we will see again if there are beds available. If there is not, the patient, go home. It should be noted that the patient goes home only at this stage and not at the beginning of the process.

From the observation of Figure 10 we can see that only in January 2021 would the insurer not cover the hospital expenses of all its customers due to the number of beds allocated to it. In this scenario, patients without the need for ICU treatments, i.e., those in the sickbay, have always responded. However, in those who need to go through the ICU, the capacity is more limited.

After the simulation was completed, the bootstrap percentile confidence intervals for the number and the costs of hospitalizations were done.

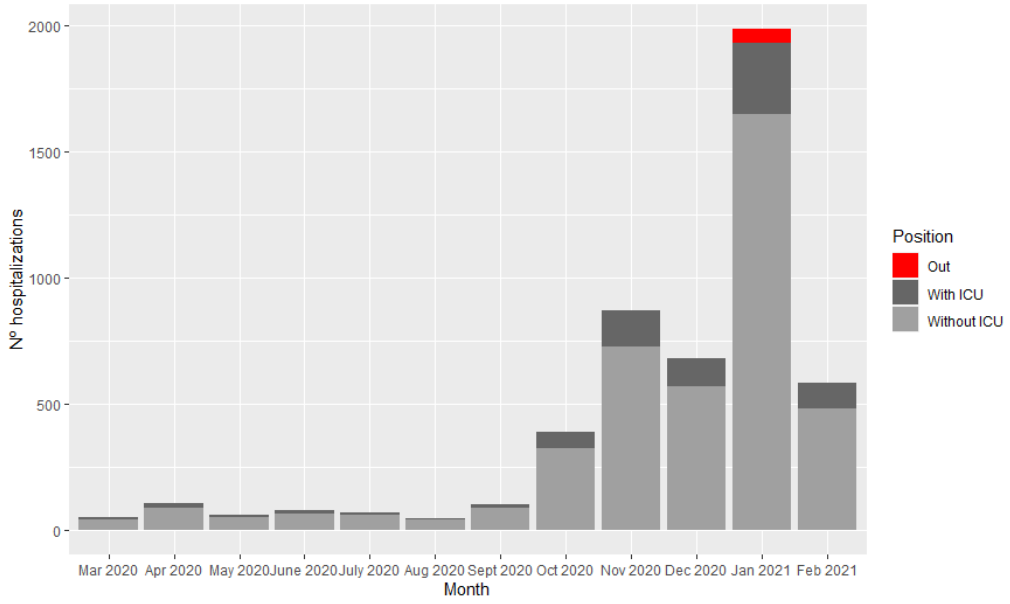

FIGURE 10: Estimated hospitalizations in the portfolio per month considering the number of beds available

4.3 Total estimated cost

The total annual cost was calculated considering the PCR tests performed, the IPEs, and the estimated hospitalizations. Since the beginning of the Pandemic, Multicare has paid for its customers' PCR tests, IPEs and hospitalizations. However, despite paying hospital admission expenses, the hospitalizations observed during this period were much lower than they could have occurred for various reasons, such as the fact that some private hospitals did not accept COVID-19 patients. For that reason, the estimated hospitalizations were used in the calculus of the annual total cost. The PCR tests and the IPEs costs are real.

From now on the following notation was used: Nº of estimated infections in the reference scenario (I); Nº of estimated hospitalizations in the reference scenario (H); Nº of estimated hospitalizations w/o ICU in the reference scenario ($H^{w/o\ ICU}$); Nº of estimated hospitalizations with ICU in the reference scenario ($H^{w\ ICU}$); Estimated total cost in the reference scenario (C); Estimated aggregated risk premium in the reference scenario (P).

Therefore, considering this reference scenario, we obtained €C and €P.

5 STRESS TESTS

In this chapter, some stress tests were taken to see the impact on the portfolio if the pandemic was not exactly as it occurred in Portugal. At first, we though that it would be interesting to see the situation of other countries. A survey was conducted to see which countries had more cases, per million people, than Portugal. Second, some assumptions were changed, making two extreme scenarios. To conclude, rehearsals with the worst characteristics of other pandemics were taken.

To calculate the total cost of the following scenarios we consider that the number of PCR tests is directly proportional to the estimated cases. A simulation procedure was also made for each of these scenarios in order to limit, from hospitalizations, those that the insurer pays.

5.1 Countries with more COVID-19 cases

At this point, we ask the following question: *"What would be the impact on the portfolio if we were in another country?"*.

In order to answer this question, it was necessary to research about COVID-19 situation in other European countries. The last two months of the first year of the pandemic (January and February 2021) were so bad for Portugal that few European countries were worse than Portugal. According to the WHO dashboard "WHO Coronavirus (COVID-19) Dashboard", 2019 and the population "Population Pyramids of the World from 1950 to 2100", 2019 of each country, it can be seen, at Table VII, that Czechia and Slovenia had more cases per million in the first year of the pandemic.

Country	Cases at Feb. 2021	Population	Cases per 1M
Czechia	1235480	10 708 982	115369
Slovenia	189627	2 078 932	91214
Portugal	803844	10 196 707	78834
Spain	3132233	46 754 783	66993
United Kingdom	4170523	67 530 161	61758
Italy	3671208	60 461 828	60719
France	2907825	65 273 512	44548

TABLE VII: PROBABILITIES OF AN INFECTED INDIVIDUAL NEEDING HOSPITAL CARE

Czechia had 91,214 cases per million, and Slovenia had 78,834 cases also per million. At the same time, Portugal had 77,834 cases. Then, using the information of cases by sex and age of each of these countries and the population distribution, it was possible to calcu-

late the incidence rate and apply it to the portfolio. After that, cases and hospitalizations in the portfolio were estimated.

If Multicare were in Czechia, there would be 1.47I cases and more than 1.60H hospitalizations. Using Slovenia data, the scenario was not so bad, but worst than Portugal. There were 1.18I cases and 1.31H hospitalizations. In order to see the impact on the first year of the pandemic costs, we estimated the total cost. As more hospitalizations mean more costs, the total estimated cost would be €1.49C for Czechia and €1.26C for Slovenia.

5.2 Using other metrics

At the beginning of this study, it was said that 12% of infected individuals were hospitalized. In the beginning of the pandemic, it was considered that 15% of patients were hospitalized without ICU and 5% with ICU. These metrics are displayed on the DGS website since the beginning of the pandemic ("Perguntas Frequentes Categoria - COVID-19", 2019). For this case, there would have been 1.67H hospitalizations (10.82%), and the total cost would be €1.36C.

The portfolio tendency is to converge for the Portuguese population. The distribution of clients per region was maintained, i.e., for the number of clients in each region, the distribution per age and sex of each of these Portuguese regions was applied. In this scenario, the incidence rate is also higher than the Portuguese rate because 70% of the portfolio lives in urban centres where the incidence rate is higher. There would be 0.96I infected people, which is 8.55% of the portfolio, and 11.87% of those would be hospitalized. This means that the estimated total cost would increase to €1.45C because the number of hospitalizations increase.

5.3 Other pandemics and epidemics

For the WHO, an epidemic is *"the occurrence in a community or region of cases of an illness, specific health-related behaviour, or other health-related events clearly in excess of normal expectancy. The community or region and the period in which the cases occur are specified precisely. The number of cases indicating the presence of an epidemic varies according to the agent, size, and type of population exposed, previous experience or lack of exposure to the disease, and time and place of occurrence"* ("WHO | Definitions", 2019). Therefore a pandemic is an epidemic that spread around the world.

The history of epidemics and pandemics is extensive. The first pandemic we know occurred between 430 B.C. and 427 B.C, the Plague of Athens, which killed two-thirds

of the Athens population. In 1347, began, in Central Asia, The Black Death plague, the largest pandemic in history until 1918 with the Spanish Flu. The geographic origin of this Spanish Flu is unknown. During the first world war, as Spain was a neutral country, it was the first country to report cases. This pandemic infected 500 million people and killed almost 5% of the world population. In Portugal, there were estimated between 50 to 70 thousand deaths. In 1957, the Asian Flu emerged in Northern China and extended to almost every country in 10 months, killing 1.1 million people. In 1968 appeared in Hong Kong a flu called The Hong-Kong Flu, which killed one million people, 500 thousand just in Hong Kong. At the end of the first quarter of 2009 came the Swine Flu, later called Influenza A Pandemic ("As Epidemias e as Pandemias na História da Humanidade", 2020). There were 166 922 infected in Portugal and 122 people died ("Questions and answers on the pandemic (H1N1) 2009", 2009; "Spanish flu", 2021).

Because of the extensive history of epidemics and pandemics, it would be relevant to study deeply some of the most recently epidemics and pandemics such as Measles, Severe Acute respiratory syndrome (SARS)-CoV and Middle East respiratory syndrome (MERS)-CoV. Next, we describe briefly each of these epi-/pan-demics and see their impact.

Measles

Measles is a viral infection characterized by high fever, cough, conjunctivitis, runny nose, and red spots in the skin. The virus infects the respiratory system and spreads throughout the body. The symptoms begin about 10 to 12 days after the infection and persist 4 to 7 days. It is highly contagious and can be transmitted from 4 days prior to 4 days after the onset of the symptoms. Serious complications occur more frequently in children under the age of 5. Until 1963, before the vaccine, measles caused 2.6 million deaths every year. It was one of the main causes of infant mortality ("Measles", 2019; "Sarampo", 2021). One-third of infected people had serious complications as acute encephalitis and pneumonia. The transmission rate is 23.9%, 10% of the infected individuals died, and 67.5% recovered from the disease. In 2019, 364 811 cases of measles were reported - the highest number in the last two decades. The reproduction number, R_0, is higher than 3.7 ("Modeling of measles epidemic with optimized fractional order under Caputo differential operator | Elsevier Enhanced Reader", 2021).

SARS-CoV

The Epidemic SARS appeared in November of 2002 in China and it caused severe acute respiratory syndrome. There were more than 8000 infected people, only 135 were

children, and 774 of those died, resulting in a mortality rate of 9.5%. This virus occurred more frequently in adults between 25 and 70 years, and there were a few cases in 15 years old children. The latent period of the disease (period since the individual's infection until he could infect others) is between 2 to 7 days (Aleebrahim-Dehkordi et al., 2021). The symptoms appear 2 to 12 days after the infection. The disease killed primarily elderly people, but it was more severe in children with 12 years or less (Bradley and Bryan, 2019). The reproduction number $R(t)$ of SARS is 1.7-1.9 (Petrosillo et al., 2020).

MERS-CoV

MERS-CoV, Middle East Respiratory Syndrome Coronavirus, was a new virus that caused respiratory diseases. It appeared on the 3rd of September 2012 in Saudi Arabia. There were more than 2400 registered cases, 2% children, most were between 40 and 50 years old, and 65% were men. Death from MERS is higher in men and patients with underlying diseases (Aleebrahim-Dehkordi et al., 2021; "MERS-CoV", 2019). The patients had symptoms 2 to 14 days after being infected [21]. 25% of the cases were asymptomatic. The mortality rate was 35.5%, which means at least 852 deaths by this virus. The incubation period was 2 to 14 days [19]. 70% of patients needed hospitalization for mechanical ventilation [21]. The average time from the beginning of symptoms to hospitalization is 4 days, and the average time until admission to the UCI is 1 day. The patient stays in the UCI for 30 days, on average. Death occurs 12 days after the beginning of symptoms (Bradley and Bryan, 2019). The R_0 is less than 1 (Petrosillo et al., 2020).

	Measles	SARS	MERS	COVID-19
R_0	>3.7	1.7-1.9	<1	2.0-2.5
Age with higher incidence	<5	25-70	40-50	20-29
Woman (%)		64-68	35	40-55
Mortality Rate	10%	9.50%	35.50%	2.11%
Hospitalization Rate	33.33% (*)	30%	70%	20%
Length of stay at ICU			30	20
Incubation Period (days)	7-14	2-12	2-14	1-14

TABLE VIII: SUMMARY TABLE OF THE CHARACTERISTICS OF OTHER EPIDEMICS

The Table VIII summarises up all the information that was manageable to collect. Looking at the mortality rate and the basic reproduction rate, it is possible to observe that, despite the similarity between COVID-19, SARS and MERS, COVID-19 shows a significantly lower mortality rate and spreads much more easily. The MERS pandemic had a higher mortality rate than COVID-19 but did not spread very much. This is probably because 70% of the cases were hospitalized, and therefore it was easier to control the disease.

The worst characteristics of each of the outbreaks were taken, and more trials were performed to estimate the cost.

The most impactful characteristics in the case of Measles are R_0, which is higher than COVID-19, and the hospitalization rate. It was considered that all patients who develop pneumonia are hospitalized. Carrying the rate of hospitalizatons would be 33%. It would be necessary to use an epidemiological model to see the evolution of a pandemic with this R_0. However, epidemiological models did not fit the objective of this project. In this case, it was considered that there would be twice the cases of COVID-19. Therefore, a hospitalization rate of 33% and twice the number of infected was used in this scenario. There would be 2I cases and 7.02H hospitalizations. The estimated annual cost would be €3.46C.

In the scenario for the SARS-CoV epidemic, a higher hospitalization rate was considered in the younger age groups, and it was also assumed that the overall hospitalization rate would be 30%. In this case, the annual estimated cost would be €2.15C.

Based on the MERS-CoV epidemic, the last scenario, it was considered a 70% hospitalization rate, instead of the 14%. The mean length of stay (LoS) at the hospital was 3/2 of the average hospitalization time of COVID-19. Therefore, the impact of the increase in the average hospitalization time was considered proportional to the cost. In this case, the annual estimated cost would be €3.87C.

Finally, a scenario was made that contemplated the most extreme one. Therefore, it was assumed that there would be twice as many infected patients, 70% of those would be hospitalized, and that the time of hospitalization increased. Consequently, in this scenario, the estimated total cost would be €5.07C.

We consider that T is the cost of all insurer claims in a year without a pandemic. In Table IX we can see the relationship between the total cost of one year of pandemic and T. In the reference scenario, we can observe that the treatment of the pathology would amount to 14% of all costs with claims, that is, it would be an increase of 14%. Moreover, in the worst case scenario, C is higher than T. We are saying that claims costs will be more than two times the cost of all insurer claims in a year without a pandemic, which can risk the solvency of the insurance company.

	Infected	Hosp$^{w/o\ ICU}$	Hosp$^{w/\ ICU}$	Total Cost(€)	C/T
Multicare	I	H$^{w/o\ ICU}$	H$^{w/\ ICU}$	C	0.14
With the incidences of the Czech Republic	1.47I	1.61H$^{w/o\ ICU}$	1.58H$^{w\ ICU}$	1.49C	0.21
With the incidences of Slovenia	1.18I	1.31H$^{w/o\ ICU}$	1.30H$^{w/\ ICU}$	1.26C	0.18
Higher hospitalization rate (15% w/o ICU + 5% w/ICU)	1.00I	1.67H$^{w/o\ ICU}$	1.67H$^{w/\ ICU}$	1.36C	0.19
Assuming portfolio as the same distribution as Portugal	0.96I	1.59H$^{w/o\ ICU}$	2.61H$^{w/\ ICU}$	1.45C	0.20
Sarampo	2.00I	7.32H$^{w/o\ ICU}$	5.55H$^{w/\ ICU}$	3.46C	0.49
SARS	1.00I	3.69H$^{w/o\ ICU}$	2.50H$^{w/\ ICU}$	2.15C	0.30
MERS	1.00I	11.30H$^{w/o\ ICU}$	5.83H$^{w/\ ICU}$	3.87C	0.55
worst case scenario	2.00I	22.60H$^{w/o\ ICU}$	11.65H$^{w/\ ICU}$	5,07C	0,72

TABLE IX: SUMMARY TABLE OF THE CASE SCENARIOS

6 LOSS AND COMBINED RATIOS

Loss Ratio

The loss ratio represents the relationship between losses, which include paid insurance claims and adjustments expenses, and total collected insurance premiums. Thus, the loss ratio is the quotient between insurance claims paid plus adjustment expenses and the total earned premiums (Equation 34).

Loss ratios vary depending on the type of insurance. For example, the loss ratio for health insurance is higher than the same ratio for car and property insurance because it is much more used. Loss ratios help assess the health and profitability of an insurance company. The lower the ratio, the greater the profitability of the insurance company. For example, if we observed high loss ratios, an insurance company cannot produce enough revenues or income, making it impossible to meet or pay its financial obligations.

$$\text{Loss Ratio} = \frac{\text{Incurred Losses} + \text{Adjustment expenses}}{\text{Earned Premium}} \tag{34}$$

Combined Ratio

The combined ratio also measures profitability an insurance company uses to scale how well it performs in its operations. The combined ratio is the quotient of the sum of incurred losses and expenses and the earned premium (Equation 35).

$$\text{Combined Ratio} = \frac{\text{Incurred Losses} + \text{Expenses}}{\text{Earned Premium}} \tag{35}$$

Insurance companies want a ratio below 100% because it indicates that the company makes an underwriting profit. In comparison, a ratio higher than 100% means that the insurance company is paying out more money in claims than receiving from premiums.

The combined ratio can be written as the sum of the Loss Ratio and the Expenses Ratio.

In 2019 Multicares Loss Ratio was 79.1%, and the Combined Ratio was 96.7%. In the reference scenario, we observed that the pandemic related costs would increase 14% of the annual costs of all claims. It means that the numerator increases by 14% and the denominator, the premiums, remain the same. In Table X we can see that the Loss Ratio could vary from 90% to 136% depending on the circumstances. Therefore it is not difficult for the Combined Ratio to reach 100%, and consequently, the insurer ceases to

have profitability, even in the reference scenario.

	C/T	Loss Ratio (%)
Multicare	0.14	90.26
With the incidences of the Czech Republic	0.21	95.74
With the incidences of Slovenia	0.18	93.17
Higher hospitalization rate (15% w/o ICU + 5% w/ICU)	0.19	94.33
Assuming portfolio as the same distribution as Portugal	0.20	95.25
Sarampo	0.49	117.71
SARS	0.30	103.13
MERS	0.55	122.33
worst case scenario	0.72	136.05

TABLE X: LOSS RATIO FOR ALL CASE SCENARIOS

7 CONCLUSION

This project was carried during changing times due to the COVID-19 pandemic crisis. The current pandemic has brought new challenges for the entire health sector, including health insurance companies that did not cover expenses related to such pandemic crisis. Given the lack of data from previous pandemics and epidemics, and therefore, the complexity in charging pandemic-related products, the risk of these events happening was not well understood. The current pandemic has made it more relevant to discuss the possible coverage of these products. In Portugal, Multicare and other insurance companies worldwide have also come forward with products to meet the needs of their policyholders. These products ranged from access to vaccination, online consultations, tests, or even a lump sum benefit if hospitalization was required. This research aimed to estimate the total cost for health insurance during a pandemic crisis to remove the clause that safeguards insurers from this cost. COVID-19 brought us many data with which we could estimate this cost. Therefore, considering the incidences, the hospitalization rate, the limit of available beds, and using a simulation process, we estimated the cost of one year of the pandemic. The medical treatment costs in a pandemic represent 14% of all insurer claims in a year without a pandemic. Nothing guarantees us that the next pandemic will be precisely the same as the current one, so when we perform the stress tests, we conclude that in the worst case scenario, the cost related to a pandemic is 0.72 times the cost of all other insurance claims. Therefore it is very likely that the Loss and Combined Ratios are higher than 100%, and the insurer does not have profitability, which can risk the solvency of the insurance company. The steps to be taken include calculating a premium or an additional premium if it is taken into another existing product. The main challenge of this calculation is to consider that pandemics do not occur every year. However, on the other hand, we do not know the time until the next pandemic, which we will also consider as a variable.

REFERENCES

Abd-Elfattah, A. M. (2011). Goodness of fit test for the generalized Rayleigh distribution with unknown parameters [Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/00949650903348155]. *Journal of Statistical Computation and Simulation*, *81*(3), 357–366. https://doi.org/10.1080/00949650903348155

Aleebrahim-Dehkordi, E., Soveyzi, F., Deravi, N., Rabbani, Z., Saghazadeh, A., & Rezaei, N. (2021). Human Coronaviruses SARS-CoV, MERS-CoV, and SARS-CoV-2 in Children [Publisher: Elsevier]. *Journal of Pediatric Nursing: Nursing Care of Children and Families*, *56*, 70–79. https://doi.org/10.1016/j.pedn.2020.10.020

As Epidemias e as Pandemias na História da Humanidade. (2020). Retrieved April 22, 2021, from https://www.medicina.ulisboa.pt/newsfmul-artigo/99/epidemias-e-pandemias-na-historia-da-humanidade

Bradley, B. T., & Bryan, A. (2019). Emerging respiratory infections: The infectious disease pathology of SARS, MERS, pandemic influenza, and Legionella. *Seminars in Diagnostic Pathology*, *36*(3), 152–159. https://doi.org/10.1053/j.semdp.2019.04.006

Camas dos hospitais. (2020). Retrieved June 22, 2021, from https://www.observatorio-das-desigualdades.com/2020/04/03/camas-dos-hospitais/

Coronavirus FAQs. (2020). Retrieved July 15, 2021, from https://www.vitality.co.uk/coronavirus-faqs/

Gebhard, C., Regitz-Zagrosek, V., Neuhauser, H. K., Morgan, R., & Klein, S. L. (2020). Impact of sex and gender on COVID-19 outcomes in Europe. *Biology of Sex Differences*, *11*(1), 29. https://doi.org/10.1186/s13293-020-00304-9

IRDAI Welcomes You. (2020). Retrieved July 13, 2021, from https://www.irdai.gov.in/Defaulthome.aspx?Page=H1

Lemos-Paião, A. P., Silva, C. J., & Torres, D. F. M. (2020). A New Compartmental Epidemiological Model for COVID-19 with a Case Study of Portugal. *Ecological Complexity*, *44*, 100885. https://doi.org/10.1016/j.ecocom.2020.100885

López-Cheda, A., Jácome, M.-A., Cao, R., & Salazar, P. M. D. (2021). Estimating lengths-of-stay of hospitalised COVID-19 patients using a non-parametric model: A case study in Galicia (Spain) [Publisher: Cambridge University Press]. *Epidemiology & Infection*, *149*. https://doi.org/10.1017/S0950268821000959

Lusa. (2021). Covid-19: Vítimas mortais no domicílio quase triplicam em Janeiro. Retrieved May 10, 2021, from https://www.publico.pt/2021/02/08/sociedade/noticia/covid19-vitimas-mortais-domicilio-quase-triplicam-janeiro-1949835

Measles. (2019). Retrieved May 3, 2021, from https://www.who.int/news-room/fact-sheets/detail/measles

MERS-CoV. (2019). Retrieved April 22, 2021, from https://pandemic.internationalsos.com/overview/mers-overview

Modeling of measles epidemic with optimized fractional order under Caputo differential operator | Elsevier Enhanced Reader. (2021). https://doi.org/10.1016/j.chaos.2021.110766

Neves, S. (2021). Covid-19. Um terço dos internados em Portugal está entre os 40 e os 69 anos. Retrieved May 10, 2021, from https://www.publico.pt/2020/11/18/sociedade/noticia/covid19-terco-internados-portugal-estao-40-69-anos-1939711

New Pandemic Insurance solution protects consumers against the financial ramifications of Covid-19 hospitalisation – Sirago Underwriting Managers. (2021). Retrieved July 13, 2021, from https://sirago.co.za/new-pandemic-insurance-solution-protects-consumers-against-the-financial-ramifications-of-covid-19

Perguntas Frequentes Categoria - COVID-19. (2019). Retrieved May 16, 2021, from https://covid19.min-saude.pt/category/perguntas-frequentes/

Petrosillo, N., Viceconte, G., Ergonul, O., Ippolito, G., & Petersen, E. (2020). COVID-19, SARS and MERS: Are they closely related? [Publisher: Elsevier]. *Clinical Microbiology and Infection*, *26*(6), 729–734. https://doi.org/10.1016/j.cmi.2020.03.026

Population Pyramids of the World from 1950 to 2100. (2019). Retrieved May 16, 2021, from https://www.populationpyramid.net/%20world/2019/

Questions and answers on the pandemic (H1N1) 2009. (2009). Retrieved May 3, 2021, from https://www.ecdc.europa.eu/en/seasonal-influenza/2009-influenza-h1n1-faq

Rees, E. M., Nightingale, E. S., Jafari, Y., Waterlow, N. R., Clifford, S., B Pearson, C. A., Group, C. W., Jombart, T., Procter, S. R., & Knight, G. M. (2020). COVID-19 length of hospital stay: A systematic review and data synthesis. *BMC medicine*, *18*(1), 270. https://doi.org/10.1186/s12916-020-01726-3

Ross, S. M. (2013). *Simulation* [Google-Books-ID: 1Dwsyyty3P8C]. Academic Press.

Sarampo. (2021). Retrieved May 3, 2021, from https://www.sns24.gov.pt/tema/doencas-infecciosas/sarampo/

Spanish flu [Page Version ID: 1018577093]. (2021). Retrieved April 22, 2021, from https://en.wikipedia.org/w/index.php?title=Spanish_flu&oldid=1018577093

Stigler, S. M. (2012). Studies in the history of probability and statistics, L: Karl Pearson and the Rule of Three. *Biometrika*, *99*(1), 1–14. https://doi.org/10.1093/biomet/asr046

WHO | Definitions: Emergencies [Publisher: World Health Organization]. (2019). Retrieved April 29, 2021, from https://www.who.int/hac/about/definitions/en/

WHO Coronavirus (COVID-19) Dashboard. (2019). Retrieved May 7, 2021, from https://covid19.who.int

WHO Outbreak Benefit. (2020). Retrieved July 15, 2021, from https://www.discovery.co.za/corporate/covid-19-benefits-and-cover