Università degli Studi di Padova
Dipartimento di Scienze Statistiche
Corso di Laurea Magistrale in Scienze Statistiche

The Covid-19 isolation effect on the mental health of older people

Relatore: Prof. Omar Paccagnella
Dipartimento di Scienze Statistiche, Università di Padova
Correlatrice: Prof.ssa Viola Angelini
Faculty of Economics and Business, Rijksuniversiteit Groningen

Laureanda: Teresa Peronio
Matricola: 1242465

A.A.2021/2022

Contents

List of Figures

List of Tables

Introduction

At the beginning of 2020 the entire world stopped because of a virus (SARS-Cov-2) that was spreading among humans. The so-called COVID-19 disease changed and is still changing millions of lives, from a lot of perspectives. First of all, the health of citizens was threatened: people with a weak immune system had a significant chance of dying or being severely damaged by this disease, but it could happen also to those in good health. This fact led to a situation of fear and uncertainty, due also to the lack of information on this new disease.

Governments had to take emergency measures to limit the diffusion of the virus and therefore there were significant consequences both on the economic situation and social life of citizens and countries. Many activities such as restaurants and cafes were in fact closed, while for other workers the 'smart working' from home became mandatory. Travel from one place to another was reduced as much as possible and seeing family, friends or colleagues was allowed only in exceptional cases. Public transport and schools almost everywhere were subjected to periods of closure and people were forced or suggested to stay in their homes. These measures were applied with varying degrees of severity in Europe, in order to protect the health of at-risk groups. Already sick people belong to these segments, but undoubtedly the elderly as well. In Italy, for example, the average age of people who died because of the infection for coronavirus is 80 years old and the median 82[1]. Furthermore, the probabilities of acquiring a serious infection, being hospitalised or going to intensive care unit increase with increasing age.

It might be interesting to approach the pandemic issue from this point of view. How did older people experience the pandemic and what were their reactions to it? What were the effects on them? These are only some of the questions that can be asked.

The main purpose of this work is to understand and illustrate the consequences of the first wave of the COVID-19 pandemic on the health of older people in various countries; in particular the focus will be on mental health in terms of loneliness and depression. Through appropriate statistical methodologies, the situations before and after COVID-19 will be compared, in order to see whether and how much the COVID-19 has worsened the mental health of the subjects.

One goal will also be to find what are the most significant factors that have affected the mental health of seniors and what it depends on. For example, did social media use help the respondents stay in touch with family? Did it have a positive influence?
Finally, we will attempt to outline and uncover potential differences among countries, based on their respective adopted restrictions.
Panel data from SHARE (Survey of Health, Ageing and Retirement in Europe) will be used for these purposes, and in particular those from SHARE-COVID 19. The target population of the SHARE study is people aged 50 years or older and in the last wave the survey was conducted in 28 different countries, all belonging to the European Union, except Switzerland and Israel. The questions focus on the physical and mental health, the economic situation and social networks of these individuals. In the SHARE-COVID 19 survey the situation before and after the pandemic was analysed.
In the first chapter, initially we will briefly explain the context of the phenomenon of COVID-19: its evolution in terms of infections, symptomatic and asymptomatic cases, hospitalisations and deaths throughout Europe.
It will then be interesting to understand the structure of the SHARE survey and the characteristics of the population at which it is directed. The differences between the waves and the methods used by interviewers will be examined. The outbreak of COVID-19 during wave 8 has in fact forced the experts to reorganise the survey taking this change and its effects into account: an in-person interview was no longer considered an option. Therefore, the structure of the new modified questionnaire will be described.
The second chapter will summarise the first inspections of the data, both with the preliminary analysis and choice of variables, and with some descriptive and exploratory statistics. This will help understanding who are the people referred to in the research, including from a socio-demographic perspective.
In order to pursue the objective, Chapter 3 will explain the models used during the analysis, particularly with reference to their theoretical structure, assumptions and advantages.
They will then be estimated in the following chapter: the results and comparison between different models will be reported.
Finally, the main conclusions will be shown, including any problems and issues encountered. Furthermore, in Appendix A the questionnaire from which the variables are taken is reported.

Chapter 1

Case of study and SHARE survey

1.1 COVID-19 spread

At the beginning of the year 2020 a pandemic struck the whole world, changing the way of living of billions of people and leading also to dramatic consequences. In this section we will firstly report some statistics about Covid-19 and its evolution among European countries, then we will focus on the subject of mental health.

At the end of the year 2019, an abnormal pneumonia occurred in the Chinese city of Wuhan and then spread to Hubei province. It appeared to come from a seafood and live animal market and was apparently only transmissible by animals. After some time, it was discovered to be a coronavirus disease and was named by scientists 'Covid-19'. Unfortunately, it was neither an isolated case illness related to the market nor transmissible only by animals: the transmission from person to person was in fact right away confirmed. This led to an incredible increase in infections, and the consequences of the disease had not to be underestimated, as patients often found their lives in danger and the intensive care units of hospitals were crowded. At the beginning of 2020 cases were found not only in Asia, but also in Europe. On March 11th 2020 the WHO (World Health Organization) declared that the COVID-19 epidemic could be considered a pandemic, as it involved all the continents.

The medical evidence suggests that the disease spreads among people in different ways. The WHO states that the virus spreads "from an infected person's mouth or nose in small liquid particles when they cough, sneeze, speak, sing or breathe"[2]. Also touching one's own mouth, eyes and face with hands is considered a source of possible contagion. Furthermore, crowded places help to spread the disease, because aerosols can remain suspended in the air longer. Because of this rapid way of transmission of the virus and because the disease was unknown, it was decided to implement a lockdown in many countries. This action was taken to fight the

spreading of the disease and to limit as far as possible travels, both from a country to another and inside the municipality of residence in general.
In Europe the first country that was hit by Covid-19 was Italy, followed by all the others.
We will show the situation of Covid-19 infection until October 2020, that is the first wave, because in most of the countries considered in this analysis, the SHARE survey was completed by the end of summer, so subsequent developments of the pandemic are not considered in this work.

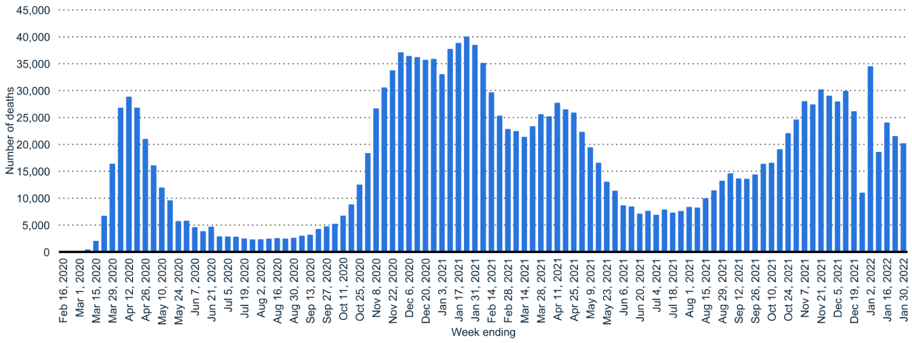
1.1.1 Overview on Covid-19 spread

The evolution of the spread of Covid19 infection is still in progress at the time of writing of this thesis, testifying to the severity of the pandemic phenomenon studied. In these two years many methods have been compared to analyse and describe the phenomenon. The main ones are: the daily or weekly number of newly infected people, the number of hospitalised people, the number of people in intensive care and the number of deaths. These indicators have been related to the number of inhabitants of the country of reference or the number of swabs executed in the period in the most accurate representations. This is not the place to discuss the appropriateness of the various methods, but we limit ourselves to the most obvious observations. We begin by saying that the spread of the virus follows a cyclical pattern, or, as they have been defined, 'in waves'.
Figure 1.1 shows the number of deaths in Europe in the two pandemic years and it highlights the four waves, alternated with periods of great reduction of the phenomenon. The present work is based on data relative to the period of the first wave, at the end of which many scientists believed that the evolution of the virus and its effects was coming to an end. We started with a representation based on the number of deaths because many believe it is the most meaningful and objective statistic, even if it is sometimes questioned because of the different criteria adopted by different countries or even by the same country in different periods to categorise deaths due to Covid-19, differentiating them or not from those occurring in people with a lot of pre-existing pathologies, including Covid.

Another parameter of great interest is the one related to the number of new infections that shows the trend of the pandemic. Even in this case we find several critical issues, which depend both on the size of the overall population, and on the number of swabs (which is the system of detection of the presence of the virus in a single person) carried out.
Observing, for example, the graph in Figure 1.2, which shows the new infections, it can be seen that the first wave seems to present a significantly lower number of sick people than the following ones, unlike the deaths in the previous graph,

Number of new coronavirus (COVID-19) deaths in Europe since February 2020 (as of January 30, 2022), by date of report

Number of new coronavirus (COVID-19) deaths in Europe since February 2020



Note(s):2020 to 2022

Further information regarding this statistic can be found on page 69.

Source(s): ECDC; ID 1102288

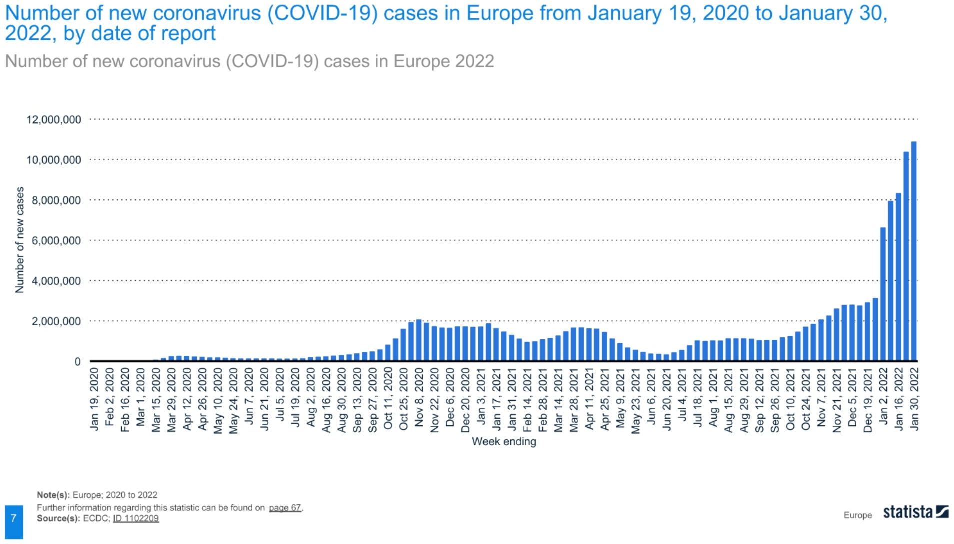
9

Europe statista

Figure 1.1: Covid-19 deaths in Europe

which are more comparable. This figure is probably strongly influenced by the fact that at the beginning of the pandemic, Europe was not prepared to detect it in large numbers of people, but the virus spread rapidly and caused the deaths that were detected. These parameters are evidently correlated because a period of high contagions determines naturally consequently, with a short delay, an increase of the deaths. Returning to the history of the evolution of the pandemic, it can be said that, as far as Europe is concerned, it initially struck Italy and then spread to all European countries, following very similar, though not exactly overlapping, dynamics as can be seen in Figure 1.3.

The apparent similarity of the curves in the various countries should not, however, deceive us about the complexity of the phenomenon. As mentioned above, in fact, it is sufficient to analyse the same parameter relative to the number of deaths in the various countries, but relating it to the number of inhabitants, to reveal trends that are also very different and still not easy to interpret. In this regard, we can look at Figure 1.4, with the number of deaths per million inhabitants in a period similar to that of the previous graph (first pandemic wave).



Number of new coronavirus (COVID-19) cases in Europe from January 19, 2020 to January 30, 2022, by date of report

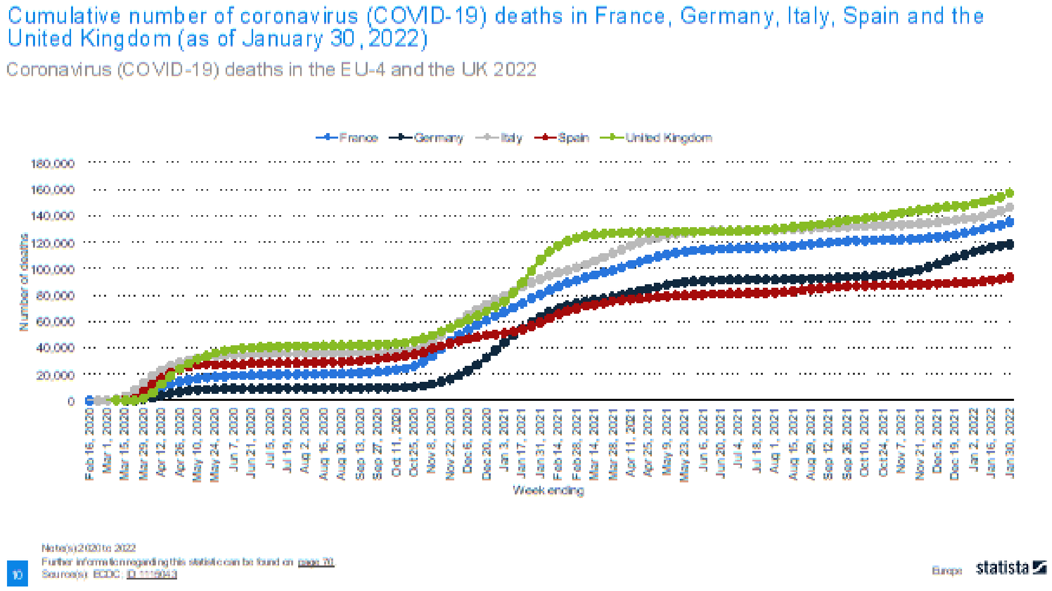
Number of new coronavirus (COVID-19) cases in Europe 2022

Note(s): Europe; 2020 to 2022
Further information regarding this statistic can be found on page 67.
Source(s): ECDC; ID 1102209

Figure 1.2: Covid-19 cases in Europe

Germany, for example, despite having a similar curve, appears to be much less affected in terms of deaths than Italy and France. There are currently no precise explanations for this, other than those already mentioned, based on the different counting criteria adopted by the various countries to differentiate in particular deaths only due to Covid from deaths of Covid and other diseases. The evolution of the pandemic in the various countries has been significantly influenced by the strategies that each State has implemented to contain it and, at the same time, these have been conditioned by the data that began to be released daily with increasing emphasis. In Europe, we have seen very different policies, ranging from 'shock' measures with drastic reductions in individual freedom of movement and assembly, to more minimal ones that relied on a natural course of the disease's expansion.

From a health care point of view, great emphasis was placed on the production and mass administration of vaccines, which, however, were only available in 2021. Therefore, returning to the measures taken to contain the pandemic, it is useful to report here a brief summary of the most significant ones that have had a serious

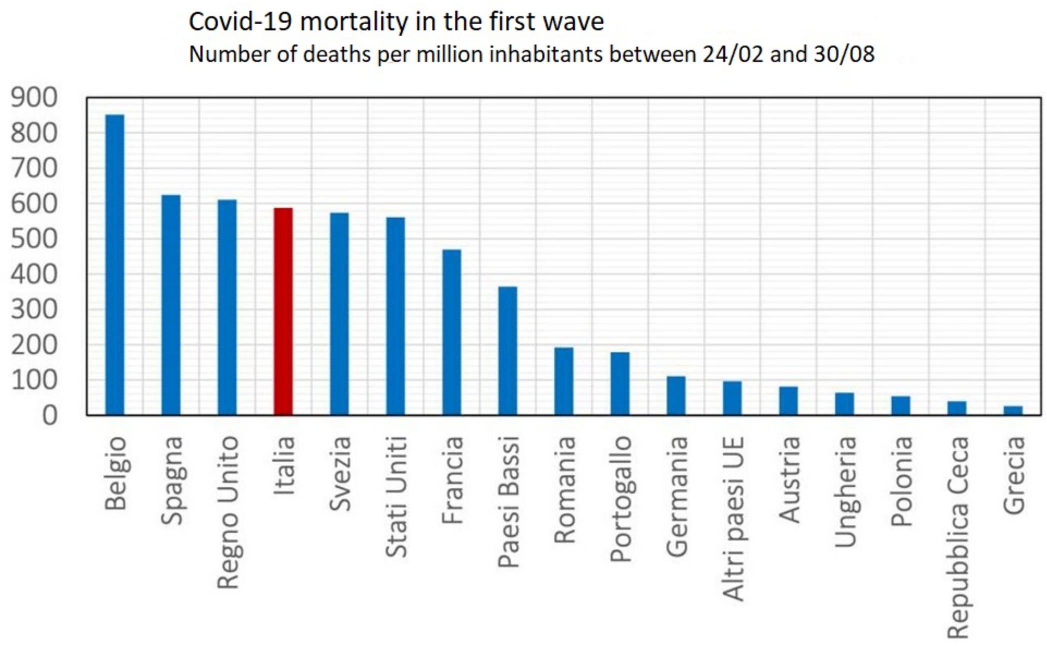


Cumulative number of coronavirus (COVID-19) deaths in France, Germany, Italy, Spain and the United Kingdom (as of January 30, 2022)

Coronavirus (COVID-19) deaths in the EU-4 and the UK 2022

Note(s): 2020 to 2022
Further information regarding this statistic can be found on page 70.
Source(s): ECDC; ID 1115043

Figure 1.3: Cumulative Covid-19 deaths in Europe



Covid-19 mortality in the first wave

Number of deaths per million inhabitants between 24/02 and 30/08

Figure 1.4: Covid-19 deaths per million inhabitants in Europe

impact on people's lives, both from a social and economic point of view. Starting with the restrictions implemented, to varying degrees and at different times in each country, there were closures of schools of all grades and the closing of many jobs; where possible, work from home was also imposed. As for public events, concerts, exhibitions, sporting events, as well as smaller events, suffered cancellations and postponements during the first wave of Covid-19. One of the fundamental restrictions in fact concerned the assemblies and aggregations of people: they were greatly restricted. In several European countries these restrictions concerned not only public places, but also private ones, such as the homes of citizens, who were forbidden to receive a number of people beyond a threshold set by the government. Both international and internal travels were affected. Travel between countries was forbidden except for reasons of absolute necessity and even within the cities for a period of time the rule was in force, which required citizens to stay inside their homes as much as possible. With reference instead to the policies undertaken to protect the health of citizens and to keep the health emergency under control, an information campaign was introduced on the attitudes and precautions to be taken. For example, the use of surgical masks and gels to sanitise the skin and environments was introduced. A system of testing and screening of the population was also used to try to find those infected early and reduce the spread of the disease. Many investments have also been made from an economic point of view to increase the number of places in intensive care and to cope with the increase in admissions as well as for the research of effective vaccines as mentioned above.

Concluding this brief summary on what is certainly the most serious epidemic evolution of these times, we are interested at this point to highlight how the period examined by the research (Jan 2020 - Sept 2020) has been characterised throughout Europe by a crescendo of concern that has pervaded governments, the media and the general population. The topic of coronavirus has become almost all-encompassing, bringing with it a thousand controversies among scientists and politicians that have often contributed to the disorientation of the frailest or oldest people. The containment measures mentioned above have had a direct impact on the quality of life of these people who, often suddenly, have found themselves unable to cultivate relationships with loved ones, to be assisted morally in places of care or even simply to be assisted in daily tasks. The data analysed in this work investigate these aspects during the first wave of the pandemic, when many still believed it would be a phenomenon limited to a few months.

1.2 SHARE survey

The data used in this research was taken from the SHARE survey. The name SHARE stands for "Survey of Health, Ageing and Retirement in Europe" and it is a "research infrastructure for studying the effects of health, social, economic and environmental policies over the life-course of European citizens and beyond"[3]. The target population of the study is the one aged 50 years or older. The idea of targeting this population originated from the fact that the population in Europe is ageing rapidly and therefore new challenges have arisen. These challenges relate to various issues in different fields: the main research areas are income and wealth, health, health care, work and retirement, and social networks. To give a general idea of the survey, we can find in the economic field for example the pension system and its possible collapse, in the social field studies of the solidarity between new and old generations, in the health field various studies on life expectancy and so on.

The first survey was conducted in 2004, and since that moment 530,000 in-depth interviews have been carried out. The data collected are micro-data, concerning the themes above mentioned. They are micro, because they are referred to individuals and their families. Most questions are answered by all household members in the survey, while for some specific topics one member is selected to respond on behalf of the whole household. For general questions, for example about the housing or the sources of income, a household respondent is chosen, while for more financial questions the financial respondent is chosen. A third type of respondent is the 'family respondent' linked to the questions on children and social support.

An important feature about the SHARE data that makes it very useful in various researches is the fact that SHARE is a cross-national survey. It is in fact directed to 27 European countries and Israel; the sample in each country must follow the specificity of the country but at the same time one of the purpose of this survey is to make the answers internationally comparable. This is possible also thanks to the use of ex ante harmonised survey tools and methodologies. All countries in fact administer the same questionnaire (with the country specific language) through the same method of interviewing, and the response rates and collected data are centrally controlled by SHARE. Probability-based sample were designed in each country, based on the fact that “Sample designs may be chosen flexibly and there is no need for similarity of sample designs. Flexibility of choice is particularly advisable for multinational comparisons, because the sampling resources differ greatly between countries. All this flexibility assumes probability selection methods: known probabilities of selection for all population elements.” [4]. Even if sample designs do not need to be equal, all are based on probability principles and have the purpose to minimise errors, in terms of coverage and precision, and to make the sample the most representative possible of the target population.

Another important feature of the SHARE survey is that it is a longitudinal survey. It does not merely provide a static representation of a situation at a certain point in time, but tries to follow the respondents through the ageing process and study their evolution in this important phase of human life. Surveys have been conducted every two years, since 2004. In this way, it is easier to grasp the dynamic character of ageing process. At the moment of writing this thesis eight waves are present, from 2004 to 2020.

As the years go by, the people eligible to take part in the survey change: some of the subjects become unable to respond or have died, while people belonging to younger cohorts grow older and thus become part of the target population. For this reason, in addition to the longitudinal sample, which includes all respondents already interviewed in any previous wave of the study, there are the refreshment samples, in countries where budget allows it, that are composed by new units which were not there in the previous waves. The refreshment sample, beside having the aim to represent also younger cohorts, is useful to compensate for the reduction of sample size due to attrition across waves of the SHARE panel. In addition, not all the countries participated in all the waves: the initial countries in fact were only eleven and gradually the others were added.

The questionnaire is administered by CAPI (Computer Assisted Personal Interview) mode: it consists in a interview done face-to-face, with the help of software to record answers, which can speed up the interview and make it more effective.

1.2.1 SHARE wave 8 and SHARE COVID-19

Wave 8 of SHARE started on October 2019 and targeted 28 countries. The sample included people from previous waves and new people, for the reasons before mentioned. As most surveys of that period it was strongly influenced by the diffusion of Covid-19. Around February 2020 in fact Covid-19 spread also in Europe and SHARE survey had to be interrupted. The situation was confusing and worrying, but on the other hand it was evident that collecting data at that time would have led to very useful research in both the short and long term, precisely because the disease seemed to affect the elderly more severely: a survey of this type, specific to the 50+ population, would have been even more interesting. SHARE therefore decided to start a new survey of the type of the previous ones but with the specificity determined by the pandemic status of Covid-19. First, SHARE started to look for an alternative way to collect data, as face-to-face interviews were prohibited in most countries and threatened the health of respondents and interviewers. The decision was to conduct the survey with a computer-assisted telephone interview (CATI), targeting the COVID-19 living situation of people aged 50 years or older. However, this decision had to take into account a number of factors, including the different use of the Internet in different countries and the older age of the subjects, who are less used to using technological devices. Furthermore, other studies have shown that mode effects on response behaviour and measurement error tend to be larger between interviewer- and self-administered modes than between modes that are both interviewer-administered such as face-to-face and telephone [5]. In

general, the advantages of the CATI interview over the CAPI interview include the fact that the costs are much lower, it is less invasive for the respondent and from an organisational point of view it can be simpler. Both methods, instead, thanks to the use of a software, have the advantage to have a high speed of detection and to minimise the errors made by the interviewer.
In addition to changing the survey technique, SHARE wanted to pay special attention to the issue of Covid-19, trying to detect and measure the main aspects of this pandemic disease and its effects. The solution found was to develop a specific questionnaire for respondents during the pandemic, shorter and with precise questions with maximum relevance for health, economic, work and family events during the first wave of the pandemic. As a result of this, SHARE ended up with different samples than in previous waves, the initial samples of wave 8 and the specific samples of the new Covid-19 survey. Specifically, as reported in the methodological note [6]:

1. the CAPI sub sample contains the data collected before the COVID-19 outbreak by the regular SHARE questionnaire of Wave 8, and its longitudinal part (about 86 per cent) can be merged with the data collected in one or more previous waves.

2. The CATI (Computer Assisted Telephone Interview) sub sample contains the data collected after the COVID-19 outbreak by the SHARE Corona Survey and can be merged with some of the previous waves of SHARE as it consists of longitudinal respondents only.

3. The CAPI & CATI sub sample exploits the full force of the survey instruments implemented in Wave 8 as it contains the data collected before and after the outbreak and can be fully merged with previous waves.

In this thesis, data from the SHARE Covid-19 were used in particular as the intended purpose was specifically related to the effects of the pandemic rather than to the evolution of the senility situation in Europe.

1.2.2 Questionnaire structure

We show briefly how the questionnaire is structured and give an overview of its contents.
The questionnaire from which answers are collected is divided into five main subject areas: specifically they are health (both physical and mental) and health behaviour, COVID-19 infections for respondents and their social network, quality of healthcare, work and economic situation, and social relationships. In various questions the focus is on a comparison between the pre-pandemic and post-pandemic

situation. Our dependent variables are derived from this comparison and are intended to capture differences between these two periods. A lot of questions of the previous waves were deleted, in order to give more space to the pandemic theme. However, questions from previous waves with immediate relevance for the COVID-19 outbreak, such as existing health conditions, medication intake, household income were maintained. Furthermore, some mental health questions from the regular SHARE survey were included and to those who declared experiencing mental health troubles was asked to compare their recent situation to before the outbreak of the pandemic. From these particular questions our dependent variables have been obtained. We want now to show the main focus of every section and describe the variables in it.

Health (physical and mental) and health behaviour

With regard to physical health, the main focus is on a comparison of health conditions from before, e.g. asking if new major illnesses were diagnosed or if certain episodes indicating health problems had occurred. With regard to health behaviour, data collection focused on frequency of contact (e.g how many times the respondent have left home since the outbreak of Coronavirus) and selected activities, such as mask wearing, distancing and hygiene measures. Regarding mental health, the questionnaire aimed to investigate mainly being sad or depressed, having problems sleeping, feeling nervous or tense and feeling lonely.

Covid-19 infections

In this section it was decided to investigate the prevalence of exposure to COVID-19 in the over-50 population in Europe and Israel. Respondents were asked if they themselves or any relations had had COVID symptoms, positive tests, negative tests or hospitalisations, and if so, the number of people. With a further question it was asked if respondents knew anyone who had died due to/with COVID-19 and who was that person.

Quality of healthcare

The section on quality of health care aims to investigate interruptions in access to care during the pandemic period, whether caused by voluntary cancellation of medical appointments by respondents, or due to cancellation or denial by health care providers. By conducting the SHARE survey early in the pandemic, the questionnaire provides only a first look at questions such as access to the types of care that were interrupted, respondents' assessment of their satisfaction or dissatisfaction with the health care they received, and any reasons for this, such as long waiting times.

Work and economic situation

Another of the main immediate consequences of the pandemic was a widespread increase in unemployment and a readjustment of work patterns in many cases. The questions in the section on work first investigate whether respondents lost their jobs due to the pandemic and, if so, for how long. It is also intended to understand the amount of hours worked, so some questions deal with the increase or decrease in the number of hours worked since the outbreak of the pandemic. In addition to the loss of jobs and the difference in hours worked, there has been a change in the way work is done, in terms of the availability of a home office or the safety measures implemented in the workplace and consequently the perceived safety of workers. Further questions therefore concern the new technological skills that had to be adopted or the provision of security measures. In addition, the COVID-19 shock led to important changes in the incomes of many households. Questions were therefore asked about household income, the difficulty of making ebds meet and the financial support received as a result of the COVID-19 crisis.

Social relations

In the final section of the questionnaire, the impact of the pandemic on social relations is analysed. Respondents are asked about the different types of contact they had with their social networks, including their children, parents, relatives and friends, and the frequency of these contacts. In this section there are also questions about receiving and providing care and volunteering work. Furthermore, to those respondents who regularly received home care before the outbreak were also asked questions about the difficulties that they could have experienced in receiving care during the pandemic.

1.3 Depression and loneliness

1.3.1 What depression is

As this thesis deals with a widely defined and studied medical phenomenon, a brief summary of current knowledge in this field is useful. According to a definition from the "State of Mind", the Journal of Psychological Science [7]: "Depression is a disorder of mood, a psychic function important for adaptation. The mood is generally flexible: when individuals experience pleasant events or situations, it bends upwards, while it bends downwards in negative and unpleasant situations. Those suffering from depression do not show this flexibility, but their mood is constantly flexed downwards, regardless of external situations." Depression in the medical sense is thus recognisable through commonly known symptoms even when they do not yet have a pathological characterisation, but simply moodiness: constant dissatisfaction and a negative view of oneself and of life's circumstances, prevalence of pessimistic views of one's own and others' futures, sloth and indolence. A significant classification of the factors characterising the illness was proposed by [8] who identified five elements:

1. A specific mood alteration: sadness, loneliness, apathy.
2. A negative self-concept associated with reproach and self-blame.
3. Regressive and self-punishing desires: desires to run away, hide or die.
4. Vegetative changes: anorexia, insomnia, loss of libido.
5. Change in activity level: retardation or agitation.

The depressed person normally tends to isolate himself from relationships by not sharing the constructiveness and optimism of other people. Conversely, it is also common that, if not compensated by important levels of affection, the attitude of those closest to them will tend to be exclusive of the relationship as it is burdensome to sustain in the medium and long term.

1.3.2 What causes depression

Depression is a phenomenon that can affect everyone. Scholars agree that it is often a feeling of loss that triggers the onset of the disorder. However, the causes of depression are manifold and vary from individual to individual (heredity, social environment, family bereavement, work problems, etc.). Two factors have been identified in the literature as the main sources of risk. First, the biological factor: some people are born with a greater genetic predisposition to depression. Second, the psychological factor: experiences and behaviours learned during one's life history (e.g. mental rumination) can make one vulnerable to depression.

In its most basic form the depressive affection corresponds to an experience of impotence, that is to say, to being helpless and desperate in the face of a situation that has occurred and that cannot be changed [9]; this accomplished fact, felt as irreversible, has introduced a negative change at a psychological level. Therefore, as mentioned above, the 'key' event that can trigger the depressive event is the loss of someone or something, considered fundamental to preserve the subject's psychological well-being. In addition to this, this change also reflects on the subject's self-assessment and self-esteem: the subject feels emptied, due to a decline in inner stability and in the perception of his own value and abilities. Depression therefore always implies a decline in self-esteem and an impoverishment of the self. However, other factors are necessary for self-esteem to be impaired and depressive affect to occur. So, [10], taking up [11], points out that, in addition to the loss, there must also be a lack of acceptance of the loss itself and the persistence of the desire for the lost object, a desire destined to remain unsatisfied. Another element that characterises all forms of depression is, in addition to the lack of acceptance of the loss of the object, the presence of aggression against oneself.

1.3.3 How depression manifests itself

Depression can manifest itself as a pathology either gradually, through an increase in typical behaviour and the increasing temporal persistence of negative attitudes (interspersed with increasingly limited periods of constructive living), or it can develop rapidly following, for example, events of deprivation or failure. Detectable symptoms are:

- biological: widespread pain, weight alteration, gastrointestinal disorders, asthenia;
- psychosomatic: constant fatigue, sleep disturbances (both insomnia and hypersomnia), sexual inappetence;
- cognitive: mental rumination (obsessive introspection of one's ailments), inability to make decisions even in unimportant matters, inability to cope and solve problems;
- emotional: despair, dissatisfaction, sadness, anxiety;
- behavioural: social self-isolation, complaining, in some cases suicide attempts.

It is usually important to detect the first symptoms described above in order to intervene before a chronic pathology develops. At the same time, it is important not to confuse the disease with the normal mood fluctuations that characterise people's lives.

1.3.4 How depression can be treated

There are various treatments and cures to help those suffering from depression. Two key therapies are antidepressant therapy and psychotherapy. Antidepressant therapy works only on symptomatic individuals and is necessary when symptoms are so severe that they affect their social, working and emotional lives. Therapeutic treatments refer to different branches of psychology: psychoanalysis, psycho-dynamic psychotherapy, cognitive-behavioural therapy and pharmacological therapy. Recent studies carried out on samples of depressed subjects (a study carried out in collaboration with four German universities, comparing cognitive-behavioural therapy with psycho-dynamic and psychoanalytic therapy in depressed patients at a follow-up of three years) have shown that cognitive-behavioural therapy is very effective in the short term, but does not guarantee the absence of relapses in the long term. On the other hand, psychoanalysis is significantly more effective than cognitive behavioural therapy (three years after the end of treatment) from several points of view: it is more effective on the symptoms of depression, on interpersonal functioning and on improving the self-scheme.

1.3.5 Differences in depression in European countries

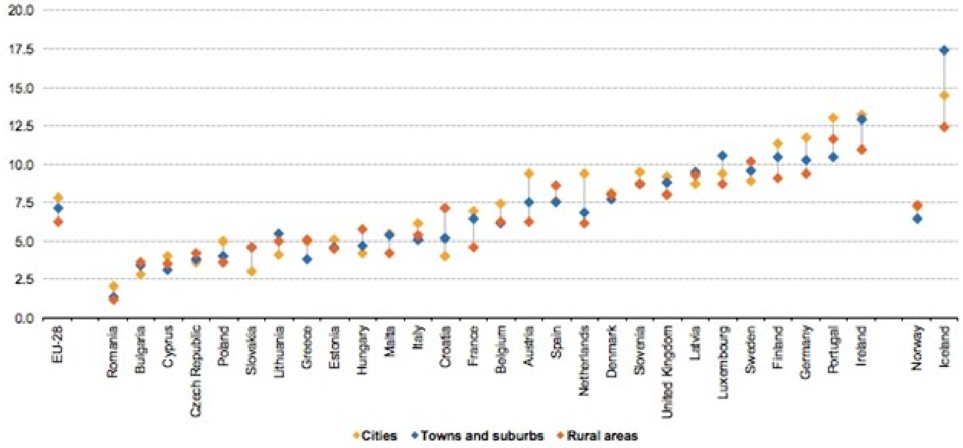
The question of whether there are significant differences in the prevalence of depression in different European countries was answered to some extent by the European Health Interview Survey (EHIS), which analysed data collected between 2014 and 2015 [12]. According to this, about 6.6% of Europeans were reported to suffer from depression. However, this average was the result of very different situations in different countries, ranging from 10 in Luxembourg to 2.7 in Czech Republic, as shown in Table 1.1. The table also shows that the predominance of the pathological phenomenon among women is a constant in every country, with a peak in Portugal (12.9% vs. 4.7%).

Another difference found by the above-mentioned research concerns the data collected within each country, distinguishing between population living in urban, peripheral and rural areas. While 7.8 % of the urban population reported suffering from depression, this value dropped to 7.1 % in the suburbs and 6.2 % in rural areas. However, there are important differences between the various states. For some of them, the figure is consistent with the general average, showing a prevalence of the disease in urbanised areas (e.g. Ireland, Portugal, Germany and Finland). For others, instead, the values are reversed, with rural areas predominating (e.g. Sweden and Spain). The data are shown in Figure 1.5.

As seen in Table 1.1, there are important differences in depression by gender, in the elderly, as in the rest of the population. Women seem to suffer from it twice as much as men. This differentiation seems to arise after puberty (previously not present) and continues throughout life. These are some of the conclusions of a major review on the subject [13], which considered 85 empirical studies conducted on all continents. The aim of that study was precisely to check whether the difference persisted into old age, and it did. The explanation for this difference must be based on various factors: some of a social nature, such as the role in the family and the different employment situation, others of a biological and physical nature such as genetic predisposition and the modulation of the neuro-endocrine system in response to fluctuations in sex hormones. Symptoms revealing the disease are also differentiated by gender: women often present physical symptoms such as fa-

Prevalence of depression symptoms by country and gender			
Country	**Women(%)**	**Men(%)**	**Total(%)**
Austria	5.1	3.4	4.3
Bulgaria	8.0	6.0	7.1
Croatia	3.4	3.4	3.4
Cyprus	5.2	3.0	4.1
Czech Republic	3.4	2.0	2.7
Denmark	9.5	5.3	7.4
Estonia	8.0	5.0	6.6
Finland	6.4	5.7	6.0
France	9.0	5.2	7.2
Germany	10.8	7.6	9.2
Greece	3.8	2.5	3.2
Hungary	9.6	7.1	8.5
Ireland	8.8	6.6	7.8
Italy	5.6	3.5	4.6
Latvia	5.8	3.3	4.7
Lithuania	4.1	2.3	3.3
Luxembourg	11.7	8.2	10.0
Malta	4.4	2.2	3.3
Poland	5.5	4.0	4.8
Portugal	12.9	4.7	9.1
Romania	5.1	4.7	4.9
Slovakia	3.4	2.3	2.9
Slovenia	7.3	4.0	5.6
Sweden	11.2	6.5	8.8
United Kingdom	8.6	6.1	7.4
EU	7.9	5.2	6.6

Table 1.1: Prevalence of depression symptoms by country and gender



Note: The figure is ranked on the share of the total population reporting that they had chronic depression.
Source: Eurostat (online data code: hlth_ehis_cd1u)

Figure 1.5: Prevalence of depression by different living area

tigue, appetite disturbances, insomnia and other sleep disorders, while for men the pathology typically has consequences such as addictions to alcohol or gambling, work obsessiveness and, more frequently than for women, suicide attempts [14].

1.3.6 Depression and loneliness in the elderly

As this work deals in particular with depression in the elderly and its connection with situations of loneliness, it is interesting to explore this in greater depth. A first obvious link between the two is the loss of significant and established relationships and consequent isolation. But loneliness can be real (elderly people living alone) or perceived (therefore linked to the lack of significant relationships). Scholars from the University of Calabria [15] have taken a closer look at this issue with the publication of the paper "Elderly people and depression: the role of loneliness". One of the main objectives of the study was to understand the relationship between social support, loneliness and depression. Depression and loneliness in the elderly are two closely related concepts: for an elderly person, cultivating relationships becomes increasingly difficult, due to deaths among peers, illnesses that may reduce the ability to move around and get out of the house, bereavements in the family (e.g. death of a spouse). All these elements may favour the development of a sense of loneliness and depression; on the contrary, in contexts rich in social interaction the elderly tend to age better, reducing the rates of senile depression. The researchers of the study distinguish between objective and subjective social support. Objective social support refers to the practical help the older person receives from friends, family and others. This can take the form of financial help or information on how to solve problems. When we consider subjective social support, we refer to the emotional side of this support, i.e. the sense of psychological closeness felt by the older person. In this sense the need to belong to a group (often the family but not only) and the affection and support provided by the group emerges clearly in old age. The interpersonal bond with the spouse or children, if any, becomes much more central in this phase of life. If these ties are broken, one finds oneself in the situation where primary relational needs remain unmet. A possible solution to this is group treatment for senile depression, which, although not very widespread, seems to be very useful.

Loneliness can therefore be a cause or effect of depression. Normally the level of loneliness in a person's life is related to the discrepancy between the number of desired and actual relationships, but this purely quantitative emphasis does not adequately value the quality of relationships and social support. If for the interest of this thesis we underline the correlation between loneliness and depression, we must not forget that various studies have highlighted the decisive influence of loneliness in the development of other pathologies such as dementia or Alzheimer's disease. In the psycho-geriatric field, therefore, the prevention of subjective or objective isolation plays a fundamental role in primary and secondary prevention.

1.3.7 Effects of the Covid-19 pandemic on depression known from the literature

Depression in the medical sense, framed in a more general mental health context, is unfortunately an increasingly widespread phenomenon in Europe and a major cause for concern, as evidenced by the 'Health at a Glance: Europe 2018' report, which notes that one in six people in EU countries - around 84 million individuals - had a problem of this kind in 2016.

The spread of the pandemic in such a context has affected the problem of depression through several mechanisms. The World Happiness Report of March 2021 by the European Parliamentary Research Service identifies four main ones:

- health-related anxieties arising directly from Covid-19, such as the likelihood of being infected, the likelihood of being hospitalised or dying, the likelihood of infecting others and the likelihood of loved ones being infected or dying;

- the mental health consequences of concerns arising from how the pandemic might affect a person's financial situation, both in the short and long term;

- the complications arising from family logistical dynamics during periods of isolation;

- the direct mental health effects of the loss or limitation of activities caused by the pandemic and various isolation policies

Similar categories were also described during a November 2020 webinar organised by the European Parliament's Committee on the Environment, Public Health and Food Safety (ENVI), by Natasha Azzopardi-Muscat, Director of the Country Policies and Health Systems Division, WHO Regional Office for Europe. Concerns identified by these reports were reflected in an iterative electronic survey by Eurofound (third round, February/March 2021) which showed that mental wellbeing was at its lowest level in all age groups since the start of the pandemic. The greatest depressive effects, however, were not found among the elderly but among those who had lost their jobs and among young people. In this context of general distress, the question of how the elderly in particular are coping with the depressive effects of the pandemic seems to require a more complex response. While there was initially concern that isolation during the pandemic might be more difficult for this group of people, both at home and in residential care facilities, and that it might worsen existing mental health conditions, recent studies seem to indicate that older people may be more resilient to the stresses of the pandemic than other age groups. However, experts warn that the elderly are a highly diverse group, and that each person's response to the stresses of the pandemic (in other words, their resilience) depends on the particular set of individual circumstances. Moreover, the long-term effects of the pandemic on the health of older people are still unclear. A November 2020 study linking loneliness with psychiatric symptoms (including anxiety and depression) in older adults found it interesting that the effect of loneliness on psychiatric symptoms was more pronounced among participants who subjectively felt older than their chronological age, while those who subjectively felt younger than their age did not show such symptoms [16].
At the European level, many questioned whether there were particular differences in the resilience of older people in different countries which also had partially different policies for dealing with the coronavirus. An analysis of data from surveys conducted in Denmark, France, the Netherlands and the UK compared patterns of loneliness, worry, anxiety and Covid-19-related behaviour among more than 200,000 participants [17]. It found that people responded in psychologically similar ways to the pandemic and associated preventive measures, despite differences in government approaches. The analysis also observed consistency in key mental health indicators across the four countries.

Chapter 2

Preliminary analysis

This chapter will describe the data cleaning operation, to understand what choices were made and what variables we considered keeping for the purposes of the research. This step is crucial in data analysis, because the decisions made create the basis for data processing and modelling. Only an adequate starting dataset, in fact, will produce a reliable result. Therefore, after comprehending the data in terms of structure and size, it is of interest to simplify the data frame, possibly by aggregating certain levels of specific variables or summarising information contained in multiple variables through a single one. The problem of missing data will then be analysed, showing how they have been handled and what decisions were taken.

Thereafter, it is suitable to find new or additional variables that would help with the research objective. Specifically, there is a dataset fusion step: variables of interest that are assumed to be fixed over time are taken from wave 7 of the SHARE survey for people interviewed in wave 8 as well. This allows for the development of a more complex and extensive analysis that will hopefully adequately explain the relationship between COVID-19 and mental health. A thorough and careful exploratory analysis through tables and graphs will then provide a first impression on the data and the distributions of different elements, for both dependent and independent variables. We recall that all the variables described in this chapter are taken from the questionnaire, reported in Appendix A.

2.1 Pre-processing

The folder related to SHARE COVID-19 downloaded from the SHARE website contains different datasets. In particular, observations for Austria are contained in a separate dataset. This is due to the fact that in Austria the survey was conducted between July 20th and September 30th 2020, while in other countries data were collected earlier, precisely between June and August 2020. However, we consider it appropriate to combine the two datasets and to analyse them together and we will properly keep it in mind for the calculation of the age. The dataset so composed contains observations on 57,303 individuals from 38,943 households. The variables present are referred to the questions of the questionnaire. We find also four identifiers: one is for the person and it is fixed between waves, the others are respectively for the household, and when is possible for the partner and the couple. This will allow comparing the observations during time, whether you want to consider the individual respondent as a statistical unit or the couple or household.
In addition, there is a specific dataset that contains information related to basic demographic data and household composition: the coverscreen module. It includes 95311 observations, 38008 more than the main dataset. This is due to two facts: first of all, in the coverscreen data also household members that did not participate to the interview are represented; secondly, it covers the entire wave 8, which started before the pandemic. Some of the respondents were interviewed both before and after the Covid-19 outbreak, while others were in the sample only before Covid-19. Specifically, 36674 units belonged to the former group and 20629 to the latter. From the coverscreen dataset the information that is deemed most necessary and useful for the analysis are incorporated.
At this point it is necessary as a first data manipulation operation to transform the encoding of some variables, for a correct interpretation. All the variables that do not take numerical values or that are not ordered are converted to factor type variables. The choice is to convert in factor also the binary variables.

2.1.1 Combining levels

As anticipated in the introduction of the chapter, it was intended to combine the levels of some variables. The changes made will be shown, starting with the dependent variables and continuing with the variables belonging to the different sections of the questionnaire.

Dependent variables

We considered as dependent variables those related to the comparison in depression and loneliness before and after the Covid-19 outbreak. These dependent variables were created from the following variables: for depression *camh002-* and *camh802*, and for loneliness *camh037*, *camh837*. The choice was to make the variables dichotomous. Analysing in detail, for the study of depression the question *camh002-* was *"In the last month, have you been sad or depressed?"* and in the case of a positive answer, *camh802* asked if it was more so, less so or about the same as before the outbreak of Coronavirus. The binary variable created from these questions takes the value 1 if the respondents state that they are more depressed or sadder than before the pandemic and 0 otherwise. In this way are considered as missing data only individuals whose answer to the questions was "don't know" or "refusal". The same logic was used for the study of loneliness: a dummy variable was created from the question *camh037*, that was *How much of the time do you feel lonely? Often, some of the time, or hardly ever or never?*. If the respondent answered often or some of the time, *camh837* asked if it was more so, less so or about the same as before the outbreak of Covid-19. Those who answered "more so" were coded as 1, while those who answered "less so" or "about the same" were codified as 0. All people that in *camh037* answered "hardly ever or never" were also coded as 0. The dependent variables will be called from now on respectively "depression" and "loneliness".

Corona-related questions

Moving on to the section regarding Coronavirus infection, it has been hypothesized for the questions about symptoms, positive or negative test, hospitalization or death of the respondent himself or someone belonging to his sentimental sphere, to create dummy variables depending on who the person was. For each of the states analyzed binary variables were generated in this way:

- the first dummy *no-one* assumes value 1 if the answer was no-one and 0 otherwise

- the second dummy *outside* assumes value 1 if the answer indicated at least one person outside the household (e.g. other relative outside household, neighbor, friend or colleague, caregiver, other) and 0 otherwise

- the third dummy *inside* assumes value 1 if the answer indicated at least one between respondent, spouse or partner, parent, child or another household member and 0 otherwise

This subdivision was made because in the field of Covid, proximity to infected people is really relevant and the closer a person is to the sick person the higher is the probability of being infected, and this could have an influence on the fear of contracting the disease and on the mental health of the subject. Furthermore, it is immediate to assume that if the person diagnosed with Covid-19 is close in an emotional sense to the respondent, he or she will be more concerned than for less close people.

Health questions

Some modifications have been introduced also in the health section.
For the variable *cah004*, indicating whether the respondent had been diagnosed with a major illness since the last interview, the levels were aggregated in this way into a new variable called *major illness*: the variable took the value 0 if no illness was diagnosed, 1 if only one of the required illnesses was diagnosed, and 2 if at least two illnesses were diagnosed; those who refused to answer the question or did not know how to answer are considered missing data.
It was therefore decided to aggregate the variables *caph089-1*,*caph089-2*,*caph089-3*,*caph089-4*, that are referred to having had episodes of falling down, fear of falling down, fatigue or dizziness into a single variable, called *falling*, indicating whether or not at least one of these episodes occurred. This decision stems from the thought that all of these are similar episodes and can be more simply represented by one index. Again, answers categorised as "don't know" or "refusal" were considered as missing data.
A similar logic was also adopted for the variable *cah006-*, that concerns having taken some prescription drugs for health. For this reason, the levels have been reduced to three: if the subject has not taken any prescription drugs, the variable takes on the value "0", if he has taken only one, it assumes the value "1", otherwise it has the value "2+". The solution for answers "don't know" or "refusal" is the same as for the previous variables. Finally, the questions *cah020* and *cah021*, that ask if a person felt nervous, on edge or anxious and to compare it to before the pandemic, were reunited into a single binary variable called *nervous*, that indicates whether or not the subject has felt more nervous, anxious or on edge after the Coronavirus outbreak.

Other questions

In the work section of the questionnaire, question *caw003*, that indicates for how long the respondent was unemployed, laid off or had to close the business, was divided into classes referred to the period of time in weeks: 0,1-4,5-9,10-20,20+.

There were some impossible values (the length of time registered is greater than the length of the pandemic) that were put in the last class. The subsequent change made in terms of merging levels concerns the section of the economic situation. In particular, the variable *cae004-* contains the answers at the question about any financial support received. It has been chosen to limit the possible levels to "employer", "governments" and "others", without discerning between relatives, friends or others.
Finally, also for the questions about social networks some modifications were made. The original questions considered 5 answers, while we considered to maintain 3 different replies.
The responses regarding the frequency of personal contact or by phone, email or electronic means contact (specifically variables *cas003* and *cas004*) were grouped creating the following levels:

- "never or almost never" if the answer was "never"
- "at least once a week" if the answer indicated "about once a week" or "several times a week"
- the variable assumes value "daily" if the answer indicated "daily"

This variable could be really useful to study the mental health, as we expect that isolation and loneliness are very related. They will be called *physical contact with children*, *online contact with children*, *physical contact with friends*, *online contact with friends*.
It is natural that all the original variables that have been grouped and aggregated are no more of interest, so they won't be used as explanatory variables in the various models, in order to avoid having variables that measure the same concept.

2.1.2 Variables from cv-share8 and deletion of some variables

As mentioned above, we chose to take some variables from the coverscreen module that can offer useful indications on the characteristics of the respondents and their households. It is important to notice that the cv-module is completed by only one household member, who provides information also for the others. In some case, an outsider person fills out the questions for some respondents, so the variable "cvresp", which indicates if the unit is the person that has done the coverscreen module, assumes value 2.
In Table 15 in the Appendix A the factors that we decided to keep are reported. However, it is important to note that some of them are used only to create other variables and not as explanatory variables in models.

2.1.3 Missing data

The original dataset, that is the one without the aggregated variables but already with data from all the countries together, contains 1513 missing values, that are observations that include at least one variable value encoded as NA. As observations with missing values are only 2.6% of the total sample (57,303 observations), we delete them. Thereby, the new dataset has a size of 55790 units. We will explain now how we handled missing data in this dataset of 55790 units. The strategy adopted to handle missing data is different depending on the type of question and consequently on the type of possible answer. Missing codes are "-2" or "-1" for all variables in the questionnaire except the financial ones. The first one expresses the refusal of the respondent to answer, while the second indicates that the respondent does not know how to answer. Instead, for financial variables, which are those in the economic section, the unwillingness to answer is codified as "-9999992", while the fact of not knowing how to respond is reported as "-9999991". We are now considering all the variables, including those generated and described in the previous paragraph.
For single answers, that are not followed by filtered questions, the decision was to unify the two codes "-2" and "-1" into a unique one, called "missing". To give an example, for the variable *caa006*, which regards temporarily moving due to Corona, there were three answers classified as "refusal" and five as "don't know". We only say that there are eight missing values, without specifying and distinguishing by reason.
For those variables that instead come from a previous question and are therefore asked conditionally to what has been answered before, the solution had to consider also this factor, so it was slightly more complicated. Those who do not access the filtered question because of the previous answer are coded as -9, whether they had answered "no" previously or if they were missing values; for financial variables they are coded as "-9999999". Also in this case the decision was to unify the answers registered as "don't know" or "refusal" in a single missing code. Again, giving an example can help to explain that.
For the variable *cah010-* the possible answers were "yes" or "no" (the question was about whether or not the respondent left the house during the pandemic) and *cah011-1* asks those who stated to have left home how often they go shopping in comparison with before the outbreak of Covid-19 . The levels of variable *cah011-1* are the followings:

- *-9* indicates that the question is "not applicable". It is in fact the sum of those who answered "no", or "-2" or "-1" in *cah010-*
- *-2* and *-1* indicate if the answer is "refusal" or "don't know"
- *1* is "not anymore"

- *2* is "less often"
- *3* is "about the same"
- *4* is "more often"
- *5* is "does not apply"

In this case, respondents whose answer to *cah010-* was "no" are placed in the "5" level of *cah011-1*, i.e., "does not apply." Finally, for both those whose response to *cah011-* is codified as "-9", because it was "-2" or "-1" in *cah010-*, and those who answered " -2" or "-1" the responses were considered "missing". For some variables that don't have the level "no" or "zero" in case of necessity it has been added a level, specific for missing values. It is, for example, the case of *caq027*, from now on called *satisfaction with hospital treatment.* It comes after the question *caq025*, which asks if the person was treated or not in the hospital (possible answers were yes or no). The possible answers to *satisfaction with hospital treatment* are from 1 (very satisfied) to 4 (very dissatisfied). The decision was in this case to put those that had answered "no" to *caq025* in a new level called "5", to distinguish them from people that were "missing" both in *caq025* or *satisfaction with hospital treatment.*

For the economic questions (we are referring in particular to *cahh017* and *caco007*), which were answered by only one household member, responses were attributed to all persons in the household, using the *hhid8.* For quantitative variables as *cahh017*, which asks the overall monthly income, some precautions had to be taken: the currency is not the same in all the considered countries, so the value was made uniform and reported in euro, dividing the answer for the variable *exchange rate*, indicating precisely the exchange rate in euro. The new *overall monthly income* variable has also been divided into five classes for values from 0 to over 1000.

After all these considerations, it was found appropriate to first eliminate all the observations for which there was a missing value in at least one of the dependent variables. Thereby the dataset consists of 55429 observations.

Furthermore, analyzing the distribution of the variable *age*, values below 50 years are noted, which could be considered as outliers. Probably they are present in the dataset because belonging to partners or cohabitants in the same household as the respondent. It is therefore assumed that it is appropriate to remove those observations (around 300), as we are interested in studying the effect of the pandemic on the mental health of people over 50 years old.

2.2 Merging datasets

In the previous waves of the SHARE survey the questionnaire was much more detailed and extensive, if compared with the one of wave 8. In fact, the latter mostly considers questions related to certain issues, with a particular focus on the comparison between the situation pre and post Covid. There are whole topics that are not covered in wave 8, and also the one covered have been adapted and reduced, for the reasons mentioned in the first chapter. Focusing on the last completed wave before the outbreak of the pandemic, the issues investigated through the questionnaire were the following:

- Demographics and Networks (e.g. education, marital status,..)
- Children (detects the presence of children and collects information about them)
- Physical Health
- Behavioral Risks (e.g smoking and alcohol use)
- Cognitive Function (detects information about memory, concentration, numeracy and verbal fluency)
- Mental Health
- Health Care
- Employment and Pensions
- Computer Use (frequency and skills in using a computer)
- Grip Strength (measures the maximum hand grip strength with a dynamo meter)
- Social Support (help received or given to others)
- Financial Transfers
- Housing (e.g size and quality of the accommodation)
- Household Income
- Consumption (notes information on household expenditure)
- Assets and activities

- Expectations about the future
- Personality (defines five aspects of the respondent's personality)

The macro-areas that were not then measured are the ones referred to children, behavioral risks, cognitive function, computer use, grip strength, housing, consumption, assets and activities and personality. Instead, information about demographics, physical and mental health, health care, employment and economic situation is contained also in the Corona survey, even if reduced. Taking all the information from wave 7 and using it to analyse the results of wave 8 would be wrong, as it would be implicitly assuming that during the three years from wave 7 (2017) to wave 8 (2020) this information has not changed. Despite this, however, it would be interesting to be able to use some of the information contained in wave 7 to understand how certain aspects may have affected the mental health of the subjects after the start of the pandemic. For this reason, we selected those questions whose answers could be considered fixed in time and for which it could be considered worthwhile to analyse a possible association with the dependent variables. This operation was obviously only done for those subjects who were interviewed in both waves.

The information collected in wave 7 was divided into different modules, depending on the topic it was about. In addition, there is a dataset called "sharew7-rel7-1-1-gv-imputations" that takes into account missing values and through special methods imputes values where possible. There are five imputed values for every observation, to try to take into account the variability due to the imputation process. Imputation was done through two main methods: hot-deck imputation and fully conditional specification method. Briefly, the first one considers similar observations to the one missing and imputes the value of these observations; the similarity is provided by some metric. The fully conditional specification method, instead, is based on Gibbs algorithm and it is an iterative process. For each variable, imputations are derived from imputations on the other variables, that are used as predictors in a regression model.

Using the variable *dn004*- present in the dataset of demographic and networks- which refers to the question asking if the country of the interview is the same as the country where the respondent was born, would have been interesting, for example to see if an inability to fly and return to the country of origin could have had an effect on the individual's mental health. However, this was not possible, as the variable is not present in the imputation file and the original file contains an excessive percentage of missing data (around 74%).

All other variables that have been selected are present in the imputation module, so the choice was made to look directly at the information contained in the imputation dataset, instead of the individual modules, in order to have a reduced

amount of missing data. We will first consider only the values obtained with the fifth imputation (variable *implicat* assumes value 5). However, it was noted that in the "imputation" dataset of wave 7 are not present respondents of Netherlands, and neither in the one of wave 6. This is due to the fact that in waves 6 and 7 the interview in this country was conducted in a mixed way, that is called "Dutch Mixed Mode Experiment" in the SHARE context. [18], which was both by telephone and web interview (CATI and CAWI). Unfortunately, in this case there are no data sets in which all missing data have been imputed. The solution will be to do extra models without this country, to see the effect of the variables on the dependent ones.
A separate discourse is valid for the variable *number of children*: also imputing the same value for the same household, in wave 7 the percentage of missing values is around 22%, while in wave 6 there are not: for this reason we considered to take observations only from that wave, so that the information for this variable is complete.

Variables from wave 7	
Variables	**Description**
mergeid	Fixed id between waves
hhid7	Household identifier
thinc	Total household income -version A
thinc2	Total household income -version B
yedu	Years of education
isced	ISCED 1997 coding of education
nchild	number of children
cjs	Current job situation
fdistress	Household able to make ends meet
lifesat	Life satisfaction
lifehap	Life happiness
bfi10extra	Extraversion
bfi10agree	Agreeableness
bfi10consc	Conscientiousness
bfi10neuro	Neuroticism
bfi10open	Openness to experience

Table 2.1: Choice of variables from wave7

Table 2.1 shows the chosen variables. Some of them are used as explanatory variables in the models, while others are useful in the descriptive analysis for a comparison of the situation before and after Covid-19 outbreak. In particular, we will use as explanatory variables those that can be considered fixed in time.

Firstly, it is acceptable to consider *years of education* and *isced*, that are referred to education, fixed between wave 7 and 8, because the target population has long passed the usual education period. Intuitively, a similar argument can be made for the variables related to personality. They are derived from a 10-item Big-Five inventory (BFI-10) introduced for the first time by [19]. For each of the five personality aspects detected (extraversion, agreeableness, conscientiousness, neuroticism and openness to experience) there were two items. Extraversion includes some characteristics as sociability, activity, assertiveness and positive emotions. Agreeableness refers to themes such as tender-mindedness, altruism and trust. The personality trait of conscientiousness is related to self-control, to pursuing objectives, to organization and precision. Neuroticism is characterized by tension, anxiety and the tendency to be temperamental. Finally, openness to experience includes both 'open' characteristics such as curiosity and originality and 'intellectual' attributes, as intelligence and wisdom. These variables could help to understand what characteristics have individuals that were most affected by the pandemic [20].

We then find the variable *number of children*. As the reference sample consists of people aged 50 and over, it is assumed that in most cases the number of children will remain unchanged from 2017 to 2020 and therefore this variable is taken into account. However, it has some anomalies, as most individuals have a value of -99, i.e. "Not applicable (missing by design)". Some of these values, however, are present as not missing in wave 6 or previous waves, so we will take them if possible from these waves. It would have been interesting to analyse also the variable *number of grandchildren*, but it is supposed to vary also in few years for people of that age, so it is not possible to keep it.

The variable *current job situation* may be useful to supplement the information contained in the main questionnaire in question *caep805*, that asks if the respondent was employed or self-employed at the time of the outbreak of Covid-19. It allows to distinguish between retired and unemployed people; in most cases, people after entering retirement do not look for another job, so it is not wrong to assume to consider that for people in 2020 too.

Continuing with variables that are useful to make a comparison between answers in wave 7 and wave 8, we find *thinc*, *thinc2* and *fdistress*. In particular, *thinc* and *thinc2* relate to the income of the household and are available in two versions. The first one is obtained by an appropriate aggregation of all individual income components in the household; the second one, instead, was generated by the question on monthly household income. The variable *fdistress* concerns the difficulty to make ends meet and a similar variable in content is present also in the dataset of SHARE COVID-19 survey. A similar reasoning was made for the variables *lifesat* and *lifehap*: they concern life satisfaction and happiness and could be used to see what the state of happiness and satisfaction with life was like before the pandemic.

Variable *life happiness* is taken from the question "How often, on balance, do you look back on your life with a sense of happiness?" and as possible answers has often (1), sometimes (2), rarely (3) and never (4). It is considered as a nominal variable. Variable *life satisfaction* is taken from the question "On a scale from 0 to 10 where 0 means completely dissatisfied and 10 means completely satisfied, how satisfied are you with your life?" and is also considered as ordinal.

2.3 Exploratory analysis

In this section the most relevant exploratory analyses are shown, in order to get an overview of the questions and see how the answers are distributed between respondents. It is also of interest to explore possible associations between variables at a descriptive level. As in Section 2.1.1, results are divided according to the category to which the variables belong.

2.3.1 Dependent variable *depression*: being more depressed or sadder

First, we analyse the absolute frequency of the dependent variable *depression* and then we try to understand its link with other variables. In Figure 2.1 it is shown the absolute frequency of variable *depression*; it is possible to see that people who consider themselves more depressed or sadder than before the outbreak of Covid-19 are a minority, precisely 7849 out of 48425 (16.2%).

Socio-demographic features

It is interesting then to understand the link between the dependent variable and some socio-demographic features of the respondents. Firstly, if we analyse the distribution of these people by gender, as in Figure 2.2, we can observe that between women the percentage of answers that indicate a negative effect of the pandemic is higher than between men. This number results from the ratio of positive answers to total answers. We will try to understand whether this effect can be considered statistically significant through appropriate statistical models.
Another study can be conducted to understand if the percentage of people more depressed or sadder changes with the age. From what we can see from the distribution by age of this kind of people in Figure 2.3, the percentage of more depressed or sadder individuals seems to increase with age: in the last age group, in fact, which is the one containing people over 90 years old, the percentage is almost 5 points more (around 20%) than in the first age group (around 15%), containing people aged between 50 and 59 years old. Finally, we can show the same statistic for the

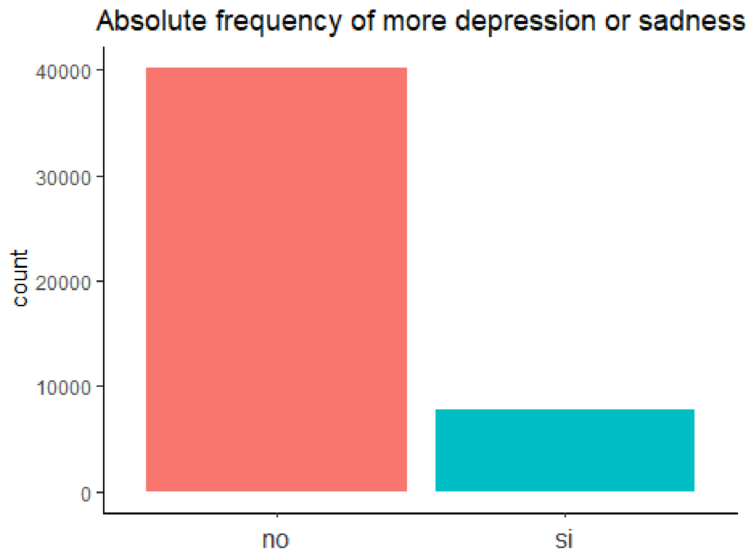


Figure 2.1: Absolute frequency of *depression*

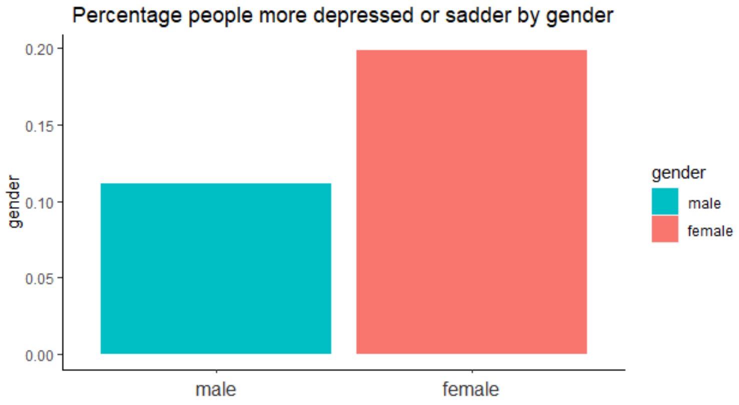


Figure 2.2: Distribution of dependent variable *depression* by gender

variable *country*. As mentioned in the first chapter, different countries adopted different measures against the outbreak, so we expect this proportion to vary among countries. In Figure 2.4 we see striking cross-country differences: the highest value belongs to Portugal, immediately followed by Italy. The lowest instead are those of Slovenia and Denmark, with a maximum difference of 20 percentage points. At first glance, it can be seen that they are in fact countries that have adopted different measures and with a different level of rigidity from Italy for example. As for the previous features, we will find out whether this difference is significant through the models, by including the variable in the explanatory set and possibly then including a random intercept.

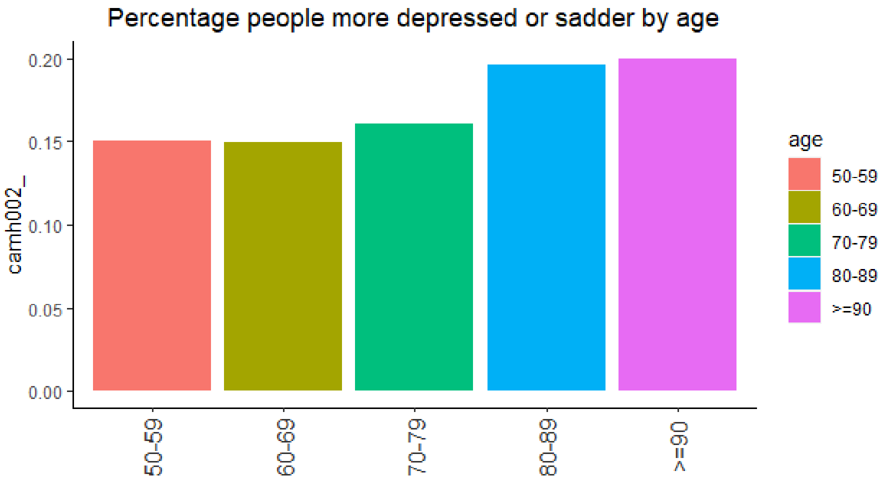


Figure 2.3: Distribution of dependent variable *depression* by age

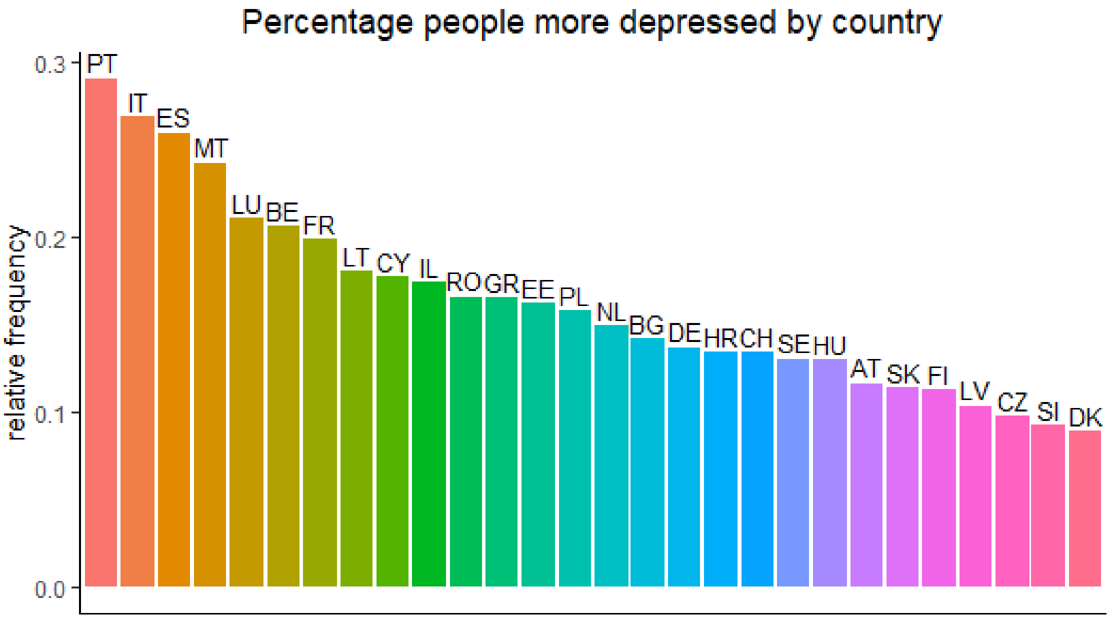


Figure 2.4: Distribution of dependent variable *depression* by country

Connection with other health variables

In addition to demographic variables, it could be interesting to have an overview also of the relation of the dependent variable with the other variables generated by the answers of the questionnaire. For what concerns the health section, we consider the variable *major illness*, created previously and explained in Section

2.1.1. From Table 2.2 we see the proportion of people that declare themselves more depressed or sadder than before the pandemic, by conditioning to the categories of the variable *major illness*. It is to notice that rows do not add up perfectly to 1 as there is also the category of people who answered "don't know" or refused to answer. It seems that with the increase of major illnesses diagnosed during the period, the percentage of depressed people increases. This is reasonable and could be a factor explaining depression, beyond the pandemic. Although this, we see that the majority of people who are more depressed or sadder have not been diagnosed with any new diseases.

major illness			
depression	**0**	**only 1**	**2 or +**
No more depressed or sadder	0.90	0.05	0.04
More depressed or sadder	0.81	0.08	0.10

Table 2.2: Dependent variable *depression* by new major illness

The same descriptive analysis is conducted for the generated variable *falling*, that indicates whether or not episodes of falling down, fear of falling down, dizziness or fatigue are occurred since the outbreak of Covid 19. As before, the response variable is conditioned to the different categories of the variable *falling*. From Table 2.3 we notice that values in the diagonals are almost the same; fundamentally it seems that having had at least one of these disorders contributes to be more or less depressed.

falling		
depression	**0**	**at least once**
No more depressed or sadder	0.63	0.36
More depressed or sadder	0.37	0.62

Table 2.3: Dependent variable *depression* by falling

Connection with Corona related variables

It is then considered important to understand whether and how much Corona related issues have influenced depression or sadness of individuals. Variables linked to corona are, as mentioned above, the presence of symptoms, the negative or

positive test, hospitalisation or death of the respondent himself (apart from death obviously) or someone that he knows. As written in the Section 2.1.1, for each of these features, dummy variables were created with respect to who the person was (precisely if the person was inside the household, outside or no-one).
We want to highlight differences in the conditioned distribution of the dependent variable on these variables. For what concerns the presence of symptoms in people inside or outside the household, in Table 2.4 the conditional distributions are reported.

symptoms inside household and outside household				
depression	**no inside**	**yes inside**	**no outside**	**yes outside**
No more depressed or sadder	0.95	0.04	0.92	0.07
More depressed or sadder	0.93	0.07	0.90	0.09

Table 2.4: Dependent variable *depression* by symptoms inside and outside household

We can see that in the category of more depressed or sadder, the percentages of people who indicated in the questionnaire that they had had someone in or outside the household with symptoms attributable to Covid infection are slightly larger than in the category of people who stated that they were not more depressed or sadder than before. Intuitively, this difference will increase with the severity of the Covid-related infection (e.g. we expect that for the question about death, the difference will increase). Furthermore, comparing the relative frequencies inside the groups, we can see that 80% of who knew someone with symptoms outside the house was not more depressed or sadder, while the 20% had a positive answer to question *depression*. The same values for who had someone with symptoms inside the household are respectively 77% and 23%, so the difference is small, but not irrelevant.
Therefore, we analyse the distributions relating to a positive test, to hospitalisation and to death. For the first, conditional distributions are reported in Table 2.5, always with the distinction between a positive test of someone inside or outside the household. Also in this case the percentage of missing data for the answer relative to positive test is the same for who is more depressed and who is not, and it is around 0.5%. The percentage of people who had someone that tested positive and is more depressed or sadder than before the outbreak of the virus is higher than the one of those who are not more depressed - both for the inside case and the outside one. As before, for who had someone that tested positive inside the household the relative frequencies of being more depressed or sadder are higher than for those who knew someone outside the household that was positive.

positive test inside household and outside household				
depression	**no inside**	**yes inside**	**no outside**	**yes outside**
No more depressed or sadder	0.978	0.016	0.936	0.057
More depressed or sadder	0.970	0.025	0.922	0.073

Table 2.5: Dependent variable *depression* by positive test inside and outside household

Connection with other variables

We can now consider other variables, related to different aspects that could influence the mental health of individuals and that are affected by the Coronavirus emergency.
For what concerns the field of work, we take into consideration answers to *caw002*, *caw021* and *caw024*. The variable *caw002* is a dummy variable and detects whether the respondent became unemployed, was laid off or had to close his business due to Covid-19.
Firstly, the absolute frequency of who had a positive answer is 1793, against the 46602 that answered in a negative way. To see the association with the dependent variable, we analyse the relative frequencies within the groups of respondents who answered negative or positive to *caw002*. For the first group, the percentage of people that declare themselves more depressed or sadder after the outbreak of the disease is 16%, while in the second one is almost 20%.
Moving now to the section on the economic situation, we consider questions *cae003* and *caco007* and their link with the dependent variable. The first one asks if the respondent or someone in his household had to receive financial support from anyone, while the second one concerns the ability of the household to make ends meet. The relative frequencies are reported in Tables 2.6 and 2.7. For the variable relative to financial support we do not find a great difference between the two categories.

received financial support		
depression	**received help**	**not received help**
No more depressed or sadder	0.082	0.918
More depressed or sadder	0.078	0.922

Table 2.6: Dependent variable *depression* by financial support

For the variable relative to the ability to make ends meet, we can notice that in the category of people that do not declare themselves more depressed or sadder after Covid-19 the percentage of who has great or some difficulty is rather lower than the one for who is more depressed. Among those who have difficulties, almost 21% is more depressed, while for who makes ends meet easily or fairly easily the percentage amounts at 14%.

difficulty in making ends meet		
depression	**difficulty**	**no difficulty**
No more depressed or sadder	0.32	0.68
More depressed or sadder	0.43	0.57

Table 2.7: Dependent variable *depression* by difficulty in making ends meet

The following section of the questionnaire concerns the social networks of the respondents during the pandemic. It is supposed to be a fundamental field connected to mental health of people. The questions asked the frequency of contact with own children, own parents, other relatives or others like friends or colleagues. In this explanatory analysis we take in consideration only the frequency of contact with own children, considering the age of the respondents and assuming that they are the closest affections outside the home. For both more depressed and not more depressed the percentage of missing data in the *cas003-1* is around 9%, so we can directly compare the relative frequencies. In Table 2.8 the relative frequencies of *depression* conditioned to the times the respondent had a personal on site contact with his own children are reported. Between the two categories we do not see large differences in the values and this is a surprising descriptive result.

contact in person with own children			
depression	**daily or several times a week**	**about once a week or less often**	**never**
No more depressed or sadder	0.18	0.30	0.43
More depressed or sadder	0.16	0.29	0.46

Table 2.8: Dependent variable *depression* by contact with own children

The same can be said about the variable *cas004-1*, which indicates the frequency of contact by electronic means as for example phone or e-mail. In people who are more depressed or sadder than before the outbreak of Covid-19 the percentage of frequent contact is higher than in the less depressed, as reported in Table 2.9.

contact by electronic means with own children			
depression	**daily or several times a week**	**about once a week or less often**	**never**
No more depressed or sadder	0.34	0.46	0.1
More depressed or sadder	0.43	0.40	0.08

Table 2.9: Dependent variable *depression* by contact with own children by electronic means

Then we can consider the variables taken from wave 7, such as the ones referred to education and to personality. For the education there are two parameters, that are the years of education and the index ISCED 1997. The boxplot 2.5 reports the different distribution depending on the years of education. People who indicate that are more depressed and sadder have a mean of 10.57 years of education, while those that are not sadder have done on average 11.44 years of education, so the difference is not big.

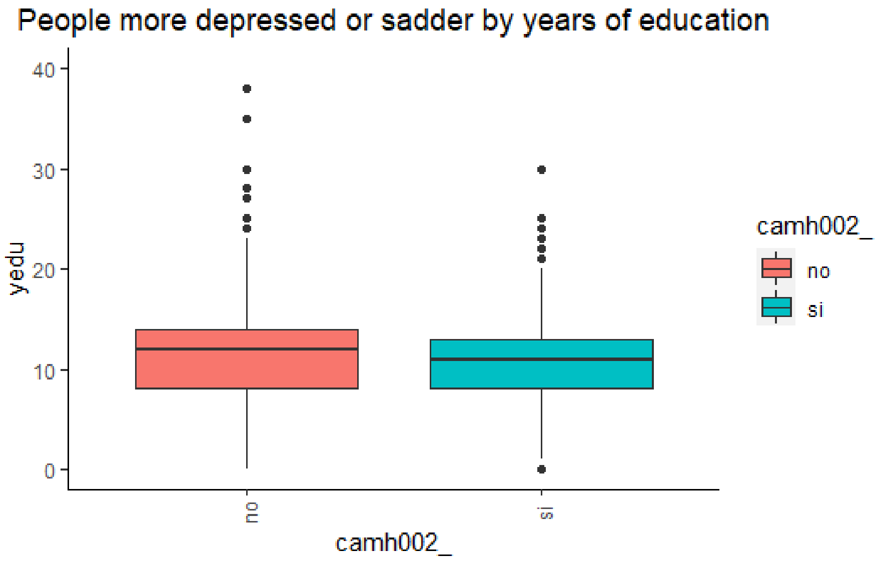


Figure 2.5: Dependent variable *depression* by years of education

Examining the distribution of the variable related to ISCED 1997, we notice that the difference seems also not relevant for this variable; the relative frequencies of each level for depressed and not depressed are similar. The levels of the index are from 0 (pre-primary education) to 6 (second stage of tertiary education). For what concerns the personality, we consider the Big Five Personalities reported in wave 7. They are measured on a 5 point Likert scale, ranging from strongly disagree to strongly agree with the questions regarding the trait of personality. In Table 2.10 the questions for each trait are reported, that are coded one in the positive and one in the negative direction of the scale.

Big 5 personalities		
Variable	**Description**	**items**
bfi10 open	openness	1) I see myself as someone who has few artistic interests 2) I see myself as someone who has an active imagination
bfi10 consc	consciousness	1) I see myself as someone who tends to be lazy 2) I see myself as someone who does a thorough job
bfi10 extra	extra-version	1) I see myself as someone who is reserved 2) I see myself as someone who is outgoing, sociable
bfi10 agree	agreeableness	1) I see myself as someone who is generally trusting 2) I see myself as someone who tends to find fault with others
bfi10 neuro	neuroticism	1) I see myself as someone who is relaxed, handles stress well 2) I see myself as someone who gets nervous easily

Table 2.10: Big Five Personality

If we compare the relative frequencies of the scores between those who said that were more depressed and the other for all the groups, we notice some differences mainly on the personality traits of neuroticism and extra-version.

2.3.2 Dependent variable *loneliness*: feeling lonelier

A similar analysis was conducted for the second dependent variable, *loneliness*. Firstly, 5515 out of 48425 people answered in the questionnaire that they felt lonelier than before the outbreak of Coronavirus; this corresponds to a percentage of 11.3%. We can notice that this value is lower than the one of *depression*, so more people declare themselves more depressed or sadder than lonelier. As in the previous paragraph, we now will consider different topics and we will see the relation with the dependent variable.

Socio-demographic features

If we compare the relative frequencies of people lonelier inside the group of men and women, we can see that among the women the percentage of a positive answer is higher, as illustrated in Figure 2.6; precisely there is a difference between the two classes of almost six percentage points (the value for men is 8% and for women 14%).

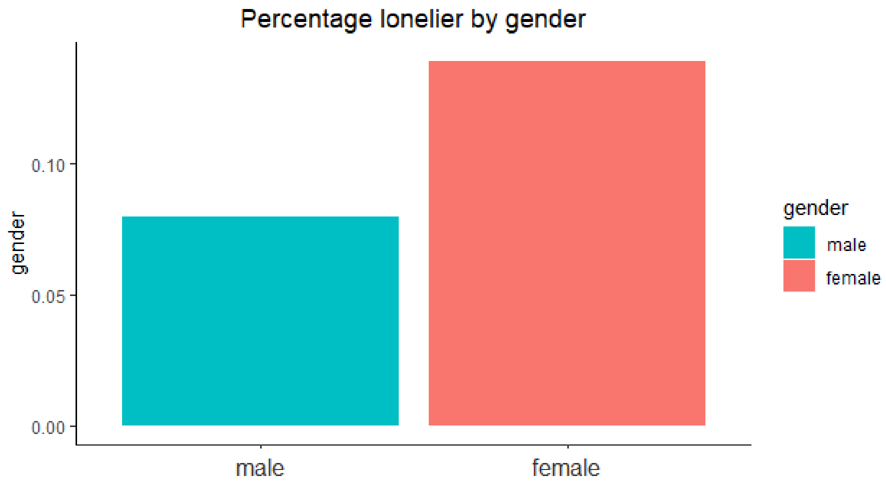


Figure 2.6: Dependent variable *loneliness* by gender

For the different values of the dependent variable by country, the first five countries in which people feel lonelier than before Covid-19 are in order Greece, Italy, Belgium, Cyprus and Sweden; apart from Italy, none of these appear in the top 5 of the most depressed people, so it would be interesting to understand what makes the difference between feeling sadder and lonelier and what are the different factors that affect both. In addition, we see in the top five countries the presence both of a country where measures were very strict (Italy) and one where the government did not enforce such strict measures. So one question we can try to answer is how much the different measures affected people's mental health. For what concerns the difference in loneliness by age, the situation in this case is quite clear. As we can see from Figure 2.7, loneliness seems to increase with age: the percentage of people that feel lonelier than before the pandemic in the class of age of "50-59" is 8%, while for the eldest people that value is 17%. Furthermore we can see that from one age group to the next this percentage always increases and the highest difference is between the age group of "70-79" and the follower.

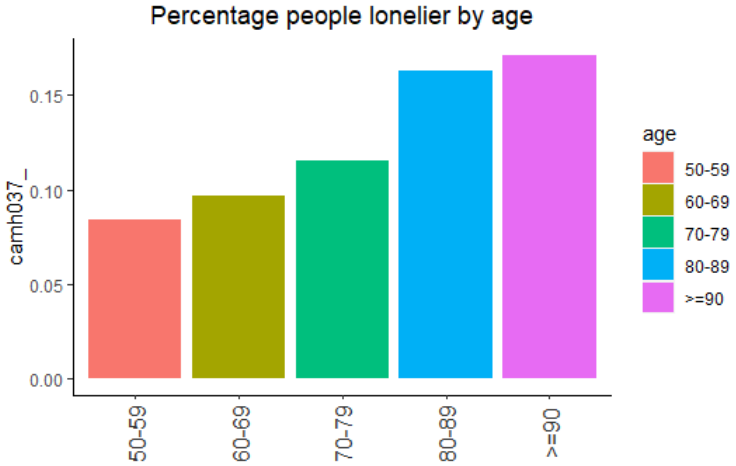


Figure 2.7: Dependent variable *loneliness* by age

Connection with other health variables

It is important to understand the link between the dependent variable and the other questions related to health. Starting with *major illness*, the percentage of people that feel lonelier increases with the increase of new diseases diagnosed, as predictable.

Results from the conditional frequencies of the variable *falling*, which indicates whether episodes of dizziness or falling happened, show us that among who is lonelier 58% had had at least one of the episodes, against the 38% of who is not lonelier. So these episodes seem to be closely related to the aspect of mental health measured by *loneliness*.

Also, the fact of feeling nervous, anxious or on edge seems to be associated with the dependent variable. As we can see in fact in Table 2.11, between those who do not feel lonelier only 17% experiences feelings of nervousness or anxiety, while this percentage is around 54% among those who feel lonelier. Also, 93% of those who do not feel nervous also do not feel lonelier than before the pandemic.

	feeling nervous, on edge or anxious	
loneliness	**no**	**yes**
Not lonelier	0.83	0.17
Lonelier	0.45	0.54

Table 2.11: Dependent variable *loneliness* by feeling nervous, anxious or on edge

A note must be made to understand also the connection between the two dependent variables. Firstly, 94% of people who do not feel more depressed or sadder also do not feel lonelier, while 37% of those that are more depressed or sadder feel also lonelier; 88% of people who do not feel lonelier, also do not feel more depressed or sadder. Therefore, the phenomenon of feeling more depressed or sadder than before the outbreak of Covid-19 seems more common than the one of loneliness.

Connection with corona-related variables

As done for variable *depression*, it is fundamental to understand how features related to Coronavirus disease have affected respondents' mental health.
For what concerns variable referred to symptoms, we may say that between those who feel more alone and those who do not, there is no relevant difference. In comparison with the fact of being more depressed or sadder, it could be less important, because knowing or not someone with symptoms intuitively may have an effect on the person's sadness or fear, but does not affect whether they feel more alone.
If we analyse then the conditional distribution of having at least one person inside or outside the household that tested positive, it is possible to see a slight difference, as shown by Table 2.12 . Among people that have a Covid-19 positive person inside the household in fact the percentage of lonelier people is the highest, compared to the other classes; this result makes sense, as having a positive in the house implies total isolation. There is also a difference in percentage between who does not know anyone that tested positive and who knows someone outside the household: this result is also consistent, because between people outside the house it could be present a "close contact", that forces to have to quarantine.

Positive test			
loneliness	**no one**	**inside**	**outside**
Not lonelier	0.89	0.85	0.87
Lonelier	0.11	0.15	0.13

Table 2.12: Dependent variable *loneliness* by positive test

A similar type of analysis can be conducted for hospitalisation. Table 2.13 shows the frequencies conditional on the dependent variable. We can see that the main difference is between those who had a person belonging to the family nucleus hospitalised and the others: among the former, in fact, 20% felt more alone, while in the others the value is lower than 15%. This could be due to the fact that the sample is largely made up of elderly couples, so it is likely that if one member of the household was hospitalised, the other remained alone at home.

The last of these variables is the one concerning if the respondent knew someone that died because of Covid-19. Among who did not know anyone the percentages

	Hospitalisation		
loneliness	**no one**	**inside**	**outside**
Not lonelier	0.89	0.80	0.86
Lonelier	0.11	0.20	0.14

Table 2.13: Dependent variable *loneliness* by hospitalisation

of people who was respectively not more alone and more alone are 82% and 18%, while the same values for those who knew someone dead for the virus are 89% and 11%, so, as predictable, there is quite a difference.

Connection with other variables

Other variables noted in the questionnaire could have a significant relationship with the dependent variable. In reporting the results, we will follow the same order as in the previous paragraph.
Starting with work variables, we consider that a variable that could be related to this dependent variable is *caw010*. It asks in fact if people worked only at home, in a blended way or only at the work place. Hypothetically, in fact, this could be affect the social life of the respondent and consequently his feeling of loneliness. 8% of those who declared of working only at home felt lonelier than before the Covid-19, while for the other groups the percentage is around 6%. Another difference is noted for variable *caw002*. In fact, among those who denied having lost the job or closed it due to Coronavirus, 94% does not feel lonelier, while for those who answered in a positive way this percentage decreases by six percentage points.
We consider then that variables of economic type are not so related to loneliness, therefore we move to the next set of questions, concerning social networks. We show only results for contact with own children and others like friends and colleagues. Firstly, it is possible to see from Table 2.14 that among who is more alone, the majority never sees own children; furthermore, the percentage of people that see their children at least once a week and are not more alone is five percentage points higher than for those who feel more alone. Also for what concerns friends or others, there seem to be differences: with the decreasing of contact frequency, the percentage of people that feels lonelier than before Covid-19 increases.
We also want to understand if tools such as phone or email have helped in making people feel less lonely. As for the dependent variable *depression*, the results are unexpected. For both own children and friends, in fact, percentage of people lonelier than before the outbreak of Covid-19 decrease with the decreasing of contact frequency.

We now consider some of the variables taken from the previous SHARE wave. For the field of education we found out that both for *years of education*, there

	contact in person with own children		
loneliness	**daily or several times a week**	**about once a week or less often**	**never**
No more alone	0.18	0.29	0.43
More alone	0.13	0.29	0.47

Table 2.14: Dependent variable *loneliness* by contact with own children

are no large differences between who feels lonelier and who does not, as shown in Figure 2.8. The average values for education, are respectively 10.47 and 11.41 years. Also if the differences are slight, values are higher for who does not feel lonelier.
For a general overview of the features of the sample, we can notice that the sample considered has an average of 10 years of education.

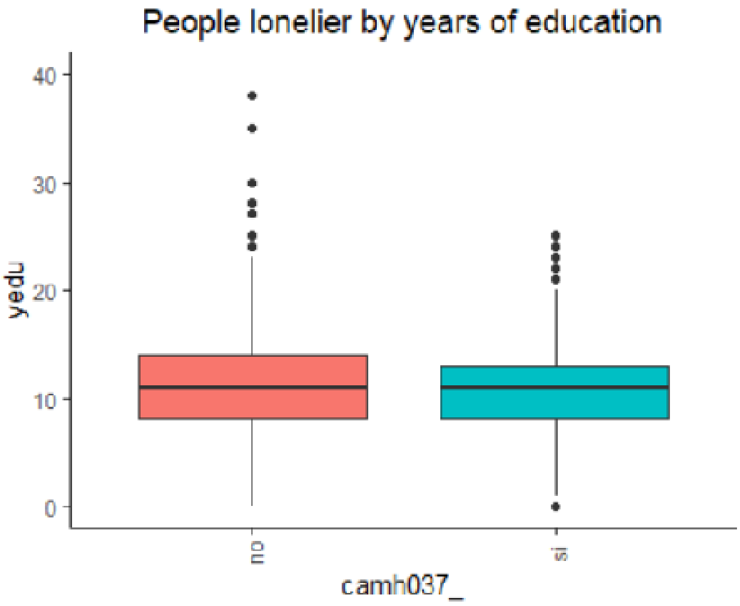


Figure 2.8: Dependent variable *loneliness* by years of education

If we then consider the variables referred to the different types of personality, comparing the relative frequencies for every type of personality, separated according to the value of the dependent variable, we find that, as for *depression*, the more relevant differences concern the aspects of extroversion and neuroticism. Also in this case, in fact, we see that those who feel more alone are less extrovert and more neurotic than who states to not be more alone.

It could be then interesting to know the average level of happiness and satisfaction in wave 7 both for the group of who then feels lonelier and for the other. For what concerns life satisfaction, for those who then feel not lonelier the value is slightly higher than for those who then feel lonelier, as predictable. Variable *life happiness* present instead a different result: for who in wave 8 declares himself more alone the level of "life happiness" was slightly higher than the one for who then says to not feel more alone.

Chapter 3

Model explanation

In this chapter we will explain from a theoretical point of view the statistical models that will be used in this thesis, whose results will be shown in Chapter 4. For each model, in addition to a brief explanation of its structure and properties, an interpretation of the results and an illustration of the advantages and disadvantages of choosing to use it will be provided.

3.1 Model 1: logistic regression

The first model that we will use in the analyses is the logistic regression.
Our dependent variables, *depression* and *loneliness*, are dichotomous and we want to study the relation between them and the other variables. Since this is a study of relation between variables, we could think to use a simple linear regression. Considering though the nature of the dependent variables, doing this would lead to some problems. To recall only one, the domain of the dependent variable is $\{0, 1\}$, and it does not fit the logical set-up of least squares because a linear regression function does not remain constrained within this set [21].

3.1.1 Explanation of logistic regression model

The response variable, called Y, is a dichotomous variable and it can assume the value 1 or 0, both for depression and loneliness, as mentioned in Section 2.1.1.

$$Pr(Y_i = 1) = \pi_i \qquad Pr(Y_i = 0) = 1 - \pi_i$$

The principal aim of the statistical analysis is so to investigate the relationship between the response probability $\pi = \pi(x)$ and the explanatory variables $x =$

$(x_1, .., x_p)$. In all generalised linear models the structure is:

$$g(\pi_i) = \eta_i = \sum_{j=1}^{p} x_{ij}\beta_j, \quad i = 1, .., n, \quad j = 1, ..., p \quad (3.1)$$

In the logistic model, specifically, a linear predictor is specified such that:

$$g(\pi) = log(\frac{\pi}{1-\pi}) \quad (3.2)$$

This function is called *logit* and from here it originates the name of the model. Therefore, we can say that:

$$logit(\pi) = \beta_0 + \beta_1 x_{i1} + ... + \beta_p x_{ip} \quad (3.3)$$

and from the inverse function of the logistic one, we can find that the probability of a positive response is:

$$\pi_i = \frac{\exp(\sum_{j=1}^{p} x_{ij}\beta_j)}{1 + (\exp \sum_{j=1}^{p} x_{ij}\beta_j)} \quad (3.4)$$

The vector of parameters $\beta = (\beta_0, .., \beta_p)$ is estimated through an iterative algorithm, called *iterated weighted least squares*; we will not go into the details of the algorithm, so please refer to [22].

3.1.2 Interpretation of logistic regression model

In this thesis, it is of great interest to interpret model estimates and to understand the relationships between dependent and independent variables. We therefore consider it appropriate to show how the parameters of the logistic regression model are interpreted.

The logistic link function allows the linear predictor to be interpreted in terms of the logarithm of the odd. The odd is represented by $\pi_i/(1 - \pi_i)$ and it indicates the ratio between the probability of success and failure. The single coefficient β_j shows the effect on the log-odds of a unit increase of x_{ij}, net of other explanatory variables. We give an example for both quantitative and factorial explanatory variables.

In the case of a quantitative variable, if we have a linear predictor of the type

$$\eta_i = \beta_0 + \beta_1 x_i \quad ,$$

and for example $\beta_1 = 0.5$, at the increasing of one unit of the explanatory variable, the variation of the linear predictor is equal to 0.5 (net of other explanatory

variables) and consequently the correspondent odds is multiplied by $exp(\beta_1)$, in this case 1.65.
In the case of dichotomous explanatory variables, so assuming that x_i can take values 1 or 0, the log odds ratio is:

$$log(\frac{Pr(Y_i = 1|x_i = 1)/Pr(Y_i = 0|x_i = 1)}{Pr(Y_i = 1|x_i = 0)/Pr(Y_i = 0|x_i = 0)}) = \beta_1$$

Therefore, the interpretation is that the ratio between probability of success and failure for units with $x_i = 1$ is $exp(\beta_1)$ times the same ratio for units with $x_i = 0$.
In the case of factorial explanatory variables with more than two levels, usually the corner parameterization is adopted: the first level of the variables is taken as reference and the parameters for other levels represent deviation from it. The coefficients are interpreted as before in terms of odds ratios, but always referring them to the reference level.

3.1.3 Pros and cons of logistic regression model

Logistic regression model is one of the most used in statistics for prediction and interpretation. It is easy to implement and it provides in most cases a really helpful interpretation: it allows the identification of the most important variables and can give information on size and direction of the association between them and the interest variable. Furthermore, it does not require assumptions concerning the distribution of independent variables, such as normal distribution, linearity and equality of variance-covariance matrix.
However, this model presents also some disadvantages. To mention just a few, firstly, even if it is less inclined to over-fitting than linear regression, the over-fitting is possible. Furthermore, as in linear regression, logistic regression requires that independent variables have no or moderate multi-collinearity: if two independent variables have a high correlation, only one of them should be used to not have repetition of information.

3.2 Model 2: Classification tree

The second model that is adapted to the data is the classification tree. It will be used for both the variables in Sections 4.1.2 and 4.2.2. As done for the logistic regression model, we will briefly show what it is and why we use it.

3.2.1 Explanation of classification tree model

The main idea of the trees (both for regression and classification) is to divide the feature space into a set of regions and then fit a simple model to each one:

the purpose is to identify homogeneous groups of units within the regions and to explore the relation between the variables. Generally recursive binary partitions are mainly used, for their easier interpretability. First the space is split into two regions, and then the response is modelled by the mean of the dependent variable in each region: the variable and split-point to achieve the best fit are chosen. This process then continue until a stop point.

In the problem of classification, we want to approximate the probability of a statistical unit to belong to one of the classes of the dichotomous variable (usually the class are coded as 1 or 0), $p(x) = Pr\{y = 1|x\}$, with a function. The structure of the function is:

$$\hat{p}(x) = \sum_{j=1}^{J} P_j I(x \in R_j) \qquad (3.5)$$

where $P_j \in (0, 1)$ is the probability that Y is equal to 1 in the region R_j. To estimate $\hat{P}_j$, it is possible to use the proportion of units of class "1" in the node j, which represent the region R_j. A choice for the function to minimise at each step of the algorithm (we want to search the minimum "within-node" variability) is the deviance of the binomial distribution, because of the binary nature of the response variable. The deviance is:

$$D = \sum_{j=1}^{J} -2n_j[\hat{P}_j log \hat{P}_j + (1 - \hat{P}_j) log(1 - \hat{P}_j)] = \sum_{j=1}^{J} D_j \quad (3.6)$$

If we rewrite this expression, we can find out that the above deviance is an average of the entropies, weighted with relative size of the leaves (the leaves are at the end of the branches and in them is reported the class to which the units are classified).

$$D = 2n \sum_{j=1}^{J} \frac{n_j}{n} Q(\hat{P}_j) \quad (3.7),$$

where

$$Q(P_j) = -\sum_{k=0,1} P_{jk} log(P_{jk}) \quad (3.8)$$

and is a measure of impurity. The entropy is an impurity measure and is referred to the inhomogeneity of the leaves with respect to the dependent variable.

It is possible to choose other impurity measures, such as Gini index or misclassification error.

After the growth phase of the tree, there is the pruning phase. It is a phase that has the aim to penalise, with a parameter, the size of the tree: in fact, if as many leaves as observations are used, we may come across the over-fitting problem. To prune the tree, the misclassification rate is mainly used, if the goal is prediction accuracy.

3.2.2 Interpretation of the classification tree

When it comes to interpret the results of the classification tree, a really useful and used instrument is the graph of the tree: it shows the split variables and the classes chosen for units are reported in the leaves. Specifically, it is possible to see to which class a unit has been assigned. We can do that if we know the specific values of the set of explanatory variables of the unit. In each node, in fact, there is an inequality, hence the units are split. If the inequality is true, we have to follow the left branch of the tree and continue in this way, otherwise we have to follow the right branch. At every new node we have to "solve" the inequality and follow the consequent path, until we reach the leaves. In every leaf there is the class to which the unit has been classified, with the relative probability. From the output of the model, furthermore, we can see the split variables, the associated class and probabilities and the loss of deviance.

3.2.3 Pros and cons of classification tree

This model is used firstly because of its logical simplicity and interpretation; furthermore, from a computational point of view, it is easy to estimate. An advantage that could be useful for this research is that the algorithm provides itself a selection of the important variables. This is consistent with the aim of the thesis, that is also to understand which were the variables that were influential for the depression and the loneliness in the pandemic. We should, however, be cautious, because it is not simple to evaluate the importance order of the variable in the tree. Other disadvantages are the instability of the results, for example due to changes in existing data, and the difficulty of adding new data without re-starting the process. Finally, it is true that this model provides robust forms of deviance, but statistical inferential procedures are not available.

3.3 Model 3: Random intercept model

We will use the random intercept model to study the particular effect of the variable related to the country and to see if there is heterogeneity between the countries. This type of model belongs to the larger class of multilevel models. They are an adequate instrument to study hierarchical structures: when analysing data structured in groups, hierarchical models are a generalisation of linear regression, where the random intercept or variables can vary between groups. The assumption is that the dataset analysed consists of a hierarchy of different populations. This class of models is characterised by different dimensions: a micro dimension, relative to the individuals, and a macro dimension, that is referred to the group or environment to which the units belong; therefore we have to distinguish between micro-units and macro-units. Units belonging to the same group share the same group-specific influences. The structure consists usually in a single response variable, which is always a first-level variable, and in one or more explanatory variables, that can belong to any level. The first-level variables aim to explain variability at the individual level, while the second-level variables aim to explain variability at the group level (in the case of a two-level model).

3.3.1 Explanation of random intercept model

The random intercept model is so a subgroup of the multi-level models. There is a dependent variable, that in our case will be *depression*, and a set of predictors, measured at the level of individuals. The model is structured in this way (we report an example with only one explanatory variables, but it is extendable to more):

$$y_{ij} = \beta_{0j} + \beta_1 x_{ij} + \epsilon_{ij} \quad , \quad (3.9)$$

where y_{ij}, the dependent variable, has the index i related to individuals and the index j related to the second-level units. The component ϵ_{ij} indicates the errors at an individual level. In our case, the different groups will be represented by the countries of residence of the respondents.

The purpose is to estimate the value of y_{ij}, considering the effect of the explanatory variables both at individual and group level. As mentioned above, the random intercept varies among the groups. The assumption, in fact, is that the effect of the group is captured through the changes in the intercept. The regression coefficients of the explanatory variables are therefore constant in the groups, but the average levels, represented by the intercept, of the predictors are different from group to group.

So, the intercept can be decomposed in this way:

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad , \quad (3.10)$$

where γ_{00} is the intercept of all groups, while u_{0j} represents the random component. This random effect is a measure of the deviation from the average of each group. Since the effect is random, it is a realisation of a casual variable U_0, that has null average and constant variance. This casual variable can be so interpreted as a variable that describes the errors at a level group. It is often assumed that the error terms are independent of each other and with the predictors.

Therefore, the total variance of the response variable, can be decomposed as the sum of the variances at the two levels, individual and of the group. The individual variance is represented by $Var(E_{ij}) = \sigma^2$, while the group variance is given by $Var(U_0) = \tau_0^2$. The total variance is so:

$$Var(Y_{ij}) = Var(E_{ij}) + Var(U_0) = \tau_0^2 + \sigma^2 \quad (3.11)$$

3.3.2 Interpretation

To understand if it is appropriate to estimate a random intercept model, we can use the coefficient of intraclass correlation. The hypothesis of hierarchical models is, in fact, that observations are not independent: the average correlation between individuals belonging to the same group is so called intraclass correlation. It is defined as:

$$\rho = \frac{\tau_0^2}{\tau_0^2 + \sigma^2} \qquad (3.12)$$

This decomposition indicates that it is the ratio of between variance and total variance. It then can be defined as the portion of variability attributable to the groups or, equivalently, as the correlation between two units of the same group[23]. This index can vary between 0 and 1. The interpretation is that the closer the coefficient is to 1, the greater is the contribution due to hierarchical structuring.

3.4 Model 4: Gradient boosting

We then choose to use another predictive model, that is the gradient boosting. Gradient boosting belongs to the category of models that combine results from different models; in this category there are also for example bagging and random forest. In particular, the gradient boosting is a subcategory of the boosting method.

3.4.1 Explanation of gradient boosting

Briefly, the main idea of the boosting is to combine the output of weak classifiers: they are classifiers whose error rates are not much better than errors given by random guessing[22]. The procedure is iterative and consists into assign a different probability of entering the sample to each unit. A greater weight is given to the units that are poorly classified in the early stage: the idea is that in this way we focus more on the subset of data that we cannot classify correctly. At every step of the procedure a new classifier is produced, from the modified weights. At the end of the process all the predictions are combined to create a final prediction, chosen through a weighted majority vote. The formula below shows the final prediction:

$$G(x) = sign(\sum_{m=1}^{M} \alpha_m G_m(x)),$$

where M is the number of iterations and α_m are the weights given to the contribute of every prediction $G_m(x)$. By doing so, the higher importance is given to the classifiers that classify in the most accurate way. The most used procedure to implement the boosting is *Adaboost* algorithm. Furthermore, as base classifier trees are usually chosen.

Gradient Boosting is an implementation of the Boosting algorithm. The models that are used are usually trees and its purpose is to minimise a generic loss function. Gradient Boosting implies a correction in the estimation of the m^{th} tree based on the gradient of the loss function: the idea is in fact to perform a gradient descent on the loss function in several steps. Through the algorithm we want to choose the split that maximises the gradient descent and that so help the approach to the minimum of the loss function. In the case of binary classification the loss function corresponds to the binomial deviance.

In summary, the algorithm function in this way: first, it is necessary to initialise the model at a constant value, then the algorithm proceeds to calculate the negative gradients and fit the first tree using these values as the new response. The algorithm then proceeds in an iterative manner, alternating between the calculation of negative gradients and the fitting of the subsequent tree based on these values. At each step the predictions are updated, with the addition of the response generated by the current step tree to the previous step value; the algorithm stops when the predetermined number of iterations M is reached.

Gradient boosting as described here is implemented in the R gbm package, that will be used for fitting the model.

3.4.2 Interpretation of gradient boosting

Gradient boosting was used in this research, because it may provide useful information on the importance of variables. In particular, in Chapter 4 will be reported a graph which shows the relative importance of variables used as predictors.

The measures are based on the number of times a variable is selected for splitting, weighted by the squared improvement to the model as a result of each split, and averaged over all trees[24]. Variable importance is determined by calculating the relative influence of each variable: whether that variable was selected to split on during the tree building process, and how much the squared error (over all trees) improved (decreased) as a result. The relative influence (or contribution) of each variable is scaled so that the sum adds to 100, with higher numbers indicating stronger influence on the response.

3.4.3 Pros and cons of gradient boosting

As all the models, also gradient boosting has advantages and disadvantages. Among the firsts, we have to mention the predictive accuracy of this model. Fur-

thermore, gradient boosting allows a lot of flexibility: it can optimise various loss functions and it is used both for regression, binary classification and multi classification.
On the other side, it presents also some disadvantages. This model in fact is not exempt from the problem of over-fitting: the GBM algorithm continues to minimise the errors; to overcome this problem we will divide the main dataset, as done for the other models. Furthermore, it is expensive from a computational point of view, because it has to estimate often many trees and it has many parameters.

3.5 How to compare models accuracy

To understand which model has the better prediction accuracy, we need to compare the results. The accuracy for all model is measured on the test set, that is a portion of data not used for the estimation. The new units will then be allocated on the basis of the results of the estimate of the model. It will be used a cross-table in which we count the number of correctly or incorrectly predicted cases, for each of the two levels. This table is called confusion matrix or misclassification table. The predicted cases are based on the probability assigned by the model: usually, if the probability is greater than $\frac{1}{2}$, the unit will be allocated to one class, otherwise to the other. As our data are not balanced, we will not consider as threshold $\frac{1}{2}$, but the relative frequency of "positive" answers in the training set. From this table we can obtain different indexes, useful to compare the results. First, we can calculate the misclassification error, that is given by the fraction of cases correctly classified on the total. We can then consider false positives and false negatives. Respectively, we can estimate the probability of false positives with the fraction of cases predicted positive but actually negative on the total of actual negative cases and call it $\hat{\alpha}$, while the probability of false negative is called $\hat{\beta}$ and is given by the fraction of cases predicted negative but actually positive.

Chapter 4

Model estimation and results

In the following chapter the main results of this research will be shown for both dependent variables. We recall that our main purposes are to understand the main drivers of experiencing depression or loneliness during the pandemic and identify the vulnerable groups, based on the answers to the SHARE questionnaire. We are more interested in the interpretation of the variables than in the accuracy of the prediction.
As for Chapter 2, we will first show the result for depression and then we will look at loneliness.

4.1 Results for dependent variable *depression*

Initially we will use the main data set, whose construction and composition has been illustrated in Chapter 2. We recall how the dependent variable is constructed: it takes the value 1 if the respondents state that they are more depressed or sadder than before the pandemic and 0 otherwise. The data set comes from merging the data set of the questionnaire of SHARE wave 8 and the one of SHARE wave 7. It is composed of 48425 observations. From a brief summary of the variables, it is possible to see that some of them contain a really high number of missing data. It is the case of variables like "drugs" and the one referred to the household income. We decide to firstly delete these two variables, to avoid having to delete important observations, only due to these two variables.
Firstly, we randomly divide the data set in two subsets: the training set, used to fit the different models, and the test set, which is instead used to evaluate the performance of these ones. We assign 80% of the observations to the training set (composed in fact of 38740 observations) and the remaining 20% to the test set, that has so 9685 observations. This operation is possible because of the large number of observations and it is appropriate to circumvent the plausible problem of

over-fitting. This problem often occurs when the number of observations is high and it consists in the fact that the model will adapt too well to the specific data, maybe fitting some unnecessary features, that will not necessarily reoccur in another sample of data of the same phenomenon.

4.1.1 Logistic regression model

The first model that has been adapted to the data is the logistic regression model, with the set of independent variables listed in Table 4.1 :

Variables for logistic regression model		
gender	usual home	health before Covid 19
health improvement	major illness	falling
left home	wash hands	hands sanitizer
drugs vs Covid 19	symptoms inside household	positive case inside household
hospitalised person inside household	dead person of the household	forgo treatment
postponement of medical appointment	negative response med appointment	satisfaction with hospital treatment
satisfaction with medical office	employment	reduce working hours
working location	safety workplace	ends meet
physical contact with child	online contact with child	physical contact with friends
life satisfaction wave 7	life happiness wave 7	openness
agreeableness	neuroticism	extra-version
consciousness	country	household size

Table 4.1: Variables for logistic regression for *depression*

Three variables were automatically eliminated due to singularities: this is the case of *negative response to a medical appointment*, *reduce working hours* and *safety of workplace*. In the estimation of the models also some observations with missing data were automatically deleted.

As explained in Chapter 3, we can interpret the estimates of the coefficients as odds ratios. We can therefore observe some interesting results: we consider the effect of the single variables, net of other explanatory variables. In Tables 4.2, 4.3, 4.4 we report the estimates of the coefficients and their relative level of significance. The significance is represented by asterisks. Three asterisks indicate that the estimate is significant at a level of 0.1%, two at 1%, one at 5% and a point is a significance level of 10%. If none of these signs is present, it means that the estimate is not statistically significant. This notation will be used for all the tables that report estimates of coefficients.

For what concerns the variable *gender*, we can say that the odds for females are about 67% higher than the odds for males. This is a really relevant result and for this reason in the section 4.1.5 we will consider it more deeply, choosing to adapt to the data different models by gender.
We first start by studying the effect of the variables affecting the health of the subjects. Firstly, the status of health before the pandemic is important: in fact, if health before Covid-19 was worse, the likelihood of being more depressed after it increases. This would therefore lead one to think that the pandemic hit those who were already physically fragile the hardest from a mental health point of view. This is a result that might be expected as a person who already had physical problems before the pandemic has to add the worries of the pandemic and the lack of health care. Analysing results in detail, for those who considered their physical health to be poor before Covid-19 the odds of being more depressed are about 84% higher than for those with very good physical health. Scaled up, for those who considered their physical health to be fair the odds of being more depressed are 50% higher, for those who considered their physical health to be good the odds of being more depressed are 25% higher, again taking very good physical health as the baseline. We can see, therefore, that this variable not only has an obviously important effect, but also that the differences from one level to another of it are considerable.
Secondly, we study the effect of *falling*; we can see that for those who have had at least one episode of falling or dizziness the odds of being more depressed are 1.99 times as large as the odds for those who have had zero episodes being more depressed. Again, we see the importance of physical health, which plays a key role during the pandemic and which, as can be expected, closely correlates with mental health.
Finally, the variable *major illness*, which indicates whether the subject has been diagnosed with one or more new diseases, does not appear to be statistically significant, contrary to what might be expected; intuitively, in fact, being diagnosed with a new disease during a pandemic in which the mortality rate appears to be increasing with co-morbidity might be further cause for concern and depression. In this case, although, the model does not show this effect.
We consider then behaviour variables; the estimates of the coefficients are -0.27 and -0.17 for the change in the frequency of washing hands and of using a hand sanitizer (the reference level is "more frequently than before"), respectively. Therefore, we can assert that those who do not increase hand washing or gel use with respect to before the pandemic are less likely to become ill with depression. This

Estimates of coefficients for logistic regression model		
Variables	**Estimates**	**significance**
intercept	-2.5	***
gender	0.51	***
usual home	0.104	
health before Covid 19 very good	0.15	
health before Covid 19 good	0.24	**
health before Covid 19 fair	0.41	***
health before Covid 19 poor	0.61	***
health improvement worsened	1.21	***
health improvement same	0.06	
only one major illness	0.09	
two or more major illnesses	0.07	
falling	0.69	***
not having left home	0.03	
wash hands same as before	-0.27	***
hands sanitizer same as before	-0.17	***
drugs vs Covid 19	-0.10	
symptoms inside household	0.17	*
positive case inside household	0.04	
hospitalised person inside household	0.24	
dead person of the household	1.04	*
not having forgone treatment	-0.39	***
postponement of medical appointment	-0.11	**
satisfaction with hospital treatment 2 (somewhat satisfied)	0.05	
satisfaction with hospital treatment 3 (somewhat dissatisfied)	0.39	
satisfaction with hospital treatment 4 (very dissatisfied)	0.46	
satisfaction with hospital treatment 5 (was not treated in hospital)	- 0.16	*

Table 4.2: Estimates of logistic regression coefficients for *depression* part 1

result might make sense when linking the fact of washing hands more or using a sanitising gel more to an increased anxiety about the pandemic situation.

Then, we want to find out which is the effect of the most characteristic variables of the Covid pandemic, that are referred to the presence inside the household of symptomatic, positive, hospitalised or dead person. We can interpret the estimates in this way: for those who have a person with symptoms inside the household the odds of being more depressed are 1.18 times as large as the odds for those who do not. For the case of death, the odds for those who have had a bereavement in the household due to Covid-19 are 2.82 times as large as the odds for those who have had no deaths inside the household. Variables related to hospitalisation and positive cases inside the household do not appear to be statistically significant, but this could be due to the limited number of respondents who reported having a positive case or hospitalised person at home. For this reason, we report anyway a possible interpretation, that suggests that the odds of being more depressed for those who have a positive case or an hospitalised person of the household are higher than the odds for those that are not in this situation.
One of the main consequences of the pandemic was the pressure on hospitals, which led to the temporary interruption of many visits and medical operations. This situation could have had an impact on the mental health of people, that could have felt neglected and ignored: in our study, both the variables respectively related to having forgone a treatment and have had to postpone a medical appointment result statistically significant. In particular, those who did not forgo treatment compared to those who did are 32% less likely to be depressed. Those who did not postpone a medical appointment are 11% less likely to be depressed. Also, the satisfaction with the hospital treatment or with the medical treatment could influence the depression of the subjects: we might expect that if a person does not have to wait for hours inside a hospital (where the probability of getting the coronavirus could be slightly higher) or receives satisfactory treatment despite the ongoing pandemic, he or she will tend to be less depressed than who does not consider himself satisfied with the treatment received. From the results of the logistic regression model, it is inferred that for those who have a lower satisfaction with treatment received at the hospital, the odds of being more depressed are higher than the odds for who has an higher level of satisfaction. The same holds true for satisfaction with the medical treatment received at the doctor's office, so we can assume that our initial hypotheses are confirmed.

Estimates of coefficients for logistic regression model		
Variables	**Estimates**	**Significance**
somewhat satisfied with medical office	0.13	
somewhat dissatisfied with medical office	0.30	.
very dissatisfied with medical office	0.27	
did not go to medical office with medical office (5)	-0.08	.
having become unemployed	-0.24	*
working location usual	- 0.16	
working location both	0.08	
working location none	0.097	
working location not employed	-0.03	
ends meet with some difficulty	- 0.25	***
ends meet fairly easily	- 0.42	***
ends meet easily	-0.51	***
not having received financial support	-0.08	
physical contact with child at least once a week	0.16	**
physical contact with child at least never or almost never	0.28	***
online contact with child at least once a week	-0.18	***
online contact with child never or almost never	- 0.32	***
physical contact with friends at least once a week	0.026	
physical contact with friends never or almost never	0.08	
online contact with friends at least once a week	-0.08	
online contact with friends never or almost never	-0.12	.
life satisfaction wave 7 (levels from 1 to 10)	0.55, -0.07, 0.15, 0.31, 0.12, 0.24, 0.11, -0.04, -0.14, -0.12	
life happiness wave 7 level 2	0.03	
life happiness wave 7 level 3	0.03	
life happiness wave 7 level 4	-0.01	

Table 4.3: Estimates of logistic regression coefficients for *depression* part 2

For what concerns answers related to work, from the estimates of coefficients it seems that the odds pf being more depressed for those who became unemployed due to the coronavirus are 20% higher than the odds for those who did not lose their job: the work component appears to have an effect on the mental health of the subjects. In fact, the pandemic has affected all areas, including the labour market. Indeed, the labour market has not only faced a severe crisis in which many people have lost their jobs or had to stop working, but has also had to change its traditional patterns. The fact that the virus was transmitted from person to person meant that the way of working had to be adapted to the new situation; smart working, that is working from home, was introduced.
How did this fact influence people's mental health? Actually, from the models appears that the coefficient on the variable "workplace" is not significant, so we can assume that this was not one of the most decisive variables for the depression.
Also the economic situation is connected with depression, and it is possible to see from variable *ends meet* that for those who have less difficulty making ends meet, the odds of being more depressed are lower than the odds for those who have difficulty.
Another issue addressed in the questionnaire was the frequency of social contacts. Social networks have in fact played a fundamental role during the pandemic, because they had to develop: physical contacts were in fact very reduced. Both relationships with own children and friends or relatives outside the household changed and influenced the mental health of people. For those who have physical contact with their children at least once a week, but not daily, the odds of being more depressed are 1.17 times as large as the odds for those who continue to see them daily compared to before the period before Covid 19. Then, for those who see them never or almost never the odds of being more depressed are 1.32 times as large as the odds for those who see them daily. On the other hand, with regard to contact via telephone or other means of communication, the relationship seems to be inverse. The odds of being depressed for those who have a more frequent contact, in fact, are higher than the odds for those who have less frequent contact. The same results apply to friends as well, although the odds differ less and the effect is not considered significant. This result is quite interesting, as one might think that the more a person was able to stay in contact even virtually with their children, the less likely they were to be more depressed. However, this is refuted by the model; the reason for this is unknown, but it could be probably due to the lack of physical contact. Continuing in the field of social relations, we consider the dimension of the household (this variable has a range from 1 to 12 in our data set). The variable related to this appears statistically significant and it suggests that the odds of being more depressed than before the Covid-19 are about 7% lower for those who have one more unit in the household than for the others. It will be

interesting to compare this effect with the one for loneliness.

Estimates of coefficients for logistic regression model		
Variables	**Estimates**	**significance**
openness	0.034	.
agreeableness	0.012	
neuroticism	0.174	***
extra-version	-0.04	.
consciousness	0.07	**
country DE	0.33	**
country SE	0.47	***
country NL	0.55	**
country ES	1.12	***
country IT	1.13	***
country FR	0.45	***
country DK	0.16	
country GR	-0.06	
country CH	0.58	***
country BE	0.72	***
country IL	0.13	
country CZ	-0.18	
country PL	0.29	*
country LU	0.88	***
country HU	0.14	
country PT	0.72	***
country SI	-0.09	
country EE	0.28	**
country HR	0.08	
country LT	0.23	.
country BU	0.004	
country CY	0.29	.
country FI	0.27	.
country LV	-0.57	***
country MT	0.99	***
country RO	0.27	*
country SK	0.001	
household size	-0.07	***

Table 4.4: Estimates of coefficients for logistic regression for *depression* part 3

Arriving therefore to the variables considered initially from the previous wave to that of the Covid, life satisfaction and happiness do not appear statistically significant, while for what concerns the variables of personality there are some interesting results. Starting with openness, it seems that for those who have one more point in openness to others, the odds of being more depressed are about 3% higher than the odds for those who have one less point. The pandemic in fact for sure did not help those who were instinctively inclined to open up to others: for those who have one more point in it, the odds of being more depressed are about 1% higher than the odds for those who have one point less. For what concerns neuroticism, it is possible to see that for those who have one point more in emotional instability the odds of being more depressed are about 18% higher than the odds for those who have one less point. It is a very large effect, but it seems also reasonable because as said in Section 2.2, neuroticism includes also anxiety, that was one of the main issues of the pandemic. Finally, for those who have one more point in extra-version the odds of being more depressed are 3% lower than the odds for those who have one less point, while for those who have one more point in consciousness the odds of being more depressed are 7% higher than the odds for those who have one less point.

It remains to be considered only one variable, that is the country of residence of the interviewed: it is a factorial variable with 27 levels; many of the coefficients result statistically significant, so we can hypothesise that levels of depression differ from one country to another. To understand better this phenomenon, we will compare in Section 4.1.6 the depression in the different countries, depending on the stringency index.

The total error of prediction of the model has been calculated setting the threshold at 0.16, that is the relative frequency of the class "yes" of the dependent variable. The total error of classification is around 0.30 and it is calculated as the ratio of incorrectly predicted observations to total observations. False positive and false negative rates are also around a value of 0.30. This values will be compared to the ones of the subsequent models, in order to see which model has the best predictive ability.

4.1.2 Classification tree

The second step was to fit another type of model to the data, keeping the same explanatory variables, to see how the results could eventually change.

Therefore, we resorted to the classification tree, which, as explained in Chapter 3, has characteristics that facilitate a simple interpretation of the effect of the variables.

The estimation set was divided into two further subsets; one was used for the tree growth phase, the other for the pruning phase. In the growth phase, the model was adapted to the data: we used the deviance of the binomial distribution as function to minimise. After both these phases, we arrive to a tree with 8 leaves, as shown in Figure 4.1.

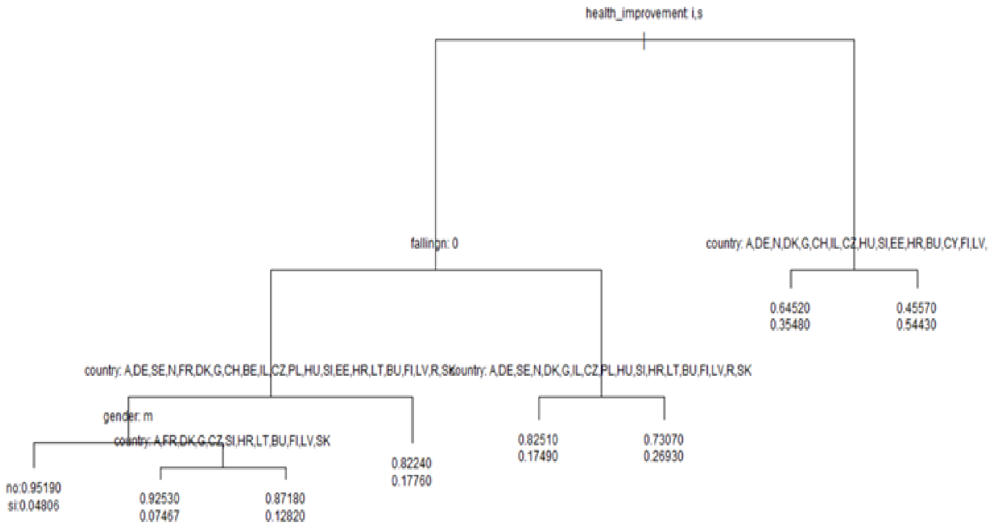

health_improvement: i,s
fallingn: 0
country: A,DE,N,DK,G,CH,IL,CZ,HU,SI,EE,HR,BU,CY,FI,LV,
0.64520
0.35480
0.45570
0.54430
country: A,DE,SE,N,FR,DK,G,CH,BE,IL,CZ,PL,HU,SI,EE,HR,LT,BU,FI,LV,R,SK
country: A,DE,SE,N,DK,G,IL,CZ,PL,HU,SI,HR,LT,BU,FI,LV,R,SK
gender: m
country: A,FR,DK,G,CZ,SI,HR,LT,BU,FI,LV,SK
0.82240
0.17760
0.82510
0.17490
0.73070
0.26930
no:0.95190
si:0.04806
0.92530
0.07467
0.87180
0.12820

Figure 4.1: Classification tree for *depression*

This tree demonstrates the importance of the variables *health improvement* and *falling*, as seen in the logistic regression model; also *country* and *gender* appear as split variables, therefore they can be considered relevant. In every leaf, the class indicator is shown (in this case "No" or "Yes", that corresponds to the classes of the dependent variable) and the probability associated to that. Precisely, starting from the root node, that is in this case related to *health improvement*, it is possible to follow the different paths, that originate different subsets. Once reached the leaf node, that is the one located at the end, it appears the predicted outcome with the respective probability.

For example, from the Figure 4.1 we can give an interpretation of the variable that determines the first split, that is *health improvement*. On the right, there are the cases when the health worsened from before the pandemic and it is possible to see that the probabilities of not being more depressed are lower than on the other side of the tree (we are referring to values as 0.64 and 0.45, compared to 0.95, 0.92, 0.87, 0.822, 0.825, 0.737). To understand from the classification tree graph alone the importance of a certain feature we can look at the length of the tree branches:

the longer the branches are, the greater is the drop in deviance determined by that variable.
The total classification error of this model, always considering the threshold 0.16 as before, is higher than the one of the logistic regression model: it is, in fact, around 0.44. The problem seems to be in the rate of false positives, that is near to 0.5, while the rate of false negatives is lower than before, that is 0.21.
So, considering all this factors, from a predictive point of view, the logistic regression model seems to be better than this one.

4.1.3 Gradient boosting

Another method that can help us to understand which are the most important variables is the gradient boosting. We choose to adapt it to the data set, maintaining the subdivision between training set and test set. The set of variables used as explanatory variables is the one found through the logistic regression models. To apply the model it is necessary to transform the dependent variable in type numeric. It is then used the gradient boosting model, in which the distribution of Bernoulli has been specified (typically used for dichotomous variables); it has been chosen a number of trees of 5000. With the gradient boosting it is possible to see the relative importance and influence of variables. The relative importance of each variable is shown in Figure 4.2. Since these measures are relative, it is customary to assign the largest a value of 100 and then scale the others accordingly[22]. The variable that appears to have the most relative influence is *country*, and its value is 58%, while it appears that the less relevant variable is the one related to have received a financial support.

The total error of the gradient boosting model is 0.30, and so are the rate of false negatives and false positives. The precision rate is around 0.30, while the recall rate (it is the ratio of correct predictions for a class to the total number of cases in which it actually occurs) is 0.69. Another possible measure of accuracy is the F1 score, that is the harmonic mean of the model's precision and recall; in this case it is 0.43. Also in this case, we will compare these errors with the other models' errors.

4.1.4 Modifications of previous models

It could be reasonable to modify the initial set of considered variables, looking both at the statistical significance and at the meaning of each one. For this reason, we decide to delete *life happiness* and *life satisfaction*, that do not appear significant. Furthermore, we include two variables that before were excluded because could be seen as leakers of the dependent variable, to see how the results change. The new estimated logistic regression model gives results in agreement with the

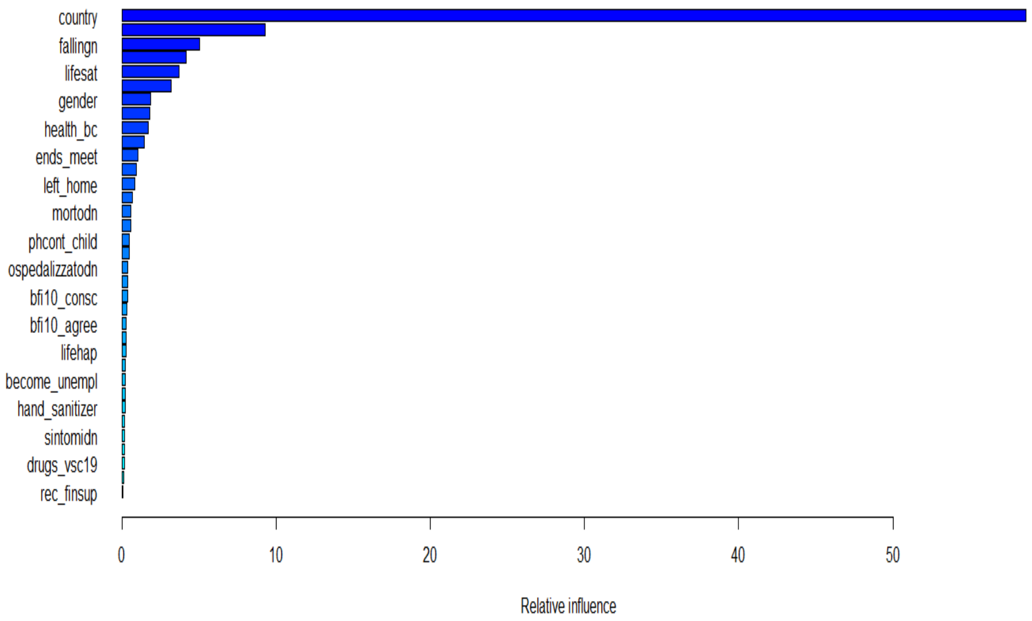


Figure 4.2: Gradient boosting for *depression*

previous one: the estimates of the coefficients give indications similar to those commented above. As regards the two added variables, they are both statistically significant at a significance level of 5%. Specifically, for those who reported feeling more nervous than before the pandemic the odds of being more depressed are nine times as large as the odds for those who did not feel more nervous. Also with respect to having trouble sleeping the result is important: for those who in fact have more problems sleeping the odds of being more depressed are almost four times higher than the odds for those who do not make this statement.

For what concerns the errors, we can see that the total error is now around 17%, so 13 percentage points lower; also the rates of false positives and false negatives are slightly lower (respectively of 0.16 and 0.22). If we adapt a classification tree with the same variables, the situation in comparison with the previous tree changes: in fact, the new variables *nervous* and *trouble sleeping* are at the first nodes of the tree: they seem the most important ones in influencing the drop in deviance.

There are some variables that we did not consider in the first analysis, because of the large amount of missing data. Deleting observations that had missing values in those variables would have led to a significant reduction in sample size and even to the elimination of almost all observations related to the Netherlands. However, it remains of interest to understand how these variables are related to the dependent variable.

In the data set without the Netherlands, in fact, the values of *years of education* and *ISCED 1997* which are missing are 7, while in the specific data set of the Netherlands the percentage of missing values for this variables is 45%. For this reason, we consider a subset of our initial data set, without the observations of Netherlands. Since both the variables measure the same phenomenon, that is education, it is considered to keep only one of the two variables, precisely *ISCED 1997*.

We estimate a logistic regression model and interpret the estimate as it has been done before. The estimate of the coefficient is equal to -0.031725 and it is statistically significant at a level of 10%, so we can say that for every level of education, the odds of being more depressed are about 3% lower than the odds for the lower level.

We want then to consider the effect of the variable *number of children*; as mentioned before, in fact, the sociability really changed due to restrictions against Coronavirus, so intuitively also the number of children or in general the presence of one or more children could have an effect on depression. This variable presents 10353 missing values in the whole data set, so about 21% of the total. After making a subset with only those observations that do not have a missing value for this variable, we estimate a logistic regression model. As the number of children increases, the odds of being depressed are 1% lower but this estimates does not result significant. For this reason, it is reasonable to create a dummy variable that distinguishes between not having children and having at least one child: the estimate appears significant at a level of 10% and it suggests that for those who have at least one child the odds of being more depressed are 55% higher than the odds for those who do not have them. This could result from the fact that the physical contact with own children is really reduced, and this result is consistent with what found previously about the variable *physical contact with own children* in the first logistic regression model.

4.1.5 Different models by gender

As seen in Section 4.1.1, the prevalence of depression among females and males is different. For this reason, we decide to treat separately these two classes and to estimate the logistic regression model, in order to see how the estimates of coefficients change and if for each group there are some more relevant factors than for the others.

Firstly, we adapt to the data of men the logistic regression model; the total observations of males are 20311 and, as done before, we randomly divide the data set in two subsets, one used for the estimation of models (which contains the 80%

of the observations), the other for the validation. Then we do the same operation with the subset composed only of females observations (the training set has 22941 observations and the test set 5623). Analysing the estimates of coefficients and their statistical significance from Tables 4.5, 4.6 and 4.7, we can firstly see some differences in the significance of the variables. For the group of males, the variable *major illness* appears significant at a significance level of 1%. For those who have been diagnosed with 2 or more serious diseases the odds of being more depressed are 51% higher than the odds for those who have not been diagnosed with a new disease. For those who have discovered they have only one new disease, the odds of being more depressed than before the pandemic are about 11% higher than the odds for the reference group, but this coefficient does not appear significant. For the group of females, instead, the variable is not at all significant. Remaining in the health area, the coefficients related to levels "good" and "very good" of *health before Covid-19* are significant for males and not significant for females, while the other levels are significant for both the groups.

Other differences in the presence of significance appear for the variables *drugs against Covid 19*, *hospitalisation of the household member*, *death of household member*, *satisfaction with received treatments*, *physical contact with friends*, *on-line contact with friends*, *extra-version* and *consciousness*. For what concerns the fact of taking some drugs for Covid-19, it seems that the relative variable is significant only for the group of females and that in particular the odds of being more depressed for those who take some drugs against Covid-19 are 17% higher than the odds for those who do not take some drugs of that type. With regard to the variables of satisfaction with hospital or medical office treatment, we notice some differences. Males who did not go to the hospital are 23% less likely to be depressed in comparison with who went and was satisfied, while for females the values are not significant; for medical office it is true the same but for females (the percentage drops to 10%). With regard to the variables of hospitalisation and death of household members, the first is significant only for males, while the second only for the group of women. However, this is a result that must be taken with caution, as the size of the different classes in these questions is very unbalanced.

We can then notice an interesting result for what concerns the contacts with children and friends. The absolute values for the variable related to physical contact with children are slightly lower for the group of females.

Effect of variables by gender				
Variables	**Estimates for males**	**Significance for males**	**Estimates for females**	**Significance for females**
intercept	-2.47	***	-1.76	***
health before Covid 19 very good	0.48	*	0.02	
health before Covid 19 good	0.63	**	0.15	
health before Covid 19 fair	0.94	***	0.32	**
health before Covid 19 poor	1.28	***	0.49	***
health improvement worsened	1.09	***	1.29	***
health improvement same	0.007		0.05	
only one major illness	0.11		0.10	
two or more major illnesses	0.41	***	-0.02	
falling	0.66	***	0.67	***
wash hands same as before	-0.31	**	-0.30	***
hands sanitizer same as before	-0.18	*	-0.19	**
drugs vs Covid 19	-0.13		-0.18	*
symptoms inside household	0.40	**	0.15	.
positive case inside household	0.01		-0.06	
hospitalised person inside household	0.81	.	0.20	
dead person of the household	1.37		1.22	*
not having forgone treatment	-0.48	***	-0.40	***
postponement of medical appointment	-0.17	**	-0.10	**

Table 4.5: Estimates of logistic regression coefficients by gender for *depression* part 1

Effect of variables by gender				
Variables	**Estimates for males**	**Significance for males**	**Estimates for females**	**Significance for females**
somewhat satisfied with hospital treatment	-0.05		0.04	
somewhat dissatisfied with hospital treatment	0.34		0.54	.
very dissatisfied with hospital treatment	0.13		0.46	
did not receive hospital treatment	- 0.26	*	-0.03	
somewhat satisfied with medical office	0.36	**	0.03	
somewhat dissatisfied with medical office	0.30		0.27	
very dissatisfied with medical office	-0.01		0.16	
did not go to medical office	0.05		-0.10	*
having become unemployed	-0.51	***	-0.24	*
ends meet with some difficulty	- 0.25	*	-0.37	***
ends meet fairly easily	- 0.58	***	-0.48	***
ends meet easily	-0.56	***	-0.59	***
physical contact with child at least once a week	0.31	**	0.11	.
physical contact with child at least never or almost never	0.32	***	0.28	***
online contact with child at least once a week	-0.20	**	-0.20	***
online contact with child never or almost never	- 0.13		-0.35	***
physical contact with friends at least once a week	0.026		0.20	*
physical contact with friends never or almost never	0.08		0.25	**
online contact with friends at least once a week	-0.08		-0.12	
online contact with friends never or almost never	-0.12	.	-0.21	**

Table 4.6: Estimates of logistic regression coefficients by gender for *depression* part 2

Effect of variables by gender				
Variables	**Estimates for males**	**Significance for males**	**Estimates for females**	**Significance for females**
openness	0.015		0.01	
agreeableness	0.0015		0.03	
neuroticism	0.20	***	0.19	***
extra-version	-0.05		-0.06	**
consciousness	0.07	.	0.04	
country DE	0.37		0.28	*
country SE	0.68	**	0.25	
country NL	0.72	*	0.50	*
country ES	1.24	***	1.12	***
country IT	1.24	***	1.13	***
country FR	0.35		0.51	***
country DK	0.30		0.18	
country GR	0.12		-0.10	
country CH	0.50	*	0.55	***
country BE	0.76	***	0.79	***
country IL	0.13		0.07	
country CZ	-0.14		-0.17	
country PL	0.31		0.37	**
country LU	1.03	***	0.96	***
country HU	0.22		0.02	
country PT	0.80	***	0.81	***
country SI	0.20		-0.08	
country EE	0.43	*	0.25	*
country HR	0.40	.	0.05	
country LT	0.76	**	0.10	
country BU	0.13		0.17	
country CY	0.38		0.18	
country FI	0.09		0.30	*
country LV	0.03		-0.56	**
country MT	1.17	***	0.85	***
country RO	0.60	*	0.34	*
country SK	0.59	*	-0.03	
household size	-0.08	*	-0.09	***

Table 4.7: Estimates of logistic regression coefficients by gender for *depression* part 3

While in the male group there appears to be almost the same difference between seeing children never or almost never and at least once a week compared to a daily frequency (for both groups the odds of being more depressed than before Covid-19 are about 37% higher than the odds for those who see children daily), for females these values change. For them, in fact, the odds of being more depressed than before Covid-19 for who sees their children never or almost never are about 32% higher than the odds for those who see children daily, while for those who see their children at least once a week are about 11% higher. For what concerns the online contact for females the absolute value is higher than for males only for the level "have online contact never or almost never", so we could assume that for mothers the real burden lies in not contacting their children at all, rather than in the frequency of contact. Another argument applies to contacts with friends, both physical and not. These variables in fact seem to be statistically significant only for females; the direction of the estimates of coefficients seems consistent with what found earlier: the less frequently you see friends physically, the more likely you are to be more depressed than before Covid, vice versa for non-physical contact.
Finally, we can see some differences in the two groups also for what concerns the variables related to the personality. Firstly, the trait of personality of neuroticism appears statistically significant at a level of 1% for both groups; extra-version is significant only for females (as a point increases in extroversion the odds of being more depressed than before are about 6% lower), while consciousness only for males (as a point increases in consciousness the odds of being more depressed than before increases by about 6%).
The results just described are the main ones in terms of significance and differences between the two groups. Differences can then be considered at the absolute value level of the effects; in particular, for what concerns *symptoms inside household*, the effect seems greater for males, as well as for *having become unemployed*. For all the other variables the coefficients do not differ much between the two groups. However, to confirm that it is actually worth performing two different models depending on the gender, we can conduct the Chow test. We do that by controlling if the null hypothesis that the estimated parameters in the two different models are equal. We find a value of test F equal to 10.62, that is really high, and if compared with the F distribution with the appropriate degrees of freedom leads us to reject the null hypothesis; so it is appropriate to generate two different models, dividing by gender.

4.1.6 Differences by country

As previously anticipated, we want to understand the effect of depression between different countries.

Stringency index

Initially it could be helpful to study the different situations in the 28 considered countries. To do this, we can make use of an index built by researchers of Oxford and called the "stringency index" [25].
It consists in an average of individual indicators, that are measured for different countries every day in a given time. In the specific case nine individual indicators are considered, eight referred to closure and containment, while one to a health measure. All the individual indicators have a score between 0 and 100. The first eight include school and workplace closure, cancellation of public events, restrictions on gatherings, closure of public transport, stay at home requirements and restrictions on internal and international movement. The health indicator is instead referred to public information campaign. These scores are then averaged to get the composite index. We consider the time between 01-02-2020 and 30-09-2020 and see how this index varies among countries. From Figure 4.3 it appears that there were some differences, both in the intensity of the measures taken against Covid-19 and in the timing of onset and trend in general.

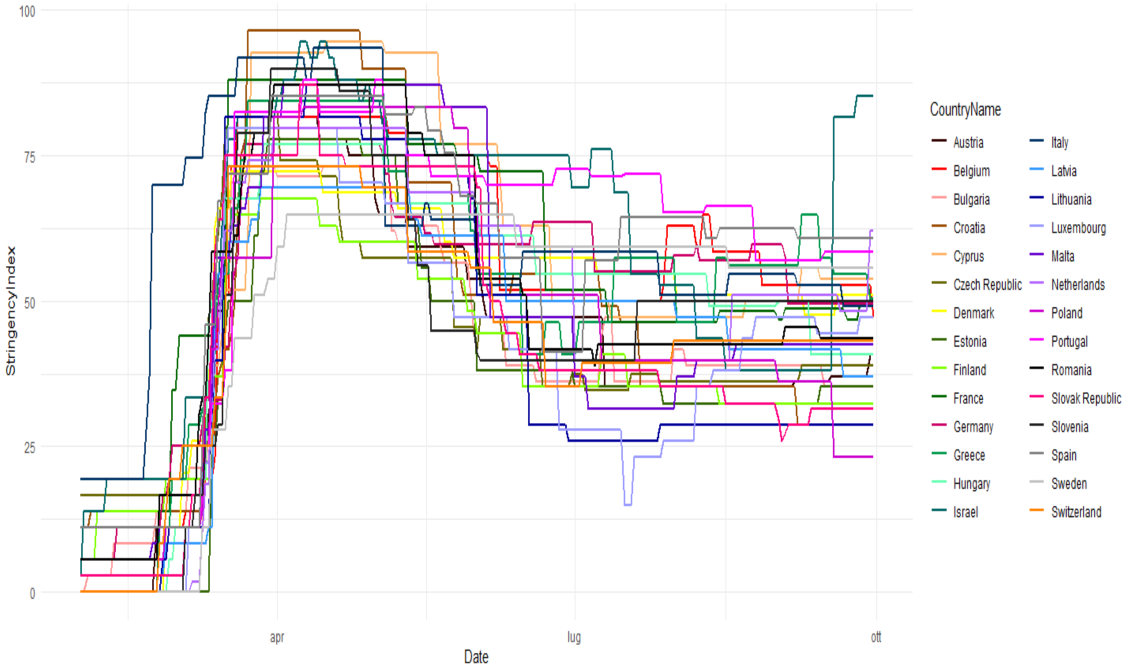


Figure 4.3: Stringency index by country

If we consider two countries, for example Italy and Sweden, it is possible to see that the mean of the value of the stringency index is respectively 61.88 and 49.12; in a 0 to 100 range 12 points are quite relevant. If we then consider the relative frequencies of people that declare themselves more depressed than before the pandemic, for Italy the frequency is around 27%, while for Sweden it is 13%.

Random intercept model

Therefore, taking these considerations as a starting point, we can assume correct to estimate a model that takes into account a possible heterogeneity between groups. This model can be used to assess the differences that exist between groups of statistical units belonging to different countries[26].
We consider the whole dataset and the explanatory variables used before: for country we decided to put a random effect, while for the other variables was estimated only the fixed effect. The variance of the random intercept is equal to 0.1381 and the standard deviation is around 0.37. Using the value of the variance of random intercept we can obtain a measure that can help to interpret this result. It is the intra-class correlation coefficient, which is given by the ratio of the variance between groups (countries in this case) and the total variance. It can take values between 0 and 1 and in this case it is equal to 0.12. This coefficient is used as a measure of the relationship that exists between the values of the depression within each group. The more the coefficient is near to 1, the higher is the proportion of variance between groups on the total variance. The value of 0.12 confirms the presence of heterogeneity between countries. The estimates of the coefficients of the explanatory variables were really similar to the logistic regression model considered in Section 4.1.1 and consequently also the interpretation.

Logistic regression model for country

However, to understand the difference in depression in different countries, we can once again use the logistic regression model and compare the effect with the stringency index of the country.

From the estimates of coefficients reported in Table 4.8 we can interpret the results as odds ratios. The country of reference in the logistic model is Austria, that has one of the lowest values of stringency index.
Besides the fact that it is evident that the coefficients for the different countries differ one from each other (and most of them result significant at a statistical level of 1%), we want to understand if there is a kind of correspondence between the stringency indexes and the coefficients. We can notice that Italy has the highest value of stringency index and also one of the highest absolute values of the coefficient: for Italian people the odds of being more depressed than before are 3.09 times as large as the odds for Austrian people. Also for the Spain we can do the same considerations, although a level of stringency index slightly lower. This corre-

Effect of variables			
Country	**Average stringency index**	**Estimate of coefficient**	**Significance**
Austria	44.30	reference level	
Belgium	55.04	0.72	***
Bulgaria	43.06	0.004	
Croatia	49.64	0.08	
Cyprus	55.16	0.29	.
Czech Republic	43.75	-0.18	
Denmark	50.47	0.16	
Estonia	38.97	0.28	**
Finland	40.16	0.27	.
France	56.70	0.45	***
Germany	53.51	0.33	**
Greece	53.1	-0.06	
Hungary	50.44	0.14	
Israel	60.34	0.13	
Italy	61.88	1.13	***
Latvia	46.05	-0.57	***
Lithuania	41.70	0.23	.
Luxembourg	41.98	0.88	***
Malta	49.47	0.99	***
Netherlands	49.28	0.55	**
Poland	47.43	0.29	*
Portugal	60.21	0.72	***
Romania	50.75	0.27	*
Slovak Republic	44.06	0.001	
Slovenia	47.19	-0.09	
Spain	58.28	1.12	***
Sweden	49.11	0.47	***
Switzerland	44.73	0.58	***

Table 4.8: Stringency index and coefficients for *depression* by country

spondence is not always followed, because for example Israel has a high stringency index, but the coefficient is only equal to 0.13 (the policies implemented by this country in terms of vaccinations could also come into play).

4.2 Results for dependent variable *loneliness*

We now analyse the relation between the explanatory variables and the other dependent variable, that is *loneliness*. We recall that the above variable is a dichotomous variable that measures whether or not a subject feels more alone after the Covid-19 pandemic outbreak than before.
The total of the observations amounts at 48425, and as done for the variable depression we divide the dataset in two subsets, the training set and the test set. We expect that the set of variables whose effect will be significant will be slightly different from that found for depression because the two variables measure two different features, but in a first moment we consider the same set, with the addition of the variable *partner in household*, that indicates if there is the partner in the household. We make this addition because we imagine that this could really be an important factor for feeling more alone, more than for depression, because it concerns specifically the fact of being physical alone. For what concerns other variables, intuitively, we can expect that a fundamental role in influencing the loneliness of people is played by the variables related to social contacts. Also the variables of personality could differ in significance compared to those significant for depression, because depression and loneliness could be influenced by different traits of personality.

4.2.1 Logistic regression model

We first adapt a logistic regression model to the data, which uses the same explanatory set of variables used for the analysis of depression.
The results shown in Tables 4.9, 4.10 and 4.11 suggest firstly that the difference between females and males remains statistically significant at a level of 1%, but the effect seems slightly lower compared to the one for depression: the odds for women to feel lonelier than before the pandemic were 1.39 times as large as the odds for men, while for depression they were 1.68 times. According to this, we will consider later also in this case a subdivision between males and females, with the aim to see and study any differences between the two groups.
We focus then on the variables related to health, in particular to the comparison between physical health before and after Covid-19. As for depression they appear statistically significant; in particular for those who have a poor health before Covid, the odds of felling more alone are greater than for those with good health. This result is reasonable but not entirely expected. In fact, while for depression this result was intuitive, in this case it was not so obvious. For the former it is indeed perceivable that already having a sub-optimal health status could lead to even more worries during a pandemic, while for the latter dependent variable it is more

difficult to interpret. First of all it is necessary to say that the variable is composed by 5 possible levels (excellent, very good, good, fair, poor) and that the level taken as reference in the logistic regression model is "excellent"; consequently, there are four coefficients in the model. So we find out that for a person whose health before Covid was poor, the odds of feeling more lonely than before the Covid-19 are about 42% higher than the odds for those in excellent health; always compared to this category, respectively for those with fair, good and very good health the odds of feeling more lonely are 35%, 18%, and 2% higher. We can hypothesise that for a person who already had health problems, especially an elderly person, and who perhaps struggled to get out of the house or who already did not have many contacts, having those few contacts "taken away" further amplified the problem of loneliness, which was perhaps already present. Still remaining in this theme, we find that the variable related to the worsening or improvement of health turns out to be significant. For those who have had a deterioration in health, the odds of feeling more alone than before are even 2.66 times as large as the odds for those who have had an improvement. The variable referred to having had episodes of falling or dizziness appears significant as for depression: for those who have had at least one of these episodes the odds of feeling lonelier are 43 % higher than the odds for those who have not. We can make this variable part of those related to physical health, so the same assumptions considered above may apply.

For what concerns the behaviour, all variables *washing hands*, *hand sanitizer* and *assuming drugs against Covid-19*, result statistically significant and all have an estimate of the coefficient around -0.13. We can then assert that for those who do not increase the frequency of these behaviours compared to before the odds of being more lonely are about 15% lower than the odds for those who increase these behaviours. This result could be interpreted as that maybe people who increased these behaviours were more worried about the disease and even when allowed tended to have as few contacts as possible. Proceeding with the analysis of the coefficients, we note that, differently from what we saw for the depression variable, none of the variables closely related to Covid-19, which we recall being having a symptomatic or positive person in the home and having had at least one family member hospitalised or deceased because of this disease, turns out to be statistically significant. This is a rather unexpected result, since one might have thought that having had to witness one of these events might accentuate the feeling of loneliness.

Estimates of coefficients for logistic regression model		
Variables	**Estimates**	**significance**
intercept	-2.71	***
gender	0.33	***
usual home	-0.05	
health before Covid 19 very good	0.02	
health before Covid 19 good	0.17	.
health before Covid 19 fair	0.30	**
health before Covid 19 poor	0.35	**
health improvement worsened	0.98	***
health improvement same	-0.03	
only one major illness	-0.16	.
two or more major illnesses	-0.03	
falling	0.36	***
not having left home	-0.03	
wash hands same as before	-0.13	.
hands sanitizer same as before	-0.13	*
drugs vs Covid 19	-0.14	.
symptoms inside household	0.008	
positive case inside household	0.01	
hospitalised person inside household	0.40	
dead person of the household	0.056	*
not having forgone treatment	-0.39	***
postponement of medical appointment	-0.04	
satisfaction with hospital treatment 2 (somewhat satisfied)	0.10	
satisfaction with hospital treatment 3 (somewhat dissatisfied)	0.11	
satisfaction with hospital treatment 4 (very dissatisfied)	0.09	
satisfaction with hospital treatment 5 (was not treated in hospital)	0.03	

Table 4.9: Estimates of logistic regression coefficients for *loneliness* part 1

Estimates of coefficients for logistic regression model		
Variables	**Estimates**	**significance**
somewhat satisfied with medical office	0.14	.
somewhat dissatisfied with medical office	0.19	.
very dissatisfied with medical office	0.12	
did not go to medical office (5)	-0.04	
not having become unemployed	-0.18	
working location usual	- 0.15	
working location both	-0.16	
working location none	0.05	
working location not employed	0.16	
ends meet with some difficulty	- 0.24	***
ends meet fairly easily	- 0.38	***
ends meet easily	-0.53	***
not having received financial support	-0.01	
physical contact with child at least once a week	0.29	***
physical contact with child at least never or almost never	0.50	***
online contact with child at least once a week	-0.05	
online contact with child never or almost never	- 0.29	***
physical contact with friends at least once a week	0.15	
physical contact with friends never or almost never	0.34	***
online contact with friends at least once a week	-0.03	
online contact with friends never or almost never	-0.11	
life satisfaction wave 7 (from 1 to 10)	-0.02, -0.23, -0.19, 0.10, -0.08, -0.10, -0.05, -0.28, -0.37, -0.46	all ., apart level 10:*
life happiness wave 7 level 2	0.03	
life happiness wave 7 level 3	-0.08	
life happiness wave 7 level 4	-0.10	

Table 4.10: Estimates of logistic regression coefficients for *loneliness* part 2

With regard to medical care, a significant variable is *having forgo a treatment*: those who did not have to do this are about 32% less likely to feel lonely than before than those who have not had to postpone anything. For what concerns the satisfaction, the one with hospital treatment does not result significant, while for those who were treated in a medical office and were somewhat satisfied or somewhat dissatisfied the odds of feeling lonelier are respectively 15% and 20% higher than the odds for who was satisfied. Proceeding with the analysis of the coefficients we can see that the economic component also seems to be influential for the dependent variable. In particular, the variable referred to the ability of the household to make ends meet is significant. For those who manage to make ends meet easily, the odds of feeling lonelier compared to before the pandemic are 41% lower than the odds for those who make ends meet with great difficulty. As regards the other classes, again taking as reference the class of those who reach the end of the month with great difficulty, the odds of feeling lonelier are respectively 32% and 21% lower for those who reach the end of the month fairly easily and with some difficulties.

We can then analyse the variables related to social relationships and physical and virtual contacts. As anticipated above, we expect these variables to be central to the study of loneliness. Indeed, the coefficients of all four variables referring to physical and virtual contact with children and friends turn out to be statistically significant at the 1% level. For what concerns the physical contact with children, the result suggests that for those who have never or almost never a contact with their own children, the odds of feeling lonelier are 65% higher than the odds for those who see them daily, while for who sees them at least once per week the value drops to 34%. These values, if compared with the ones of depression, that were respectively 33% and 19%, are quite higher; this is consistent with what was expected. Also the physical contact with friends or relative outside the household appears statistically significant: for those who hardly ever or never see their friends the odds of feeling lonelier are 40% higher than the odds for those who see them daily, while for those who see them at least once a week are only 15% higher (really this is not a significant coefficient, probably because the frequency compared to before has not changed so radically). Also in this case, the effect is greater than for depression. For what concerns the online contacts, the effect is for loneliness as for depression inverse: in fact, the estimates of coefficients have a negative sign. This means that both for friends and children the virtual way of contact has not the desired effect, but the opposite. As mentioned before, this could be explained by the fact that elderly people could be less used to social media and suffer more from the difference of contact.

The composition of the household also turns out to be very important in studying subjects' loneliness. In fact, both the variables referred to the size of household and the presence of the partner are statistically significant at a level of 1%. For what concerns the household size, for those who have one more member in the household, the odds of feeling lonelier are about 16% lower than the odds for who does not. Then,for those who do not have a partner in the household the odds of feeling lonelier are 2.11 times as large as the odds for those who have one. From a descriptive point of view, 50% of people that assert to feel lonelier after COVID outbreak do not have a partner, while the same relative frequency for those who do not affirm to feel lonelier is 28%. In the training set only 4453 out of 38740 subjects feel lonelier than before and among them precisely 2194 did not have a partner also before Covid-19. This confirms our expectations and therefore we can say that in some way loneliness seems to have affected in a greater measure those who already before the pandemic were alone, in the sense that they did not have a partner in the household and so probably lived alone.

The last block of variables to be dealt with is that relating to personality. From the output of the model we derive that the only trait of personality that turns out to be significant is the one related to neuroticism. This result should not surprise us, as it is in line with the results found previously about depression. In particular, it seems that for those who have one more point in this variable the odds of feeling more lonely than before the pandemic are 13% higher than for those who have a point less.

Finally, also for *loneliness* it appears that the variable country is significant: we can then assume that in every country the effect of pandemic was different, as had been hypothesised in Chapter 1 and as seen from the descriptive analysis in Chapter 2.

To compare the different models, as done before, we report the total classification error, that in this case amounts at 0.315, using the threshold of 0.11, that is the relative frequency of positive answers to the question "Do you feel lonelier than before the outbreak of Covid-19?". Also the rates of false positives and negatives are around a value of 0.30. We will compare the results with those of other models.

Estimates of coefficients for logistic regression model		
Variables	**Estimates**	**significance**
openness	0.025	
agreeableness	-0.02	
neuroticism	0.125	***
extra-version	-0.02	
consciousness	0.02	
country DE	0.31	*
country SE	1.07	***
country NL	0.56	**
country ES	0.14	
country IT	1.11	***
country FR	0.22	.
country DK	0.43	**
country GR	0.98	***
country CH	0.54	***
country BE	0.87	***
country IL	0.48	**
country CZ	-0.44	**
country PL	-0.19	
country LU	0.72	***
country HU	-0.5	*
country PT	0.12	
country SI	-0.02	
country EE	-0.59	***
country HR	0.29	*
country LT	-0.55	**
country BU	-0.01	
country CY	0.53	**
country FI	0.34	*
country LV	-0.68	***
country MT	0.88	***
country RO	-0.02	
country SK	0.46	**
household size	-0.17	***
no partner in household	0.75	***

Table 4.11: Estimates of logistic regression coefficients for *loneliness* part 3

4.2.2 Classification tree

We proceed subsequently trying to adapt to the data a different model, in order to have a comparison of the results. Therefore, we decide to adapt to the data a classification tree. As done before, we divide in two subsets the training set; of these two subsets, one was used to grow the tree and contains 60% of the observations, while the other is used for pruning and consequently contains 40% of total observations. Also in this case as for depression we have used the deviance as function to minimise. In the pruning phase, the most appropriate number of leaves was found, that is equal to 13, and which was then used in the final tree. Looking therefore at the graph of the final tree, several considerations can be made regarding the variables that seem to discriminate most between one outcome and the other.

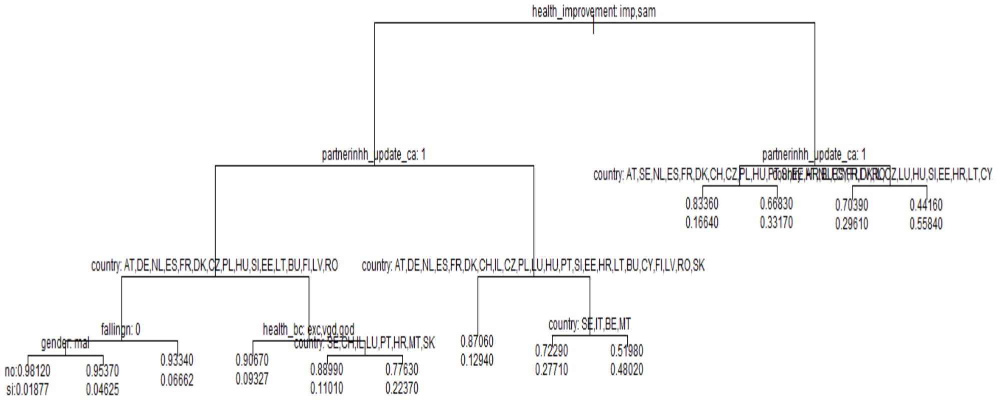


Figure 4.4: Classification tree for *loneliness*

From Figure 4.4 we can notice that the majority of people are classified as not lonelier than before; this result reflects the initial proportions in the data sample, because, as mentioned earlier, only 11% of people stated that they felt more lonely. It is although interesting and appropriate to compare the different probabilities associated, that are visible at the end of every branch. The first split variable is *health improvement*. From the values of the probabilities we can see that if the health is improved or is remained around the same level, the probability of not be lonelier than before is around values of 0.90, with some exceptions, while if we look at the right part of the tree, that is referred to those who who have had a deterioration in health, these probabilities are lowered to values around 75%. The second influential variable is the one referred to having the partner in the household. In the left part of the graph there are the probabilities of who lives with the partner, that as said before, are higher for the answer "not feeling more alone than before". Then, some differences are noticeable through the variable country. It seems that in every branch there are some groups based on it. For example, we can see that countries like Italy, Malta, and Sweden are often in the same final leaves. For these countries the probability of not being more lonely is lower than for the other countries, and we can compare this result with the stringency indexes of these ones. Another country that seems different from the majority is Greece. If we look, in fact, at the final branch which has as split variable country with the levels Sweden, Italy, Belgium and Malta, we can observe that there is a division: observations of these countries (obviously taking into account also the previous subdivisions) are associated with a probability of feeling lonelier that is about 0.28. For Greece the same probability is quite higher: it is indeed equal to 0.48.

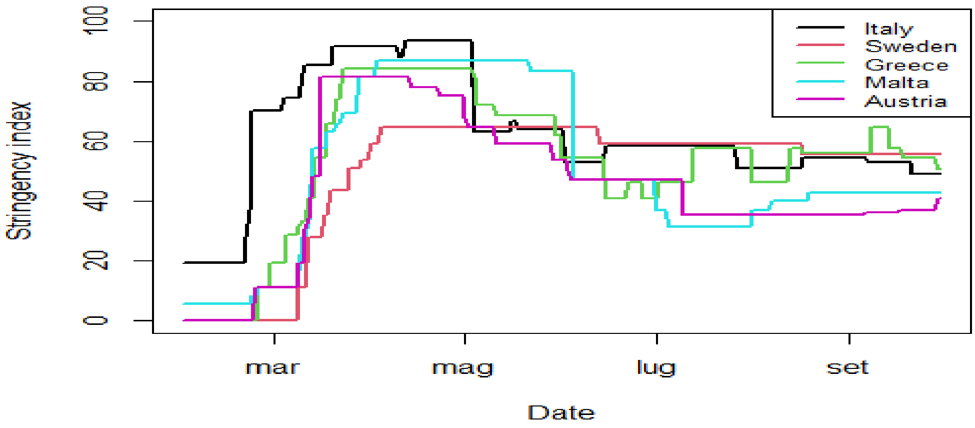


Figure 4.5: Stringency index for selected countries

From Figure 4.5 it is possible to see the different trends of the stringency index of the countries cited above, with the addition of Austria, to consider also one of the other countries. In fact, with the exception of Sweden, it seems that all these countries have a similar stringency index trend and that the values of this index is at higher levels for these countries than for the others.

Thus analysing the predictive ability we notice that for this model, although, 36% of the observations are misclassified, always taking as threshold the relative frequency of people that feel lonelier than before. This result leads to prefer in terms of predictive ability the logistic regression model, but is however useful to use this model to interpret the effects of the variables.

4.2.3 Gradient boosting

As was done for the depression variable, we want to adapt the gradient boosting model for the variable loneliness as well. The variable has to be transformed into numerical and we proceed by estimating this model to the training set. This type of model can supply us useful information regarding the order of importance of the variable ones. In Figure 4.6 all the variables are reported, from the first in relative importance to the last. With the classification tree we found that the variable *country* was discriminant for loneliness. This figure can in some way confirm this result: *country* has in fact the 60% of relative importance and so it seem quite more relevant than other variables. Among the last variables there are instead the one referred to having in the household someone with symptoms and having lived in the usual home. Both these results were found also through the previous models, because these variables neither were statistically significant nor appeared in the tree.

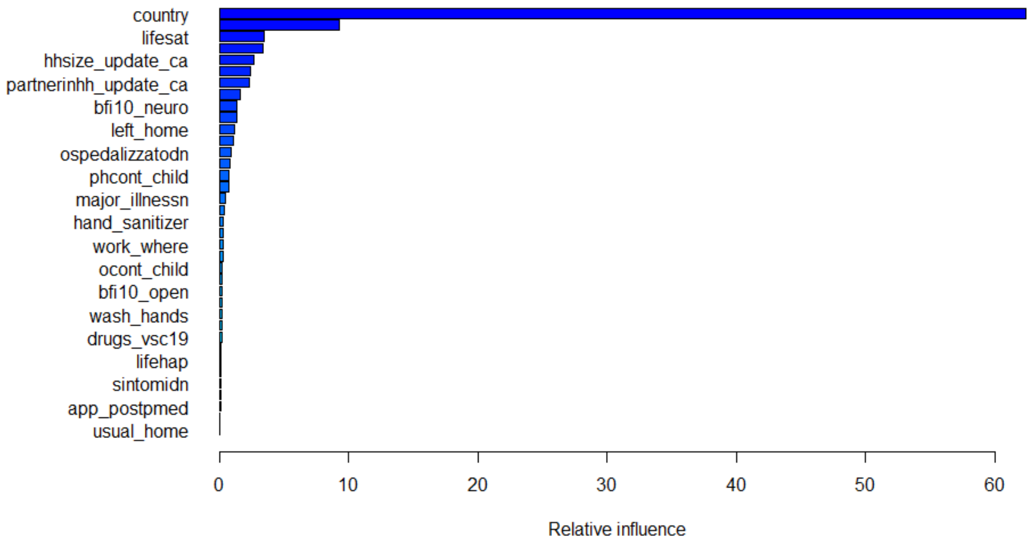


Figure 4.6: Relative importance of all variables for *loneliness* according to gradient boosting

It is then interesting to take into consideration and try to interpret the five variables that are considered most important, which are illustrated in Figure 4.7.

Apart from the already commented country, we find *health improvement*, *life satisfaction*, *having forgone a treatment* and *household size*. As already said previously mental health and physical health are strictly correlated: this is in part confirmed by the fact that *health improvement* appears in the second position and it has almost 10% of total importance. Also the satisfaction with life seems relevant and this is a result reasonable but new in comparison with the previous analyses. In line with what has already been found are instead the fact that the size of the household is important for the feeling of loneliness in a person and that having forgone a treatment could influence the mental health of the subject in this particular situation.

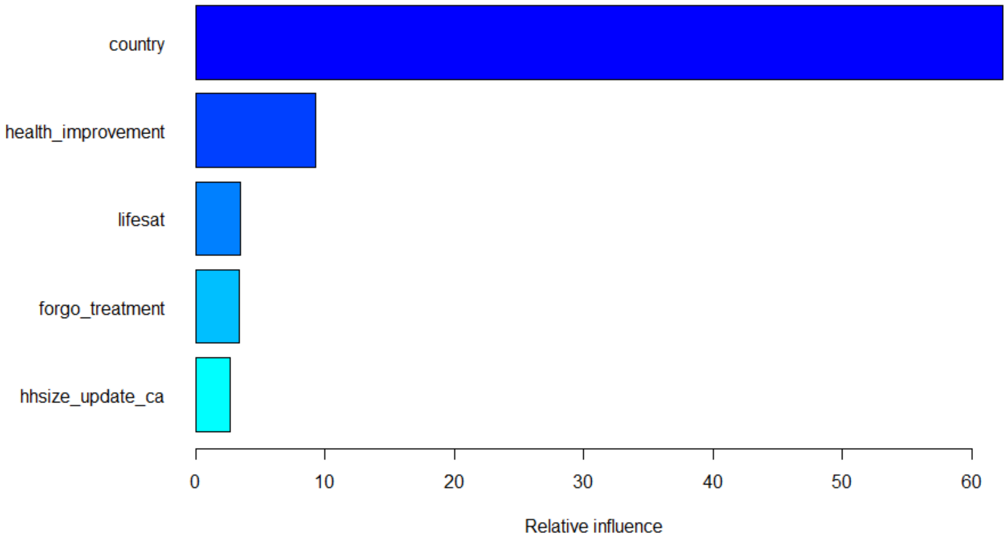


Figure 4.7: Relative importance of five most influential variables for *loneliness* according to gradient boosting

4.2.4 Different models by gender

As anticipated in Section 4.2.1 we want to understand if there are relevant differences between women and men in loneliness; in the logistic regression model in fact we found that the coefficient associated to the variable *gender* resulted statistically significant.

Hence, we decided to re-estimate the model with the same explanatory variables, except obviously the one of gender, separately for these two groups and highlight possible differences. For this reason in Tables 4.12, 4.13 and 4.14 are reported the estimates of parameters and their associated significance. The differences in the presence of significance regard the variables *health before Covid-19*, *drugs against Covid-19*, *hospitalisation of household member*, *satisfaction with medical office* and *online contact with children and friends.* The most relevant difference worth commenting on concerns the field of social relations. For both the groups the physical contact with children is important and the less frequent it is more they are likely to feel more alone, but for women not seeing their children on a daily basis seems to have a slightly greater effect than for men. For mothers, having virtual contact with their children is also significant, in contrast to fathers; as for depression, the effect is also negative for loneliness. As well as physical contact with children, contact with friends and relatives outside the home seems to be important. Again, however, for women it seems to have a greater effect. The difference appear for those who see friends at least once a week compared to those who see them daily: for males, for those who see friends at least once a week the odds of feeling lonelier are 5% higher than the odds for who see them daily, while for females the odds are almost 21% higher. For what concerns the virtual contact with friends the variable is not statistically significant for any group. With regard to personality, the only trait that appears significant is neuroticism, and the effect is really similar for men and women: as the variable increases by one unit, the odds of feeling more lonely are about 14% for men and 17% for women higher than the odds for those who have one less point in this personality trait. Another comment can be made about the variable of hospitalisations: for men it is significant and the coefficient estimate is greater than 1, whereas for women it is not significant. This could actually be due to the fact that for this variable the number of positive answers is very low. Finally, also the coefficients of the different countries appear different between males and females in terms of values and significance. In order to understand what the predictive capacity of both models is, we can analyse the misclassification table as we did before. For the women's model the total misclassification rate is 0.30 (the threshold is set at 0.139 which corresponds to the relative frequency of positive answers), while for the men's model it is slightly lower (the threshold is set at the corresponding relative frequency).

4.3 Predictive accuracy of models

For what concerns the predictive accuracy of the models for *depression*, as reported before, the total misclassification errors of the logistic regression and of the gradient boosting are similar and about 0.30; also the rate of false positives and negatives are around that value in both models. The classification tree leads us to greater errors, precisely of a total of 0.44. For this reason, from a predictive point

of view, the logistic regression model and the gradient boosting seem to provide a better prediction.
With regard to *loneliness*, the same considerations are true: the total misclassification error for both logistic model and gradient boosting is around 0.31, while for the tree is about 0.36.

Effect of variables by gender				
Variables	**Estimates for males**	**Significance for males**	**Estimates for females**	**Significance for females**
intercept	-3.85	***	-2.76	***
health before Covid 19 very good	-0.17		0.16	
health before Covid 19 good	0.002		0.37	**
health before Covid 19 fair	0.32	.	0.40	***
health before Covid 19 poor	0.47	*	0.43	**
health improvement worsened	1.38	***	0.99	***
health improvement same	0.44	.	0.009	
only one major illness	-0.19		-0.05	
two or more major illnesses	0.02		0.07	
falling	0.41	***	0.29	***
wash hands same as before	-0.26	*	-0.16	.
hands sanitizer same as before	-0.28	**	-0.13	*
drugs vs Covid 19	-0.15		-0.16	.
symptoms inside household	-0.06		0.11	
positive case inside household	0.154		0.01	
hospitalised person inside household	1.16	*	-0.56	
dead person of the household	-0.53		0.27	
not having forgone treatment	-0.35	***	-0.48	***
postponement of medical appointment	-0.05		-0.05	

Table 4.12: Estimates of logistic regression coefficients by gender for *loneliness* part 1

Effect of variables by gender				
Variables	**Estimates for males**	**Significance for males**	**Estimates for females**	**Significance for females**
somewhat satisfied with hospital treatment	-0.15		-0.13	
somewhat dissatisfied with hospital treatment	-0.78		0.20	.
very dissatisfied with hospital treatment	0.59		-0.09	
no treated in hospital	- 0.01		0.06	
somewhat satisfied with medical office	0.29	*	0.10	
somewhat dissatisfied with medical office	0.53	.	0.27	
very dissatisfied with medical office	0.11		0.26	
did not go to medical office	-0.01		-0.010	
not having become unemployed	-0.29		-0.10	
working location usual	-0.47	.	-0.29	.
working location both	-0.124		-0.22	
working location none	-0.002		-0.08	
working location not employed	0.05		-0.09	
ends meet with some difficulty	- 0.20	.	-0.33	***
ends meet fairly easily	- 0.35	**	-0.48	***
ends meet easily	-0.54	***	-0.61	***
physical contact with child at least once a week	0.21	.	0.36	***
physical contact with child at least never or almost never	0.43	***	0.54	***
online contact with child at least once a week	0.09		-0.03	
online contact with child never or almost never	- 0.11		-0.22	*
physical contact with friends at least once a week	0.05		0.19	.
physical contact with friends never or almost never	0.37	*	0.38	***
online contact with friends at least once a week	-0.06		-0.05	
online contact with friends never or almost never	-0.07		-0.13	

Table 4.13: Estimates of logistic regression coefficients by gender for *loneliness* part 2

Effect of variables by gender				
Variables	**Estimates for males**	**Significance for males**	**Estimates for females**	**Significance for females**
openness	0.008		0.035	
agreeableness	-0.0012		-0.04	
neuroticism	0.13	***	0.16	***
extra-version	-0.004		-0.04	
consciousness	-0.013	.	-0.012	
country DE	0.16		0.29	.
country SE	0.79	**	0.99	***
country NL	0.22		0.69	**
country ES	0.24		0.09	
country IT	1.17	***	1.05	***
country FR	0.07		0.39	*
country DK	0.06		0.39	*
country GR	1.40	***	0.84	***
country CH	0.23		0.57	***
country BE	0.87	***	0.81	***
country IL	0.84	**	0.32	
country CZ	-0.36		-0.47	**
country PL	-0.29		-0.15	
country LU	0.48		0.66	**
country HU	-1.42	*	-0.43	.
country PT	0.05		0.12	
country SI	0.18		-0.02	
country EE	-0.70	**	-0.57	***
country HR	0.41		0.27	
country LT	-0.30		-0.52	**
country BU	0.35		0.18	
country CY	1.21	***	0.23	
country FI	0.09		0.31	.
country LV	-0.77	.	-0.40	.
country MT	0.63	*	0.86	***
country RO	-0.05		-0.08	
country SK	0.90	**	-0.28	
household size	-0.19	***	-0.21	***
no partner in household	0.55	***	0.32	***

Table 4.14: Estimates of logistic regression coefficients by gender for *loneliness* part 3

Conclusion

The work described in this thesis has led to articulated results which have been presented in the previous chapters. In this last part we propose to summarise the most significant results.

With respect to the main goal, which was to identify the consequences of the Covid-19 pandemic and its isolation on depression and loneliness among the elderly, a first significant result is related to the gender difference. Females appear to be significantly more affected than males, particularly by depression; gender also appears to be more important than other elements that seem to have a strong influence on the onset of depression, such as previously compromised health or psychological elements such as nervousness, anxiety and sleep disorders. In addition to these inner contributory factors, one can also add elements related to one's own behaviour or caused by circumstances, such as a concern for hand hygiene or a lack of attention from health care facilities. There are a number of factors that even in normal times could favour the onset of depression and that in times of the pandemic are simply accentuated. These certainly include the temporary or permanent loss of a job or the physical separation from friends or family, particularly children. Surprisingly, the frequency of contact with children and relatives via technological means of communication was found to have a negative influence on depression: the more frequent these contacts were, the more they tended to be depressed. An explanation for this, which should perhaps be further investigated, could lie in the feeling of nostalgia caused by them or in the frustration at the lack of ability to make proper use of technological means of communication.

With regard to loneliness, the results partly coincide with those for depression, but not completely. Similar effects are noted for gender difference and pre-pandemic health conditions. Other factors, such as the occurrence of Corona related symptoms or the consequences of Covid-19 in one's social life, as well as knowing people who are positive, hospitalised or died because of the disease, do not seem to be statistically significant for the variable loneliness in the logistic regression model and do not appear among the most important variables either in the classification tree or in gradient boosting.

Similar effects on loneliness and depression are found in the frequency of contact with children and friends and in the importance of the personality trait of neuroticism. In these cases the effect is greater for the former, but still follows the same dynamic. Finally, it is worth mentioning the differences found on both variables in the different countries considered. Comparing, for example, the results of the model estimates with the trend of the stringency index, it could be observed that in several cases the countries that imposed more restrictions favoured an increase in both loneliness and depression among the elderly.

Given the continuous updating of the SHARE surveys and of the individual states' policies towards the pandemic, a future survey could investigate in a more refined way the possible correlations between policies undertaken and mental health, also to provide recommendations to governments. Other aspects that might be interesting to note are those linked to the arrival of vaccinations, with their load of reassurance, but also of polemics and social tensions. In the current historical phase, finally, we are beginning to speak of 'pandemic fatigue', since the continuation of an emergency state beyond two years has certainly affected the resilience of everyone, and in particular the elderly. Undoubtedly, the data provided by SHARE opens up a wide range of possibilities for research, of which this work is only one particular example.

Bibliography

[1] *Median age in Italy.* URL: https://www.epicentro.iss.it/en/coronavirus/bollettino/Report-COVID-2019_10_january_2022.pdf (visited on 01/13/2022).

[2] *Spread of Coronavirus.* URL: https://www.who.int/news-room/questions-and-answers/item/coronavirus-disease-covid-19-how-is-it-transmitted (visited on 01/13/2022).

[3] *Definition of SHARE.* URL: http://www.share-project.org/home0.html (visited on 01/13/2022).

[4] Leslie Kish. "Multipopulation survey designs: five types with seven shared aspects". In: *International Statistical Review/Revue Internationale de Statistique* (1994), p. 173.

[5] Mick P Couper. "The future of modes of data collection". In: *Public Opinion Quarterly* 75.5 (2011), pp. 889–908.

[6] Axel Börsch-Supan and Michael Bergmann. "SHARE Wave 8 Methodology: Collecting Cross-National Survey Data in Times of COVID-19". In: (2021).

[7] *Definition of depression.* URL: https://www.stateofmind.it/tag/depressione/ (visited on 02/13/2022).

[8] Aaron T Beck and Brad A Alford. *Depression: Causes and treatment.* University of Pennsylvania Press, 2009.

[9] Edward Bibring. "The mechanism of depression." In: (1953).

[10] B Bleichmar Hugo. "Some subtypes of depression and their implications for psychoanalytic treatment". In: *International Journal of Psycho-Analysis* 77 (1996), pp. 935–961.

[11] Sigmund Freud. *Inhibitions, symptoms and anxiety SE 20 [→].* 1926.

[12] Ulfert Hapke, Caroline Cohrdes, and Julia Nübel. "Depressive symptoms in a European comparison–Results from the European Health Interview Survey (EHIS) 2". In: *J. Health Monit* 4 (2019), pp. 57–65.

[13] Joan S Girgus, Kaite Yang, and Christine V Ferri. "The gender difference in depression: are elderly women at greater risk for depression than elderly men?" In: *Geriatrics* 2.4 (2017), p. 35.

[14] Martin Kockler and Reinhard Heun. "Gender differences of depressive symptoms in depressed and nondepressed elderly persons". In: *International journal of geriatric psychiatry* 17.1 (2002), pp. 65–72.

[15] Pamela Santaera, Rocco Carmine Servidio, and Angela Costabile. "Anziani e depressione: il ruolo della solitudine". In: (2017).

[16] *UK survey.* URL: `https://www.mentalhealth.org.uk/our-work/research/coronavirus-mental-health-pandemic/key-statistics-wave-8` (visited on 02/13/2022).

[17] *European survey.* URL: `https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/696164/EPRS_BRI(2021)696164_EN.pdf` (visited on 02/13/2022).

[18] M. de Bruijne and al. "Dutch Mixed Mode Experiment. Version: 1.0.0. SHARE–ERIC. Dataset. https://doi.org/10.6103/SHARE.w6NLmmExp.100". In: (2017).

[19] Beatrice Rammstedt. "The 10-item big five inventory". In: *European Journal of Psychological Assessment* 23.3 (2007), pp. 193–201.

[20] Axel Börsch-Supan et al. *Health and socio-economic status over the life course First results from SHARE Waves 6 and 7.* De Gruyter, 2019.

[21] Adelchi Azzalini and Bruno Scarpa. *Data analysis and data mining: An introduction.* OUP USA, 2012.

[22] Trevor Hastie et al. *The elements of statistical learning: data mining, inference, and prediction.* Vol. 2. Springer, 2009.

[23] Marta Nai Ruscone. "Modelli gerarchici: aspetti metodologici e ambiti di applicazione". In: (2011).

[24] Jane Elith, John R Leathwick, and Trevor Hastie. "A working guide to boosted regression trees". In: *Journal of animal ecology* 77.4 (2008), pp. 802–813.

[25] Thomas Hale et al. "Variation in government responses to COVID-19". In: *Blavatnik school of government working paper* 31.2020-11 (2020).

[26] Gilberto Ghilardi and Nicola Orsini. "Modelli lineari ad intercetta casuale, stimatori e valutazione di sistemi formativi". In: *Statistica* 62.4 (2002), pp. 695–713.

Appendix A: COVID-19 Questionnaire

Variables from CV-SHARE8	
Variables	**Description**
mergeid	Fixed id between waves
firstwavehh	First Wave in which the household was sampled
cvresp	Coverscreen respondent
age2020	Age in 2020
ageint	Age at the time of interview (only for those before the pandemic)
partnerinhh	Partner in household
hhsize	Household size
interview	Index of interview done in wave 8
intyear	Year of interview
intmonth	Month of interview
interview-ca	Index of interview done in wave 8 after Corona
intyear-ca	Corona Interview year
intmonth-ca	Corona Interview month
deceased-update-ca	Index of death- update Coronavirus
hhsize-update-ca	Household size- update Coronavirus
partnerinhh-update-ca	Partner in household- update Coronavirus
hhmoved	Household moved
nursinghome	Living in a nursing home

Table 15: Choice of variables from SHARE-cv8 dataset

.1 A – Intro and basic demographics

CAA001

Some time ago, we sent you an invitation letter, which also included a data protection statement. Have you received the statement? 1. Yes 5.No

IF CAA001 = 5

CAA002

In this case, I will then summarise the most important points of the statement for you.Furthermore, I will be pleased to answer any question regarding the protection of your data that you may have now. The [FILL in name of CTL institution] in cooperation with SHARE-ERIC are responsible for the implementation of the survey. We, [FILL in name of Survey Agency], are commissioned to carry out the interviews. The purpose of the study is to provide scientists with data on health, socio-economic status and social and family networks to address their research questions in relation to the process of population ageing. Participating in this interview is voluntary and the information is kept confidential. We will not record the conversation. During the interview, I will enter your answers in a computer. They will be stored together with a code number only. I.e., your contact details and names are strictly stored separately from the information provided by you during the interview. Your contact details and names will be stored until the end of the SHARE study's last wave of data collection only. After the collection of the individual interviews, they will be compiled and later on be researcher knowing your identity. The results of the analyses will be presented in an anonymised form only. If we should come to any question you don't want to answer, just let me know and I will go on to the next question. Non-participation will not lead to any disadvantages for you. You can also withdraw consent at any time with effect for the future. Furthermore, you have several other data protection rights. In the next step, I will tell you how you can receive more information about your rights.Do you agree to participate in this study?

IWER: Answer all questions of R.

1. Yes, R consented to participate.
2. No, R refused to participate. No interview possible.

IF CAA002 = 1

CAA003

Thank you. For further information, you can contact us by calling [FILL in telephone number of survey agency]. Furthermore, we can send the data protection statement to you again. Do you want us to send you the statement once more?

END IF

ELSE IF CAA001 = 1

CAA004

If you have questions regarding the data protection statement, I will be pleased to answer them. Let me stress that participating in this interview is voluntary and that the information is kept confidential. We will not record the conversation. Instead, during the interview, I will enter your answers in a computer. Your answers will be used only for research purposes in different analyses, without the individ-

ual researcher knowing your identity. If we should come to any question you don't want to answer, just let me know and I will go on to the next question. Do you agree to participate in this study?
IWER: Answer all questions of R.
1. Data protection statement has been provided; R consented to participate.
2. Data protection statement has been provided; R refused to participate. No interview possible.
END IF
IF CAA002 = 2 || CAA004 = 2
CAA005
IWER: Are you sure that Respondent has refused to participate?
1. Yes, R refused. Terminate interview.
2. No, R consented. Continue interview.
END IF
IF CAA002 = 1 || CAA004 = 1 || CAA005 = 2
CADN042
IWER: Note sex of respondent (ask if unsure).
1. Male
2. Female
CADN002
In which month were you born?
CADN003
In which year were you born?
CAA006
Are you in your usual home now or have you temporarily moved elsewhere due to Corona?
1. Usual home
2. Lives now temporarily elsewhere
CAA010
Now I have a set of questions about how you were affected by Corona

.2 H – Health (physical and mental) and health behavior

CAPH003
Before the outbreak of Corona, would you say your health was excellent, very good, good, fair, or poor?
1. Excellent
2. Very good
3. Good

4. Fair
5. Poor

CAH002

If you compare your health with that before the outbreak of Corona, would you say your health has improved, worsened, or stayed about the same?

1. Improved
2. Worsened
3. About the same

CAH003

Since we last interviewed you, were you diagnosed with a major illness or health condition?

1. Yes
5. No

IF CAH003 = 1

CAH004

Do you have any of the following illnesses or health conditions? Please answer yes or no:

IWER: With this we mean that a doctor has told you that you have this condition, and that

you are either currently being treated for or bothered by this condition.

IWER: READ OUT.

CAH004-1 Hip fracture?
CAH004-2 Diabetes or high blood sugar?
CAH004-3 High blood pressure or hypertension?
CAH004-4 A heart attack including myocardial infarction or coronary thrombosis or any other heart problem including congestive heart failure?
CAH004-5 Chronic lung disease such as chronic bronchitis or emphysema?
CAH004-6 Cancer or malignant tumor, including leukemia or lymphoma, but excluding minor skin cancers?
CAH004-7 An other illness or health condition

1. Yes
5. No
-1. Don't know
-2. Refusal

END IF

CAPH089

For the past six months at least, have you been bothered by any of the following health conditions?

Please answer yes or no:

IWER: READ OUT.

CAPH089-1 Falling down
CAPH089-2 Fear of falling down
CAPH089-3 Dizziness, faints or blackouts
CAPH089-4 Fatigue
1. Yes
5. No
-1. Don't know
-2. Refusal

CAH006

Do you regularly take prescription drugs?
1. Yes
5. No
IF CAH006 = 1

CAH007

Do you take any of the following drugs? Please answer yes or no: Drugs for...
IWER: READ OUT.
CAH007-1 High blood cholesterol?
CAH007-2 High blood pressure?
CAH007-3 Coronary or cerebrovascular diseases?
CAH007-4 Other heart diseases?
CAH007-5 Diabetes?
CAH007-6 Chronic bronchitis?
1. Yes
5. No
-1. Don't know
-2. Refusal
END IF

CAH010

Since the outbreak of Corona, have you ever left your home?
1. Yes
5. No
IF CAH010 = 1

CAH011

Since the outbreak of Corona, how often have you done the following activities, as compared to before the outbreak? Not any more, less often, about the same, or more often?
IWER: Read out each activity and check the appropriate answer.
CAH011-1 Going shopping?
CAH011-2 Going out for a walk?
CAH011-3 Meeting with more than 5 people from outside your household?

CAH011-4 Visiting other family members?
1. Not any more
2. Less often
3. About the same
4. More often
5. Does not apply
-1. Don't know
-2. Refusal

CAH012

How often did you wear a face mask when you went outside your home to a public space? Was it always, often, sometimes, or never?
1. Always
2. Often
3. Sometimes
4. Never

CAH013

How often did you keep distance to others when you went outside your home? Was it always, often, sometimes, or never?
1. Always
2. Often
3. Sometimes
4. Never

END IF

CAH014

Did you wash your hands more frequently than usual?
1. Yes
5. No

CAH015

Did you use special hand sanitizer or disinfection fluids more frequently than usual?
1. Yes
5. No

CAH016

Did you pay special attention to covering cough and sneeze?
1. Yes
5. No

CAH017

Did you take any drugs or medicine as a prevention against Corona?
1. Yes
5. No

CAH020

In the last month, have you felt nervous, anxious, or on edge?
1. Yes
5. No
IF CAH020 = 1
CAH021
Has that been more so, less so, or about the same as before the outbreak of Corona?
1. More so
2. Less so
3. About the same
END IF
CAMH002
In the last month, have you been sad or depressed? IWER: If participant asks for clarification, say 'by sad or depressed, we mean miserable, in low spirits, or blue'.
1. Yes
5. No
IF CAMH002 = 1
CAMH802
Has that been more so, less so, or about the same as before the outbreak of Corona?
1. More so
2. Less so
3. About the same
END IF
CAMH007
Have you had trouble sleeping recently? IWER: DO NOT READ OUT.
1. Trouble with sleep or recent change in pattern
2. No trouble sleeping
IF CAMH007 = 1
CAMH807
Has that been more so, less so or about the same as before the outbreak of Corona?
1. More so
2. Less so
3. About the same
END IF
CAMH037
How much of the time do you feel lonely? Often, some of the time, or hardly ever or never?
1. Often
2. Some of the time
3. Hardly ever or never
IF CAMH037 = 1 || CAMH037 = 2

CAMH837
Has that been more so, less so or about the same as before the outbreak of Corona?
1. More so
2. Less so
3. About the same
END IF

.3 C – Corona-related infection

CAC001
Now I will ask you about whether you, someone in your family or among your neighbors and friends has been affected by the Corona illness.
CAC002 Since the outbreak of Corona, did you or anyone close to you experience symptoms that you would attribute to the Covid illness, e.g. cough, fever, or difficulty breathing? IWER: Respondent can think of people who live close, and people who are close in an emotional sense, like family members.
1. Yes
5. No
IF CAC002 = 1
CAC003
Who was it? Please tell me their relationship to you. IWER: Check all that applies and enter the number of persons in the checkbox on the right.
IWER: PROBE: 'Any others?'
1. Respondent
2. Spouse or partner
3. Parent
4. Child
5. Other household member
6. Other relative outside household
7. Neighbor, friend or colleague
8. Caregiver
97. Other
END IF
CAC004
Have you or anyone close to you been tested for the Corona virus and the result was positive, meaning that the person had the Covid disease?
1. Yes
5. No
IF CAC004= 1

CAC005

Who was tested positive? Please tell me their relationship to you. IWER: Check all that applies and enter the number of persons in the checkbox on the right.
IWER: PROBE: 'Any others?'
1. Respondent
2. Spouse or partner
3. Parent
4. Child
5. Other household member
6. Other relative outside household
7. Neighbor, friend or colleague
8. Caregiver
97. Other
END IF

CAC007

Have you or anyone close to you been tested for the Corona virus and the result was negative, meaning that the person did not have the COVID disease or has recovered from it?
1. Yes
5. No
IF CAC007 = 1

CAC008

Who was tested and the result was negative? Please tell me their relationship to you. IWER: Check all that applies and enter the number of persons in the checkbox on the right. IWER: PROBE: 'Any others?'
1. Respondent
2. Spouse or partner
3. Parent
4. Child
5. Other household member
6. Other relative outside household
7. Neighbor, friend or colleague
8. Caregiver
97. Other
END IF

CAC010

Have you or anyone close to you been hospitalized due to an infection from the Corona virus?
1. Yes
5. No

IF CAC010 = 1
CAC011
Who was hospitalized? Please tell me their relationship to you. IWER: Check all that applies and enter the number of persons in the checkbox on the right. IWER: PROBE: 'Any others?'
1. Respondent
2. Spouse or partner
3. Parent
4. Child
5. Other household member
6. Other relative outside household
7. Neighbor, friend or colleague
8. Caregiver
97. Other
END IF
CAC013
Has anyone close to you died due to an infection from the Corona virus?
1. Yes
5. No
IF CAC013 = 1
CAC014
I am very sorry. Can you tell me who that was? IWER: Check all that applies and enter the number of persons in the check box on the right.
2. Spouse or partner
3. Parent
4. Child
5. Other household member
6. Other relative outside household
7. Neighbor, friend or colleague
8. Caregiver
97. Other
END IF

.4 Q – Quality of healthcare

CAQ001
Now I have some questions about your doctor visits and the healthcare system since the outbreak of Corona.
CAQ005

Since the outbreak of Corona, did you forgo medical treatment because you were afraid to become infected by the corona virus?
1. Yes
5. No
IF CAQ005 = 1
CAQ006
Which type of medical treatment did you forgo? Please answer yes or no. Did you forgo... IWER: READ OUT.
CAQ006-1 Check up at a general practitioner?
CAQ006-2 Check up at a specialist, including a dentist?
CAQ006-3 A planned medical treatment, including an operation?
CAQ006-4 Physiotherapy, psychotherapy, rehabilitation?
CAQ006-97 Some other type of medical treatment?
1. Yes
5. No
-1. Don't know
-2. Refusal
END IF
CAQ010
Did you have a medical appointment scheduled, which the doctor or medical facility decided to postpone due to Corona?
1. Yes
5. No
IF CAQ010 = 1
CAQ011 Which type of medical treatment had to be postponed? Please answer yes or no: IWER: READ OUT.
CAQ011-1 Check up at a general practitioner?
CAQ011-2 Check up at a specialist, including a dentist?
CAQ011-3 A planned medical treatment, including an operation?
CAQ011-4 Physiotherapy, psychotherapy, rehabilitation?
CAQ011-97 Some other type of medical treatment?
1. Yes
5. No
-1. Don't know
-2. Refusal
END IF
CAQ015
Did you ask for an appointment for a medical treatment since the outbreak of Corona and did not get one?
1. Yes

5. No
IF CAQ015 = 1
CAQ016
Which type of medical treatment were you denied? Please answer yes or no. Were you denied... IWER: READ OUT.
CAQ016-1 Check up at a general practitioner?
CAQ016-2 Check up at a specialist, including a dentist?
CAQ016-3 A planned medical treatment, including an operation?
CAQ016-4 Physiotherapy, psychotherapy, rehabilitation?
CAQ016-97 Some other type of medical treatment?
1. Yes
5. No
-1. Don't know
-2. Refusal
END IF
CAQ025
Since the outbreak of Corona, were you treated in a hospital?
1. Yes
5. No
IF CAQ025 = 1
CAQ027 How satisfied were you with the way you were treated? Very satisfied, somewhat satisfied, somewhat dissatisfied, or very dissatisfied?
1. Very satisfied
2. Somewhat satisfied
3. Somewhat dissatisfied
4. Very dissatisfied
IF CAQ027 = 3 || CAQ027 = 4
CAQ028 Why were you dissatisfied? IWER: Let R mention all reasons and check all that applies.
1. Long waiting time
2. Overcrowded
3. Doctor and nurses did not have time for me
4. Shortage of equipment and supplies
5. Insufficient safety measures against infections
97. Other
END IF
END IF
CAQ020 Since the outbreak of Corona, did you go to a doctor's office or a medical facility other than a hospital?
1. Yes

5. No
IF CAQ020 = 1
CAQ021 Was this related to Corona?
1. Yes
5. No
CAQ022 How satisfied were you with the way you were treated? Very satisfied, somewhat satisfied, somewhat dissatisfied, or very dissatisfied?
1. Very satisfied
2. Somewhat satisfied
3. Somewhat dissatisfied
4. Very dissatisfied
IF CAQ022 = 3 || CAQ022 = 4
CAQ023
Why were you dissatisfied? IWER: Let R mention all reasons and check all that applies.
1. Long waiting time
2. Overcrowded
3. Doctor and nurses did not have time for me
4. Shortage of equipment and supplies
5. Insufficient safety measures against infections
97. Other
END IF
END IF

.5 W - Work

CAW001
I now turn to the economic consequences of the Corona crisis, first to your work situation.
CAEP805 At the time when Corona broke out, were you employed or self-employed, including working for family business?
1. Yes
5. No
IF CAEP805 = 1
CAW002 Due to the Corona crisis have you become unemployed, were laid off or had to close your business? IWER: Business closure can be both temporarily or permanently.
1. Yes
5. No

IF CAW002 = 1
CAW003 How long were you unemployed, laid off or had to close your business?
IWER: Number in weeks.
END IF
CAW010
Since the outbreak of Corona, some people worked at home, some at their usual work place outside their home, some both. How would you describe your situation?
IWER: If R got unemployed, laid off, or had to close business since the outbreak, R should think of the remaining time he or she worked during the outbreak. None of these means that did not work at all, neither at the usual workplace nor at home.
1. Worked at home only
2. Worked at the usual work place
3. Worked from home and at the usual work place
4. None of these
IF CAW010 != 4
IF CAW010 = 1 || CAW010 = 3
CAW012 Did you learn new computer skills?
1. Yes
5. No
9. Works without computer
CAW013 Was your Internet connection adequate? Please answer yes or no:
1. Yes
5. No
9. Works without internet
END IF
IF CAW010 = 2 || CAW010 = 3
CAW016
Did you get any protection such as masks, gloves, protective screens, disinfection fluid at the work place?
1. Yes
5. No
CAW017
How safe did you feel health-wise at your work place? Was it very safe, somewhat safe, somewhat unsafe, or very unsafe?
1. Very safe
2. Somewhat safe
3. Somewhat unsafe
4. Very unsafe
END IF

CAW020
How many hours per week did you normally work before the outbreak of Corona? Please include overtime.
CAW021
Did you reduce your working hours since the outbreak of Corona? IWER: If R got unemployed, laid off, or had to close business, code 'Yes'.
1. Yes
5. No
IF CAW021 = 1
CAW022 What was the lowest number of hours in a single week? IWER: If R got unemployed, laid off, or had to close business, put 0 hours.
IF CAW022 is response
CAW023-1
When was that?
CAW023-2
In which week of the month was that?
END IF
END IF
CAW024
Did you increase your working hours since the outbreak of Corona? Please include overtime.
1. Yes
5. No
IF CAW024 = 1
CAW025
What was the highest number of hours in a single week?
IF CAW025 is response
CAW026-1
When was that?
CAW026-2
In which week of the month was that?
END IF
END IF
END IF
END IF

.6 E – Economic situation

CAE001 IWER: Are you interviewing the first respondent in this household?
1. Yes
5. No
IF CAE001 = 1
CAE002 I now want to ask you to compare your household's financial situation before and after the outbreak of Corona.
CAHH017
How much was the overall monthly income, after taxes and contributions, that your entire household had in a typical month before Corona broke out? IWER: Enter an amount in [currency of country].
CAE003 Did you or any other household member receive additional financial support due to the outbreak of Corona from your employer, the government, relatives, friends, and/or others?
1. Yes
5. No
IF CAE003 = 1
CAE004 Who gave you this financial support? IWER: Check all that applies.
IWER: Probe: "Any others?"
1. Employer
2. Government
3. Relatives
4. Friends
97. Others
END IF
CAE005
What was the lowest overall monthly income, after taxes and contributions, that your entire household had, including any financial support you may have received since the outbreak of Corona? IWER: Enter an amount in [currency of country].
CACO007
Thinking of your household's total monthly income since the outbreak of Corona, would you say that your household is able to make ends meet with great difficulty, with some difficulty, fairly easily, or easily?
1. With great difficulty
2. With some difficulty
3. Fairly easily
4. Easily
IF CACO007 = 1 || CACO007 = 2
CAE011

Since the outbreak of Corona, did you need to postpone regular payments such as rent, mortgage and loan payments, and/or utility bills?
1. Yes
5. No
CAE012 Since the outbreak of Corona, did you need to dip into your savings to cover the necessary day-to-day expenses?
1. Yes
5. No
END IF
END IF

.7 S – Social networks

CAS001
I would now like to hear about the kinds and frequency of contacts that you have with family and friends from outside your home.
CAS003
Since the outbreak of Corona, how often did you have personal contact, that is, face to face, with the following people from outside your home? Was it daily, several times a week, about once a week, less often, or never? IWER: Read out each relationship and check the appropriate answer.
CAS003-1 Own children:
CAS003-2 Own parents:
CAS003-3 Other relatives:
CAS003-4 Other non-relatives like neighbors, friends, or colleagues:
1. Daily
2. Several times a week
3. About once a week
4. Less often
5. Never
99. Not applicable
-1. Don't know
-2. Refusal
CAS004 Since the outbreak of Corona, how often did you have contact by phone, email or any other electronic means with the following people from outside your home? (Was it daily, several times a week, about once a week, less often, or never?) IWER: Read out each relationship and check the appropriate answer.
CAS004-1 Own children:
CAS004-2 Own parents:

CAS004-3 Other relatives:
CAS004-4 Other non-relatives like neighbors, friends, or colleagues:
1. Daily
2. Several times a week
3. About once a week
4. Less often
5. Never
99. Not applicable
-1. Don't know
-2. Refusal

CAS010

Since the outbreak of Corona, did you help others outside your home to obtain necessities, e.g. food, medications or emergency household repairs?
1. Yes
5. No
IF CAS010 = 1

CAS011

Compared to before the outbreak of Corona, how often did you help the following people from outside your home to obtain necessities: less often, about the same, or more often? IWER: Read out each relationship and check the appropriate answer.
CAS011-1 Own children:
CAS011-2 Own parents:
CAS011-3 Other relatives:
CAS011-4 Other non-relatives like neighbors, friends, or colleagues:
1. Less often
2. About the same
3. More often
99. Not applicable
-1. Don't know
-2. Refusal
END IF

CAS012

Since the outbreak of Corona, did you provide personal care to others outside your home?
1. Yes
5. No
IF CAS012 = 1

CAS013

How often did you provide personal care to the following people from outside your home compared to before the outbreak of Corona; less often, about the same, or

more often? IWER: Read out each relationship and check the appropriate answer.
CAS013-1 Own children:
CAS013-2 Own parents:
CAS013-3 Other relatives:
CAS013-4 Other non-relatives like neighbors, friends, or colleagues:
1. Less often
2. About the same
3. More often
99. Not applicable
-1. Don't know
-2. Refusal
END IF
CAS015 Since the outbreak of Corona, did you do any other volunteering activity?
1. Yes
5. No
IF CAS015 = 1
CAS016
Was it less often, about the same, or more often than the volunteering that you did before the outbreak of Corona?
1. Less often
2. About the same
3. More often
END IF
CAS020
Since the outbreak of Corona, were you helped by others from outside of home to obtain necessities, e.g. food, medications or emergency household repairs?
1. Yes
5. No
IF CAS020 = 1
CAS021 How often did the following people from outside your home help you to obtain necessities, compared to before the outbreak of Corona? Less often, about the same, or more often? IWER: Read out each relationship and check the appropriate answer.
CAS021-1 Own children:
CAS021-2 Own parents:
CAS021-3 Other relatives:
CAS021-4 Other non-relatives like neighbors, friends, or colleagues:
1. Less often
2. About the same

3. More often
99. Not applicable
-1. Don't know
-2. Refusal
END IF
CAS025 Did you regularly receive home care before the outbreak of Corona?
1. Yes
5. No
IF CAS025 = 1
CAS026 Since the outbreak of Corona, did you face more difficulties in getting the amount of home care that you need?
1. Yes
5. No
IF CAS026 = 1
CAS027 Which difficulties were they? IWER: Let R mention all difficulties and check all that applies.
1. I had to pay more to get the help I need
2. People who cared for me could not come to my home
3. Other difficulties
END IF
CAS028 Did the people who cared for you wear protective devices such as masks or gloves?
1. Yes
5. No
99. No caregiver visited my home since the outbreak.
END IF

.8 F – FINALE

CAF001 We now come to the end of the interview. These were a lot of questions about a hard time. But even during hard times there are some good things in life. What was your most uplifting experience since the outbreak of Corona, in other words, something that inspired hope or happiness? IWER: DO NOT READ OUT. Let respondent answer and choose appropriate option.
1. Named something right-away
2. Hesitated to name something
3. Did not name anything
CAF002
Finally, what is it that you are looking most forward to doing once Corona abates?

IWER: DO NOT READ OUT. Let respondent answer and choose appropriate option.
1. Named something right-away
2. Hesitated to name something
3. Did not name anything

CAF003

Thank you very much for your kind cooperation. Stay healthy!
END IF

CAF004

IWER: Please enter any remarks about this interview you want to tell us.

CAF005

IWER CHECK: Who answered the question?
1. Respondent only
2. Respondent and proxy
3. Proxy only