Socioeconomic Status and the Prevalence of Chronic Health Problems in Portugal

Maria Teresa Figueirôa

Dissertation written under the supervision of Professor Miguel Gouveia

Dissertation submitted in partial fulfilment of requirements for the MSc in Economics
at Católica-Lisbon School of Business and Economics,
January 2021.

Socioeconomic Status and the Prevalence of Chronic Health Problems in Portugal

Thesis written by Maria Teresa Pontes de Brito Figueirôa

Abstract

Portugal, throughout the times, has displayed high health inequalities based on socioeconomic differences, as is the case in many other countries. This work aims at estimating the correlation between each individual's socioeconomic status (SES) and the prevalence of chronic health problems. Previous studies have focused on specific diseases, but a more complete and detailed analysis is needed, so as to better capture the distribution of mental, physical and, even, mixed chronic conditions in the Portuguese population. To that end, the incidence of depression, diabetes, chronic pulmonary obstructive disease and chronic pain was analyzed. SES was measured by a composite indicator, which aggregates several variables through the use of Multiple Correspondence Analysis. The usage of this statistical method came, inevitably, as a result of the excessive aggregation of income data in the National Health Survey of 2014 (*Inquérito Nacional de Saúde*). Concentration indices were then computed and shown to be statistically significant, uncovering the disproportionately high prevalence of chronic conditions in the population with lower SES. Therefore, further research should be carried out to better comprehend the primary causes of the revealed inequality as well as to develop a more adequate policy response.

Keywords: Chronic Diseases; Chronic Obstructive Pulmonary Diseases; Depression; Diabetes; Health Inequalities; Multiple Correspondence Analysis; Socioeconomic Status

O Estatuto Socioeconómico e a Prevalência de Problemas de Saúde Crónicos em Portugal

Tese redigida por Maria Teresa Pontes de Brito Figueirôa

Resumo

Portugal, tal como outros países, tem vindo a exibir, ao longo dos anos, elevadas desigualdades em saúde, baseadas em diferenças socioeconómicas. Este trabalho tem como objetivo estimar a correlação entre o estatuto socioeconómico de cada indivíduo e a prevalência de problemas de saúde crónicos. Estudos passados têm investigado doenças específicas, mas uma análise mais completa e detalhada é necessária, para que seja possível perceber a manifestação de condições crónicas mentais, físicas, e mesmo mistas, na população Portuguesa. Nesse sentido, a incidência de depressão, diabetes, doença pulmonar obstrutiva crónica e dor crónica foi analisada. O estatuto socioeconómico foi traduzido por um indicador composto criado com base na agregação de variáveis, através do uso da Análise de Correspondência Múltipla. Esta utilização resultou, inevitavelmente, da excessiva agregação que os dados relativos ao rendimento apresentam no Inquérito Nacional de Saúde de 2014. Os índices de concentração foram calculados e revelam ser estatisticamente significativos, demonstrando a prevalência desproporcionalmente alta destas condições crónicas na população com estatuto socioeconómico mais baixo. Assim sendo, pesquisas adicionais devem ser levadas a cabo, visando melhor estudar as causas primárias para a desigualdade encontrada e as políticas mais indicadas para lidar com este problema.

Palavras-chave: Análise de Correspondência Múltipla; Depressão; Desigualdades em Saúde; Diabetes; Doenças Crónicas; Doença Pulmonar Obstrutiva Crónica; Dor Crónica; Estatuto Socioeconómico

Acknowledgements

First off, I would like to present my sincere gratitude to Professor Miguel Gouveia, my supervisor, for his immeasurable support throughout this process.

Importantly, I would like to thank Católica-Lisbon and all its faculty for becoming my second home for the last five years. A special appreciation goes to my Academic Director, Professor Catarina Reis, for her unpayable advice, and to my master teachers, who lectured me with especially interesting topics and presented me with most of the tools used to develop this thesis.

I would also like to express my appraisal to *Instituto Nacional de Estatística* (INE) for the opportunity of using the data required to develop this research work.

A huge thanks to my friends for their support during my life. A special word goes to my boyfriend, for the patience and love.

Lastly, to my family and, most importantly, to my wonderful parents for always making me feel the luckiest daughter in the world. Without them, I would not have the opportunity to pursue my dreams. This dissertation is for them.

Contents

1. Introduction 10

2. Literature Review 12

2.1. Diabetes 12

2.2. COPD 13

2.3. Depression 14

2.4. Chronic Pain 15

2.5. The SES Variable 16

3. Data Analysis 18

3.1. The National Health Survey (2014): An Overview 18

3.2. Descriptive Statistics 18

4. Methodology 27

4.1. SES Variable: Construction and Implementation 27

4.2. On the Measurement of Health Inequality 36

5. Results 40

5.1. Concentration Indices and Curves Results 40

5.2. The Case of Evitable SES-related Inequality in Health 45

6. Analysis 50

6.1. Discussion 50

6.2. Public Policy Implementation 51

7. Conclusion 53

8. References 54

9. Appendix 64

A. Additional Data Analysis 64

B. Variable Creation Procedures 72

C. Multiple Correspondence Analysis Procedure 74

D. Indirect Standardization Procedure 78

E. Additional Tables of Results 83

List of Figures

Figure 1 - Sample Gender Division... 19
Figure 2 - Sample Age Group Distribution 20
Figure 3 - Sample Education Distribution.. 20
Figure 4 - Sample Distribution by Intensity of Unmet Needs....................................... 21
Figure 5 - Sample Distribution by Self-reported Health Status 21
Figure 6 - Sample Distribution by the Number of Chronic Conditions 22
Figure 7 - Chronic Diseases' Prevalence by Age Group 23
Figure 8 - Chronic Diseases' Prevalence by Intensity of Unmet Needs 26
Figure 9 - Biplot of MCA Coordinates 30
Figure 10 – MCA's First Dimension Projection Plot... 31
Figure 11 - Kernel Density Estimates for MCA's First Component................................ 35
Figure 12 - Kernel Density Estimates for SES Final Variable.. 36
Figure 13 - Concentration Curves: Chronic Diseases and SES Variable........................ 43
Figure 14 - Depression and SES Concentration Curves with Respective Confidence Intervals
.... 44
Figure 15 - Odds Ratio for Age Groups 48

Figure A2. 1 - Life Satisfaction Self-reporting by Diseases Prevalence......................... 66

List of Tables

Table 1 - Sample Region Division 20
Table 2 - Sample Chronic Conditions' Prevalence 22
Table 3 – Prevalence of Chronic Health Problems across Population Sub-groups 25
Table 4 - Per Capita PPP by Region, in 2013 28
Table 5 - MCA Stata Output 32
Table 6 - Indicator Matrix of 5 Example Observations 34
Table 7 - SES Variable Statistics 36
Table 8 - SES Concentration Index (Gini) and Concentration Curve Deciles 40
Table 9 - Chronic Conditions Concentration Indices and Concentration Curve Deciles 42
Table 10 - Standardized Chronic Conditions Concentration Indices and Concentration Curve Deciles 49

Table A1. 1 – Population and Sample Gender Division 64
Table A1. 2 - Sample and Population Region Division 64
Table A1. 3 - Population and Sample Age Group Division 65
Table A1. 4 - Population and Sample Education Division 65

Table A2. 1 - Chronic Diseases Prevalence by Age Group, Intensity of Unmet Needs and Health Status Self-appreciation 67
Table A2. 2 - Chronic Disease Prevalence by Health Variables and Self-reporting of Life Satisfaction 68
Table A2. 3 - CP Prevalence by Pain Interference 69
Table A2. 4 - Diabetes and COPD Prevalence by Heart Conditions 70
Table A2. 5 - COPD Prevalence by Smoking Habits 70
Table A2. 6 - Depression Incidence by Depressive Symptoms 71
Table A2. 7 - Diabetes, COPD and CP Incidence by Depressive Symptoms 71

Table C1. 1 - Categorical Variables Used in MCA 74
Table C1. 2 - MCA Complete Stata Output 76
Table C1. 3 - Principal Inertias of Unadjusted Burt Matrix 77

Table C1. 4 - Principal Inertias of Adjusted Burt Matrix 77

Table D1. 1 - Auxiliary Logistic Regression for Diabetes and Depression 81
Table D1. 2 - Auxiliary Logistic Regression for COPD and Chronic Pain 82

Table E1. 1 – Concentration Index and Concentration Curve Deciles for the SES variable ... 83
Table E1. 2 – Concentration Indices and Concentration Curve Deciles for Diabetes 83
Table E1. 3 – Concentration Indices and Concentration Curve Deciles for Depression 84
Table E1. 4 – Concentration Indices and Concentration Curve Deciles for COPD 84
Table E1. 5 – Concentration Indices and Concentration Curve Deciles for Chronic Pain 85

Table E2. 1 - Standardized Results for Diabetes 86
Table E2. 2 - Standardized Results for Depression 86
Table E2. 3 - Standardized Results for COPD 87
Table E2. 4 - Standardized Results for Chronic Pain 87

List of abbreviations

CA – Correspondence Analysis

CDC – Center for Disease Control and Prevention

COPD – Chronic Obstructive Pulmonary Disease

CP – Chronic Pain

IASP – International Institute for the Study of Pain

INE – *Instituto Nacional de Estatística* (in English, National Statistics' Institute)

MCA – Multiple Correspondence Analysis

NHS – National Health Survey

NUTS - *Nomenclatura das Unidades Territoriais para Fins Estatísticos* (in English, Territorial Units Nomenclature for Statistical Purposes)

PCA – Principal Component Analysis

PPP – Purchasing Power Parity

SES – Socioeconomic Status

WHO – World Health Organization

1. Introduction

Health is one of the most scrutinized policy areas. Attached to it, there is the undeniable concern of understanding whether individuals from different socioeconomic status (SES) have similar abilities in maintaining a satisfactory health status. This concern corresponds therefore to the well-known persistent socioeconomic inequities in health, which have been repeatedly found in the literature (e.g., van Doorslaer et al., 1997; 2004; Mackenback et al., 2008).

The goal of this research is to focus on the analysis of these same inequalities in the prevalence of chronic health problems, also known as noncommunicable diseases. This large group is defined as including "*conditions that last 1 year or more and require ongoing medical attention or limit activities of daily living or both*" (CDC, 2020). According to the World Health Organization (WHO, 2020), these conditions account for 71% of all deaths in the world. As such, their utmost importance in our modern society is undeniable and, thus, the choice of putting them at the center of this research work.

Therefore, understanding whether certain socioeconomic groups are more adversely affected by chronic diseases is crucial to finding better policies, in order to tackle the affected groups in the most efficient way. In the literature, SES has been referred to as "*one of the most powerful determinants of health*" (Gershon et al., 2012).

Until now, there is no complete consensus on the joint prevalence of chronic diseases and SES. While it is true that the most general evidence is that income or education (i.e., SES measures) have a negative correlation with the prevalence of these conditions (e.g., Espelt et al., 2008; Robbins et al., 2005; Everson et al., 2002), it is also true that some studies do not find any pattern (e.g., Anderson and Horvath, 2004). For this reason, additional research is needed, especially in the case of Portugal.

To allow for a more comprehensive, yet not exhaustive, study, this thesis will devote attention to four specific chronic diseases. These were chosen in order to encompass conditions of different nature, such as physical, mental and mixed chronic diseases. These are diabetes, chronic obstructive pulmonary disease (COPD), depression and chronic pain (CP).

Through the use of the 2014's National Health Survey (NHS) - *Inquérito Nacional de Saúde* - where individuals self-reported, among other factors, about their health status, this dissertation will try to grasp whether, in Portugal, socioeconomic factors are shown to be relevant regarding the occurrence of chronic diseases.

This work hypothesizes that, in line with the main evidence in the literature, a negative correlation between SES and the prevalence of chronic diseases is to be found. Notice that this hypothesis is only a matter of correlation and, thus, causality is a whole different matter.

The major challenge in this research, however, is the lack of an appropriate income measure. As such, there was the need to employ more sophisticated techniques to overcome this information gap. One of this study's intermediate steps was the use of statistical tools to create a summary index variable that is able to estimate each individual's SES, through Multiple Correspondence Analysis (MCA). This idea was inspired by the analysis in Varela's (2020) work, who used Principal Component Analysis (PCA) to reduce the dimensionality of data. This SES measure will then be used to compute concentration indices and represent the respective concentration curves. This methodology follows the work of various research papers, where concentration indices are used to compute SES-based inequalities in the prevalence of health problems (e.g., Kakwani et al., 1997; Hong et al., 2011; Santos et al., 2014).

The following section will review past studies, providing therefore a more in-depth view of the main facts for each of the conditions studied. Additionally, a specific subsection will investigate the use of econometric and statistical tools in the creation of SES variables in previous literature.

Then, a section introducing the NHS and encompassing the major statistics of the relevant variables follows. Afterward, in the methodology section, the mechanics of the creation of the new summary variable will be explored. Later in the same section, the concept of concentration indices and concentration curves as health inequality measures will be detailed.

Next, in *Section 5*, the results will be displayed, with further analysis being addressed in *Section 6*, which includes considerations regarding public policy implementation. A specific subsection of *Section 5* will focus on standardized results and its possible interpretations. Lastly, in *Section 7*, the main conclusions will be presented.

2. Literature Review

2.1. Diabetes

According to the *Sociedade Portuguesa de Diabetologia*[1], in 2015, 7.5% of the population with ages from 20-79, had diagnosed diabetes. However, this number increases to 13.3%, when also considering the possible undiagnosed cases (*Sociedade Portuguesa de Diabetologia*, 2016). At a world level, in 2014, for individuals of 18 years of age onwards, the proportion of diagnosed cases was 8.5% (WHO, 2020). Nonetheless, Beagley et al. (2014) estimate that this number becomes 45.8% higher, by including the undiagnosed cases.

In *Relatório Anual do Observatório Nacional de Diabetes,* 2015, costs due to diabetes are reported to represent almost 1% of the Portuguese GDP, and approximately 10% of the total health expenditure in Portugal (*Sociedade Portuguesa de Diabetologia* , 2016).

Besides these economic facts, there are other consequences. Falcão et al. (2008), in their sample, find that approximately 7.3% and 4.1% of diabetes patients had, respectively, a stroke and a heart attack, as a diabetes-related complication.

Considerable research has been done concerning inequality and diabetes. In Canada, inequalities in diabetes prevalence do exist and have been increasing (Brown et al., 2015). This is consistent with other research, for European countries or the US (Espelt et al., 2008; Robbins et al., 2005), corroborating the hypothesis that diabetes prevalence is more concentrated among the poor.

In Portugal, Santos et al. (2014) find the same pattern, with negative concentration indices, whether SES is measured by education or by income. Nevertheless, inequality based on the individuals' education level is larger than that based on income. The authors justify this result with the existence of the free universal health system in Portugal, which narrows down possible income differences between the population. This could also be since the income variable used is extremely coarse.

Nonetheless, examples of counterevidence exist, as is the case of Mutyambizi et al. (2019), as South Africa displays a pro-poor inequality, i.e., a higher prevalence of diabetes in population from higher SES.

[1] Portuguese Society of Diabetology

However, the most common conclusions are aligned and support the inverse relationship between SES and the prevalence of diabetes. Moreover, certain studies support the idea that both resulting conditions and mortality are also unequally distributed (e.g., Sortsø et al., 2017; Espelt et al., 2008). As such, most of these works are consistent with the idea that a prominent action plan should be to reduce these differences by "*improving health and health care especially for the low-status groups*" (Mielck et al., 2005).

2.2. COPD

Chronic obstructive pulmonary disease is a "*common, preventable and treatable disease that is characterized by persistent respiratory symptoms and airflow limitation*", as defined by the Global Initiative for Chronic Obstructive Lung Disease (GOLD, 2018). According to the Global Burden of Disease Study (WHO, 2018), in 2016, there were 251 million cases in the world. More importantly, in 2010, COPD was the third disease accountable for more deaths (López-Campos et al., 2016), being responsible for 2.8 million deaths worldwide (Burney et al., 2015).

In Portugal, 5.49% of the population in the range of 35-69 years has COPD (Portuguese Pneumology Society, 2018).

Additionally, not only direct costs of the disease (e.g., medical expenses) should be accounted for, but also indirect costs (e.g., productivity loss due to disability), as Viegi et al. (2001) state in their study. Besides, as is the case with other health conditions, COPD patients also display higher probabilities of developing other diseases. In a study conducted by Mannino et al. (2015), cardiovascular diseases, diabetes and asthma were the most prevalent pathologies in patients with COPD, respectively with 34.8%, 22.8% and 14.7%. As the reader might infer, this makes it easier to witness an increasingly higher burden of COPD costs (Lin et al., 2005).

COPD prevalence in the whole population is approximately 1%, however, it abruptly increases to above 10%, if only the 40+ age range is considered (Chapman et al., 2006). Additionally, smoking habits appear as the most-important attributable cause of COPD. Nonetheless, 30% of COPD patients never smoked, showing that other factors, like passive smoking or other types of environmental exposure can also contribute to its occurrence (López-Campos et al., 2016).

Knowing that smokers and older people have a higher propensity for COPD is not enough. One important aspect is to understand its prevalence in what regards SES. Prescott et al. (1999) find evidence that "*income and education are associated with lung function independently of smoking*". Moreover, Gershon et al. (2012) surveyed the literature, finding a negative correlation between SES and COPD, for whichever the SES measure used.

Additionally, being ranked lower in the socioeconomic scale is a significant factor associated with increased COPD mortality (Sommer et al., 2015). Some additional evidence is presented in Miravitlles et al. (2011), showing that COPD patients from lower socioeconomic levels had lower Health-Related Quality of Life.

All this is enough to motivate further research. Besides age and smoking history, different patient target groups can be detected based on SES. Policies seeking the reduction of SES disparities "*could have a greater impact than any of COPD medication currently available*" (Gershon et al., 2012).

2.3. Depression

According to the WHO, depression is a "*mental disorder affecting more than 264 million people worldwide. It is characterized by persistent sadness and a lack of interest or pleasure in previously rewarding or enjoyable activities*" (WHO, 2020). Gusmão et al. (2004) state that at least one in five people will suffer no less than one episode of depression in their lifetime.

Simon and VonKorff (1998) find that 4% of depressed individuals commit suicide. As such, mortality associated with depression does not present numbers as high as other conditions, nonetheless, its importance must not be underestimated. For instance, Ohayon and Schatzberg (2002), referred to depression as the number two cause of disability in Europe. This increases this disease's costs due to the loss of capacity.

In 1992, it was estimated that 80% of society's total depression burden was due to losses in productivity, while 3% were losses due to suicide and the remaining 17% correspond to direct healthcare costs (Gusmão et al., 2004). In 2002, these total costs were estimated to be approximately 15.46 billion euros (WHO, 2005).

Past research indicates that depression in women is around 2-times more common than in men (Kessler et al., 2013), and that older individuals are more susceptible to develop depression (Blazer and Hybels, 2005).

A large amount of literature has investigated the correlation between SES and the prevalence of depression, using several indicators for SES. Andersen et al. (2009) find that the lower the socioeconomic position, the higher the propensity to display depression, whether SES is measured by employment, education or income, although unemployment and income propensities were shown to be stronger.

Other examples are available where several different countries were studied, such as South Korea (Hong et al., 2011) where results were found to be similar. In Portugal, where a group of older adults was studied, the prevalence of mental depression was found to be more pronounced in low-education individuals (Sousa et al., 2017). Rodrigues et al. (2017) also found a positive correlation between unemployment and depression.

Although, over the years, different perspectives appeared in the literature, Lorant et al. (2003) perform a meta-analysis, where they gather many of the previous studies and summarize their findings. The results are in line with the main idea that there is an inverse correlation between SES and the prevalence of depression.

2.4. Chronic Pain

In 2020, the International Institute for the Study of Pain (IASP) defines chronic pain as being: "*An unpleasant sensory and emotional experience associated with, or resembling that associated with, actual or potential tissue damage*" (IASP, 2020). Nicholas et al. (2019) state that the diagnosis is made when symptoms, that are not explained by any other condition, are present for more than 3 months, attached with psychological instability or disability.

Hoy et al., in 2014, estimated that, in 2010, 9.4% of the world population had chronic low back pain. In Portugal, this number is 10.4% (Gouveia et al., 2015) and 36.7% when considering all CP conditions (Azevedo et al., 2012).

One of the hypotheses studied is the existence of a positive correlation between CP and depression. Indeed, Ohayon et al. (2003) find that more than 40% of patients with major depression also report chronic pain. Moreover, Blumer and Heilbronn (1982) named this condition as the pain-prone disorder, that is proven to be distinct from the pain related to a "*well-defined somatic disease*", and thus, CP being a ramification of mental diseases.

Besides, its costs are increasingly high. Von Korff et al. (1991) find that individuals who suffer from CP are substantially more likely to use healthcare services than otherwise. This uncovers the increased pressure this condition deposits in societies' healthcare systems. These are the well-known direct costs, which, in Portugal, are estimated to account for 42.7% of total costs, (Azevedo et al., 2014).

The remainder is related to indirect costs, which correspond to the loss in the productivity of patients with disability. Azevedo et al. (2012) indicate that, in Portugal, "*CP subjects have high pain-related disability*" and that 49% of patients have reported CP interference in their work. Gouveia and Augusto (2011) estimate the value of this loss to be at around € 740 million, resulting from short-term disability and early retirement. These extremely high costs are one of the reasons that justify the necessity of preventive actions.

Todd et al. (2019) state that CP is more common in women, and Rustøen et al. (2005) underline major age-related differences in CP prevalence and duration, whereby older individuals are more affected.

Azevedo et al. (2012) find evidence that, in Portugal, there is a higher prevalence of CP among people positioned lower in the socioeconomic ranks. Others found similar results (e.g., Gouveia et al., 2015; Teixeira, 2018). Previous research conclusions for other countries are in accordance with this same direction (e.g., Elliott et al., 1999; Todd et al., 2019; Großschädl et al., 2015). Interestingly, Saastamoinen et al. (2005) showed similar relationships for the case of CP, but not for the case of acute pain, where none of the SES factors were shown to be significant, which makes it clear that these are two distinct conditions.

As such, as in the case of the previously reviewed diseases, if a structured analysis is conducted using SES as a baseline, we will be uncovering the most affected groups and foster their health outcomes, through policy implementation.

2.5. The SES Variable

Traditionally, equity analysis of health data relies on a measure of income. Unfortunately, the National Health Survey of 2014 only provides the respondents' income quintile, which is a very coarse measure of SES.

To obtain a better estimate of an individual's SES and to overcome the excessively aggregated income data, a synthesis indicator, that hopefully resembles a continuous variable, was created based on several social and economic variables.

The approach of creating a summary SES measure was inspired by Varela's (2020) work using Principal Component Analysis (PCA). In the literature, several techniques in the field of Descriptive Methods have been used to achieve similar goals, besides PCA. Some examples of these are Correspondence Analysis (CA) or Multiple Correspondence Analysis (MCA), which have been used, both in SES index creation and in health-related studies, for instance.

One example comes from Cortinovis et al (1993), who used MCA in the construction of a socioeconomic index and found that people with lower SES were the most vulnerable regarding basic healthcare conditions. Vyas and Kumaranayake (2006) explained how to use PCA to construct an SES index. Engels et al (2014) find that, using MCA to construct a socioeconomic index, there is a negative correlation between the latter and the incidence of strokes. Others have used MCA to compute composite health indicators (e.g., Kohn, 2012).

Knowing that these methods have been used in the literature is a great indicator of their validity and usefulness. Nonetheless, understanding which of these techniques is the most suitable to use in this case is imperative. Essentially, the three are used for the same purpose of reducing the dimensionality of the data, but in different situations. Abdi and Valentin (2007) provide a simple distinction between these three techniques:

> *"MCA is an extension of CA which allows one to analyze the pattern of relationships of several categorical dependent variables. As such, it can also be seen as a generalization of PCA when the variables to be analyzed are categorical instead of quantitative." (Abdi and Valentin, 2007)*

In this sense, Ezzrari and Verme (2013) explain that, whenever one is dealing with categorical qualitative data, the choice should be MCA, which is the generalization of CA used when we have more than two variables. *"MCA is, thus, particularly relevant in studies where a large amount of qualitative data is collected"* (Costa et al., 2013).

Nonetheless, even if one has mixed data, with certain quantitative variables, Asselin (2002) remembers that it is always possible to transform quantitative variables into categorical ones, whereas the opposite is not true. As will be seen in the Methodology section, the raw data variables used in this research are mainly categorical, and that means that MCA should be used. In this case, the work of Kohn (2012) will be the main reference.

3. Data Analysis

3.1. The National Health Survey (2014): An Overview

This thesis uses data from 2014's National Health Survey (NHS) – the fifth *Inquérito Nacional de Saúde* by *Instituto Nacional de Estatística* in collaboration with *Instituto Ricardo Jorge*. Information was summarized into 212 mostly self-reported variables, characterizing health status and habits.

A total of 22 558 households were inquired, with only one member of each family being interviewed. Survey participants were 15 years of age or older and resided in Portugal between September and December of 2014. Only 80.8% of the responses were validated, leading to a total of 18 204 valid answers. The sample was stratified by regions, with diverse weights. In other words, individuals have different weights when computing summary statistics or regressions, for instance. These weights were designed based on the multiplication of three different components. First, the initial design weight of each individual, based on sampling rates. Second, a specific factor to compensate for the effect of missing answers. Last, a general factor that acts to approximate the sample to the original population, through a method known as calibration, in terms of region, gender, education, age group and household dimension.

Nonetheless, there is one clear drawback in the way the dataset is presented, as previously mentioned: there is only information about the quintile to which each individual's household equivalent income belongs. As such, income information is limited because of excessive aggregation. Having said this, to better analyze each individual's socioeconomic position, more advanced techniques will be used, as already hinted. Interestingly, this initial limitation became the motivation for using an innovative way of characterizing each individual's position in the socioeconomic scale.

3.2. Descriptive Statistics

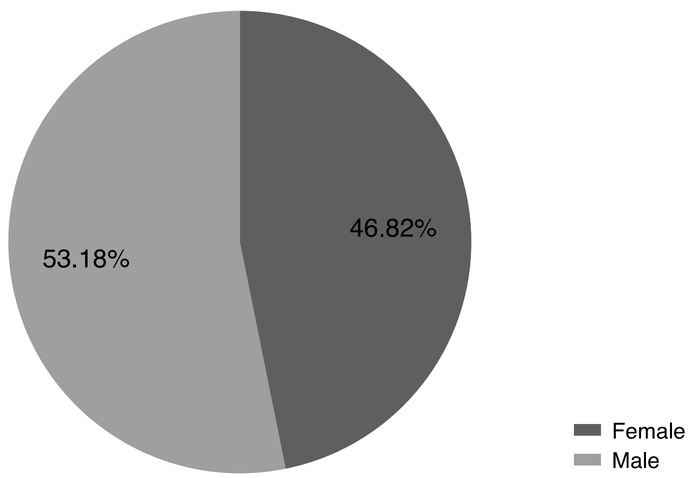
In this subsection, the main objective is to get a general overview of the NHS sample and the prevalence of chronic health problems, based on several characteristics. Note that the results we will be attaining are most certainly easily adaptable to the entire population as, in *Appendix A.1. NHS Sample as a good representation of the Portuguese population*, we understand that, demographically, this sample is representative of the Portuguese population. Even without that evidence, the main idea is that this should be the case, due to the calibration used to approximate the sample to the population statistics.

Important remarks concern tables and figures of results in this and the following sections. All displayed statistics, tables and figures were computed using the sample weights. Moreover, except for those cases where it is clearly stated, all figures and tables were computed using data from 2014's NHS. Additionally, whenever necessary, complementary tables, with complete information are presented in *Appendix A.2. Additional Results, Facts and Characteristics about the NHS Sample and Chronic Patients.*

3.2.1. Demographics

The main demographics of our sample show that it is comprised of 46.82% women and 53.18% men (Figure 1[2]). Regarding region division based on NUTS II (Table 1), the highest share of individuals lives in the North Region, accounting for 35.08% and Lisbon, corresponding to 26.69%. In turn, focusing on age, in Figure 2, it can clearly be seen that the sample includes increasingly high shares of individuals until it reaches the age range of 40-44. As such, the most common age individuals present is from 35 to 54 years, since 34.86% of our sample belongs to those age groups. And, lastly, concerning education, in Figure 3, the highest share (34.6%) has only the 1st or 2nd cycles of completed studies. Other representative groups are those with individuals with 3rd cycle or high-school education, corresponding to respectively 19.5% and 18.6% of the sample.

Figure 1 - *Sample Gender Division*



[2] All displayed figures and tables in this and the following sections were constructed using weighted data.

Table 1 - *Sample Region Division*

	Sample
North	35.08%
Algarve	4.21%
Center	22.17%
Lisbon	26.59%
Alentejo	7.17%
Azores	2.31%
Madeira	2.47%
Total	100%

Figure 2 - *Sample Age Group Distribution*

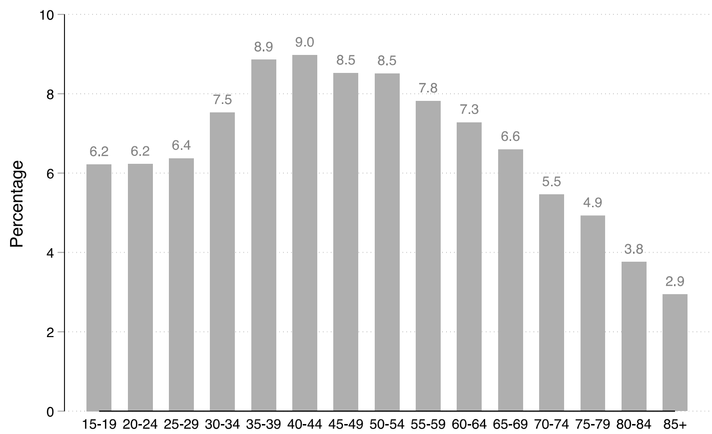
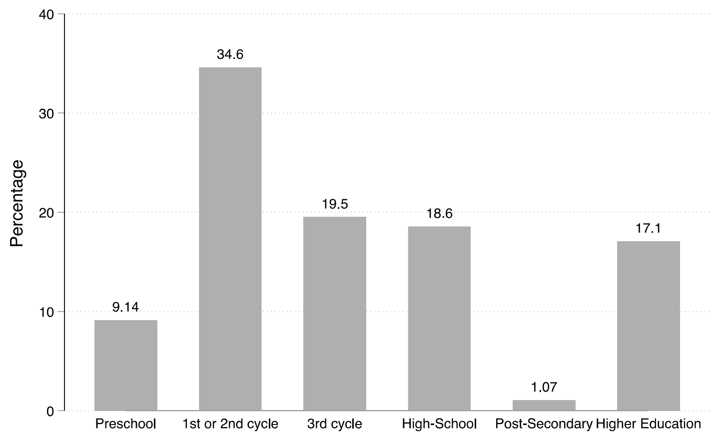


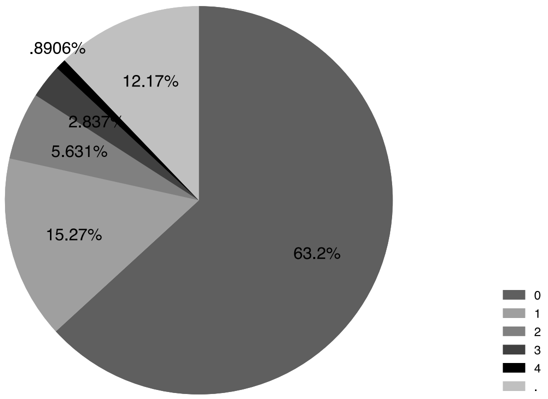
Figure 3 - *Sample Education Distribution*



3.2.2. Other population characteristics

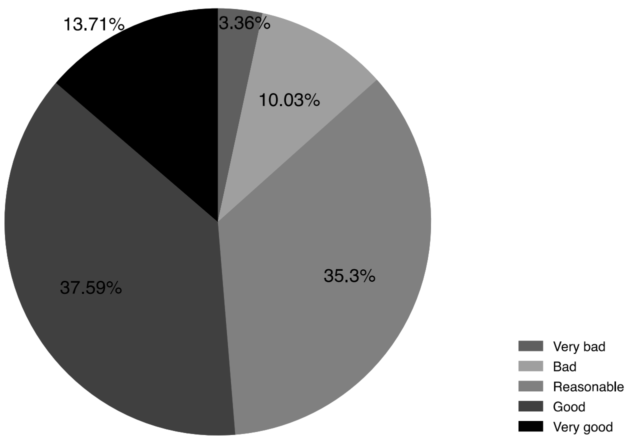
It also seems relevant to analyze some other variables used to conduct this inquiry, namely four variables that may possibly indicate that there are financial needs. These relate to the following questions: “In the past 12 months, due to financial problems, was there a need for any of the four presented areas that was not met: (1) **Medical appointments or treatments**; (2) **Dental appointments, exams or treatments**; (3) **Prescription medicines**; (4) **Psychiatric, psychology or psychotherapy appointments, or other mental health treatment.**”. A new variable was created - **Intensity of Unmet Needs –** where the value of 0, 1, 2, 3, 4 or 5[3] was assigned to each one of the respondents based on the number of needs the individual was unable to satisfy. Further details about this and other variables’ creation process are presented in *Appendix B. Variable Creation Procedures.* Through the analysis of this variable, it can be seen that 24.63% of our sample was unable to meet at least one of these needs, as Figure 4 shows.

Figure 4 - *Sample Distribution by Intensity of Unmet Needs*



As seen in Figure 5, the vast majority of our sample reports their health status as being *Reasonable* (35.30%) or *Good* (37.59%).

Figure 5 - *Sample Distribution by Self-reported Health Status*



[3] Missing values [.] were assigned to individuals who did not face any of the four needs.

3.2.3. Chronic Disease Prevalence

This subsection focuses on a data description of the general prevalence of chronic conditions, while, at the same time, evaluating the main statistics (e.g., demographics) of the population that reports their incidence.

To evaluate the incidence of the four diseases, four binary variables were created, with the value of 1 whenever the respondent reported it and 0, otherwise. Particularly, for the case of chronic pain, the survey participant was assigned 1 if he or she reported pain intensity from moderate to very intense and 0 when responses ranged from none to slight.[4]

The percentage of the sample presenting at least one of these chronic diseases is 40.37%, as seen in Figure 6. In Table 2, we can see that the rarest condition is COPD, which affects 5.79% of respondents, followed by diabetes, present in 9.33% of the sample. Depression and chronic pain are the more frequent conditions in our sample, with their incidence percentages being respectively 11.89% and 29.73%. These numbers are also aligned with information that was previously mentioned in the literature review.

Figure 6 - ***Sample Distribution by the Number of Chronic Conditions***

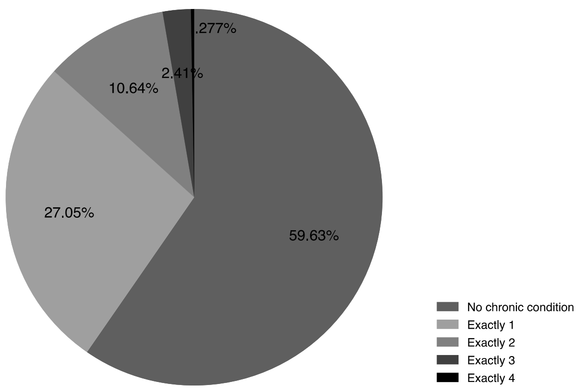


Table 2 - ***Sample Chronic Conditions' Prevalence***

	Diabetes	Depression	COPD	Chronic Pain
Prevalence	9.33%	11.89%	5.79%	29.73%

[4] For the other three diseases, the survey was already designed with three specific direct questions, inquiring respondents about the prevalence of diabetes, depression or COPD.

Referring to age distribution, all diseases present a general increasing pattern, with some minor exceptions. All of them show high differences between younger and older respondents. However, some show more evident differences. As an example, we can mention diabetes, since, for instance, from the 20-24 individuals, only 0.10% have diabetes, while in the 70-74 individuals, 26.27% present the same disease. In Table A2. 1, one can find detailed tables with the results. However, to facilitate the interpretation, spine plots were created for each of the diseases, which are presented in Figure 7.

Figure 7 - ***Chronic Diseases' Prevalence by Age Group***

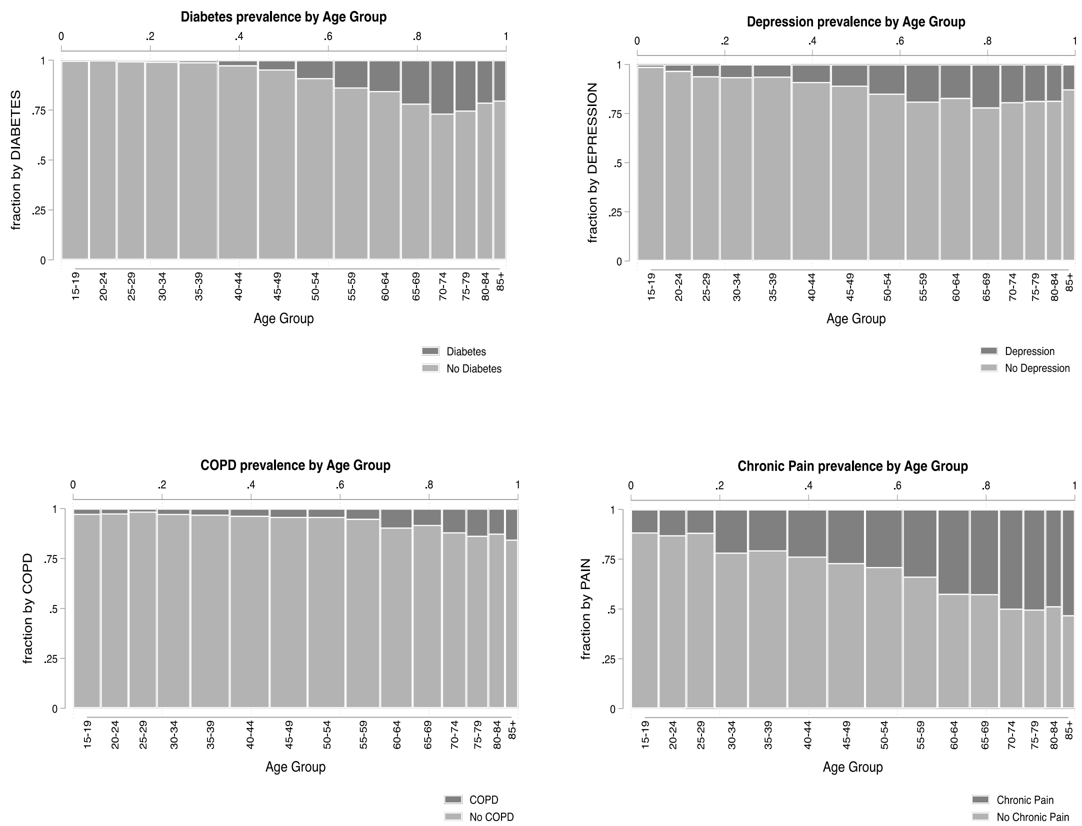


Table 3 summarizes the prevalence of these conditions according to subgroups, showing the percentage of individuals within that subgroup that suffer from that disease, i.e., for example, when considering group division by gender, it can be seen that when considering all the women in the sample as 100%, 9.4% of them suffer from diabetes. Thus, in the following paragraphs, except when stated otherwise, the relevant results are presented in Table 3.

Regarding demographic statistics, concerning gender division, all studied diseases present a higher prevalence in women than in men. The closest distribution is found in diabetes prevalence, as 9.4% of women in our sample are diabetic, whereas 9.2% of men present the same condition. The differences become much more pronounced when studying depression or chronic pain. For instance, only 5.9% of men present depression, while 17.2% of women self-report the same pathology.

Analyzing income, there is a higher prevalence of these chronic conditions in the lower quintiles of household equivalent income. For instance, depression has a prevalence of 15.5% in the first quintile, whilst only 7.4% in the fifth quintile. The same is also true for the other noncommunicable diseases.

Focusing on education there is a higher occurrence of diseases in the lower levels of education, with all conditions being the most prevalent in respondents with only preschool, when compared to higher educational levels.

We may also hypothesize whether the area in which each individual lives has any correlation with the pattern of disease prevalence. Using the NUTS II region variable, there is no clear pattern of disease prevalence. For instance, while Lisbon seems to have a higher prevalence of depression, the corresponding conclusion cannot be taken for the remaining diseases, as the Center region has the highest percentage of affected population when considering CP or COPD. Additionally, in the case of diabetes, 10.4% of the population from Alentejo self-reports diabetes, while in Lisbon the number is 7.7%.

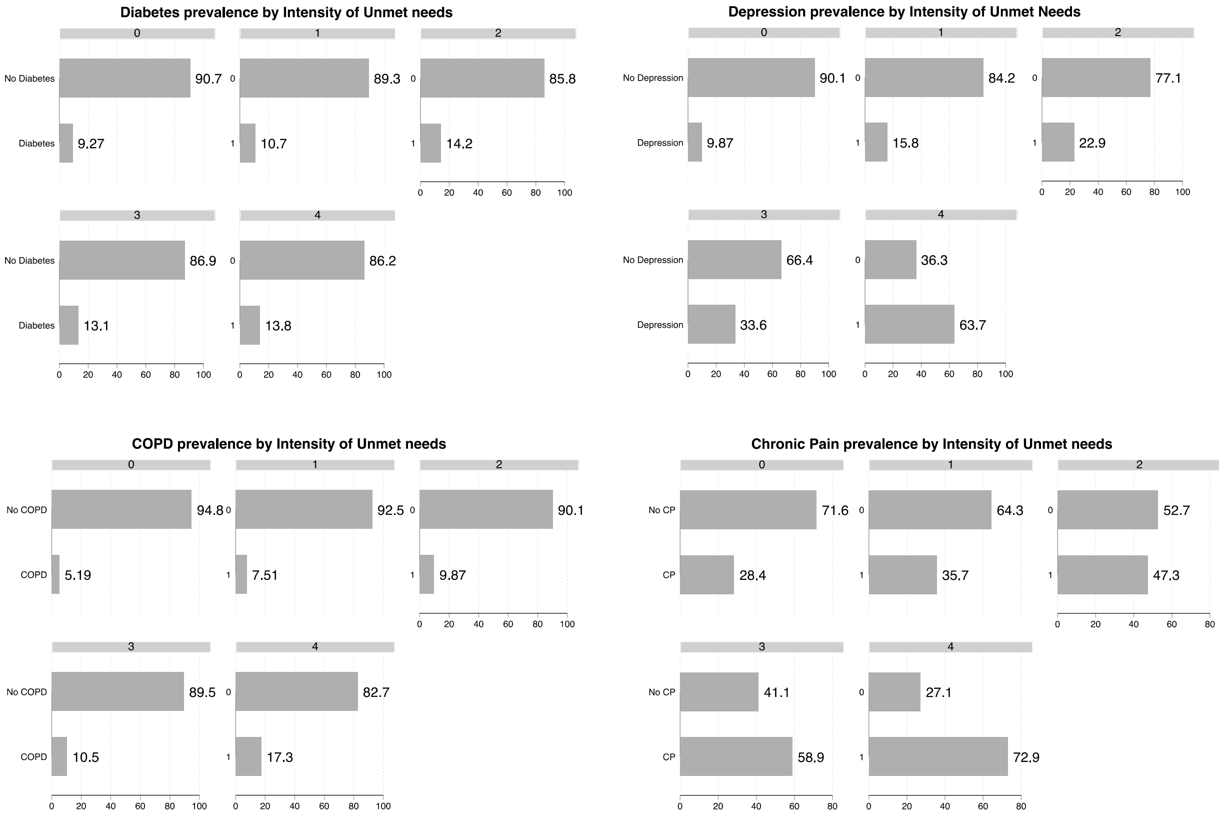
In what regards to the degree of urbanization, low-density areas seem to show a higher prevalence of the abovementioned diseases, apart from depression. Additionally, the difference between low-density areas and the other two categories is quite substantial. For instance, while 7.1% of the individuals living in low-density areas report having COPD, only 5.2% state having the same condition, in high-density areas. The exception of depression can most probably be explained by the increased pressure and stress that people from middle and high-density areas are usually subject to. These are the areas that best represent developed and modern societies. The fact that Lisbon has a higher depression prevalence rate also corroborates this premise.

In the diseases studied, there is a clear increasing prevalence pattern, as the number of unmet needs increases (Figure 8), with this being better perceptible in depression, for instance, where individuals with 0 unmet needs report the disease in 9.87% of the cases, and with 4, 63.65% presented depression (Table A2. 1).

Table 3 – *Prevalence of Chronic Health Problems across Population Sub-groups*

	Diabetes	Depression	COPD	Chronic Pain
Gender				
Female	9.4%	17.2%	6.7%	37.7%
Male	9.2%	5.9%	4.7%	20.7%
Quintile				
1st	10.5%	15.5%	7.2%	34.1%
2nd	12.8%	13.9%	7.5%	34.7%
3rd	9.0%	12.3%	5.5%	30.3%
4th	8.3%	10.4%	6.0%	26.9%
5th	6.2%	7.4%	2.9%	22.7%
Education				
Preschool	24.3%	19.4%	14.7%	55.3%
1st or 2nd cycle	15.0%	17.2%	7.4%	37.5%
3rd cycle	5.2%	8.5%	4.0%	21.8%
High-School	2.6%	7.0%	3.3%	20.3%
Post-Secondary	3.2%	3.9%	4.7%	27.0%
Higher Education	2.3%	6.8%	2.6%	20.0%
Region				
North	10.0%	11.7%	5.9%	29.6%
Algarve	8.3%	9.4%	3.9%	29.1%
Center	10.2%	12.0%	6.7%	32.6%
Lisbon	7.7%	12.8%	5.2%	28.8%
Alentejo	10.4%	10.7%	6.0%	30.0%
Azores	9.4%	12.8%	6.1%	25.9%
Madeira	8.3%	10.4%	4.9%	22.6%
Degree of Density				
Low Density	11.2%	11.7%	7.1%	32.7%
Medium Density	8.5%	10.5%	5.5%	28.0%
High Density	8.8%	13.0%	5.2%	29.1%

Figure 8 - ***Chronic Diseases' Prevalence by Intensity of Unmet Needs***



All in all, Table 3 and Figure 8 provide us with evidence that there is a pattern of disease concentration in the lower levels of the socioeconomic scale since lower income quintiles, lower education levels and a higher number of unmet needs correspond to higher disease prevalence, and thus, this information might be useful in the creation of our SES composite index.

Additional information that displays the differences between patients and non-patients is presented in *Appendix A.2. Additional Results, Facts and Characteristics about the NHS Sample and Chronic Patients,* where, for instance, the lower life satisfaction levels or lower self-reported health status are shown to be evident for chronic patients.

4. Methodology

In this section, the goal will be to introduce all the theoretical concepts and statistical tools that will be used in the analysis. As such, this fourth group acts as an introduction for the results section where the outcomes of the implementation of these methodologies will be presented and analyzed.

4.1. SES Variable: Construction and Implementation

The first objective in this section is to understand how to create an SES variable that can summarize the main given socioeconomic variables, and that, at the same time, can bring more information than just the individuals' income quintile.

4.1.1. Intermediate Variables

One must be able to understand the social and economic factors that should be included in the new summary variable. These will, from now on, be referred to as intermediate variables, since they act as a means of obtaining the final SES index. As already mentioned in the Data Analysis section, some NHS's variables seem to fit the purpose.

Undoubtedly, the household's equivalent income quintile must be present, since it is the only variable concerning **income**. It also seems relevant to include the variable specifying the individual's **education level**, since it is one of the most widely used variables to characterize SES.

Another pertinent variable is the **degree of urbanization** of each individual's region. In principle, areas where population density is higher have people with a higher SES. Obviously, this is not linear, as observed in the Data Analysis section, where depression appeared as a counterexample. Nonetheless, the assumption is that the probability of having someone from higher SES is larger for the more developed and more populated areas.

One could also mention the per capita relative purchasing power parity (per capita PPP) of each region. In the NHS of 2014, there is no information about each respondent's county. This said, we are only able to work with the variable that specifies to which NUTS II area individuals belong. According to INE and PORDATA (2020), in 2013[5], the per capita PPP was as the following table specifies:

Table 4 - ***Per Capita PPP by Region, in 2013***

	PPP, 2013
Lisbon	125.1%
Algarve	96.4%
North	92.0%
Alentejo	89.4%
Center	89.2%
Madeira	86%
Azores	84.6%

Source: Data for per capita PPP was obtained from *Instituto Nacional de Estatística* and PORDATA.
Notes: The values are in percentage, where 100% corresponds to the baseline mean value.

Only the Lisbon Metropolitan Area presents a value higher than the mean. Therefore, the chosen methodology was to create a **dummy variable to distinguish the Lisbon area inhabitants from the rest.**

Also, some of the available variables in the dataset are concerned with the possibility of each individual having unmet needs. Here, instead of using the previously created Intensity of Unmet Needs, a new variable will be created – **Unmet Needs** – that will only take the values of 0, 1 or 2, depending if the individual has any unmet need or not. Further details about this variable creation process are outlined in *Appendix B. Variable Creation Procedures.* A summary table of all the categorical variables used is present in Table C1. 1.

4.1.2. Statistical tools

MCA is the most suitable technique for the creation of the new summary variable, as hinted in the literature review. Although both PCA and MCA are useful to study the relationship between two or more variables, note that all the variables selected are non-continuous and categorical and MCA is described as the correct technique to use in the case when dealing with this type of data.

[5] Note that NHS data is from 2014. However, the closest years to which PPP was computed are 2013 and 2015. As such, 2013 PPP information was used, assuming that individuals lived in the same NUTS II region both in 2013 and 2014.

Even if it was the case that some of the variables were quantitative and continuous, the methodology advises the creation of bins, where different categories are generated according to the different values the continuous variable can take, and to which each individual will be assigned. Essentially, it is always possible to transform continuous quantitative data into a non-continuous categorical format, while the opposite is not true. As PCA is not suitable for categorical data and there are no tools to transform it into continuous data, MCA is therefore the best technique for this analysis.

Through this process, different components are created. These are summary indices with different intermediate variable weights. As such, each of them will have a different explanatory power, with the first component always being the one with the highest power. Therefore, being the combination of the intermediate variables that better acts as a summary variable. As such, the first component assigns to each intermediate variable the correct weight in the computation of the final SES measure. Thus, our final SES variable will be the first component of the MCA.

4.1.3. Implementation

Kohn (2012) is the main source of the implementation procedure. Using Stata 15.1., the first dimension of MCA is obtained, with a 64.65% explanatory power, i.e., it is able to explain 64.65% of the total variability in the data. The variable loads are displayed in Table 5, and the correspondent weights of each of the intermediate variables are visually displayed in Figures 9 and 10. Both entail the same information, although MCA's first dimension projection plot is a simplified version of MCA's biplot, which includes coordinates for the first two dimensions.

Here, we can understand how the different intermediate variables contribute to the MCA's first component. The lower the degree of density, the quintile or the level of education, the lower the final individual's SES value. Besides, if an individual resides in Lisbon, his SES value will be higher than otherwise. Furthermore, if he has not satisfied at least one of the four aforementioned needs this will have a negative contribution to his SES. The same information is understood through Table 5, where the output for the MCA procedure is displayed. The detailed information can be found in Table C1. 2. This was done through the use of the *mca* command in Stata 15.1.

Figure 9 - *Biplot of MCA Coordinates*

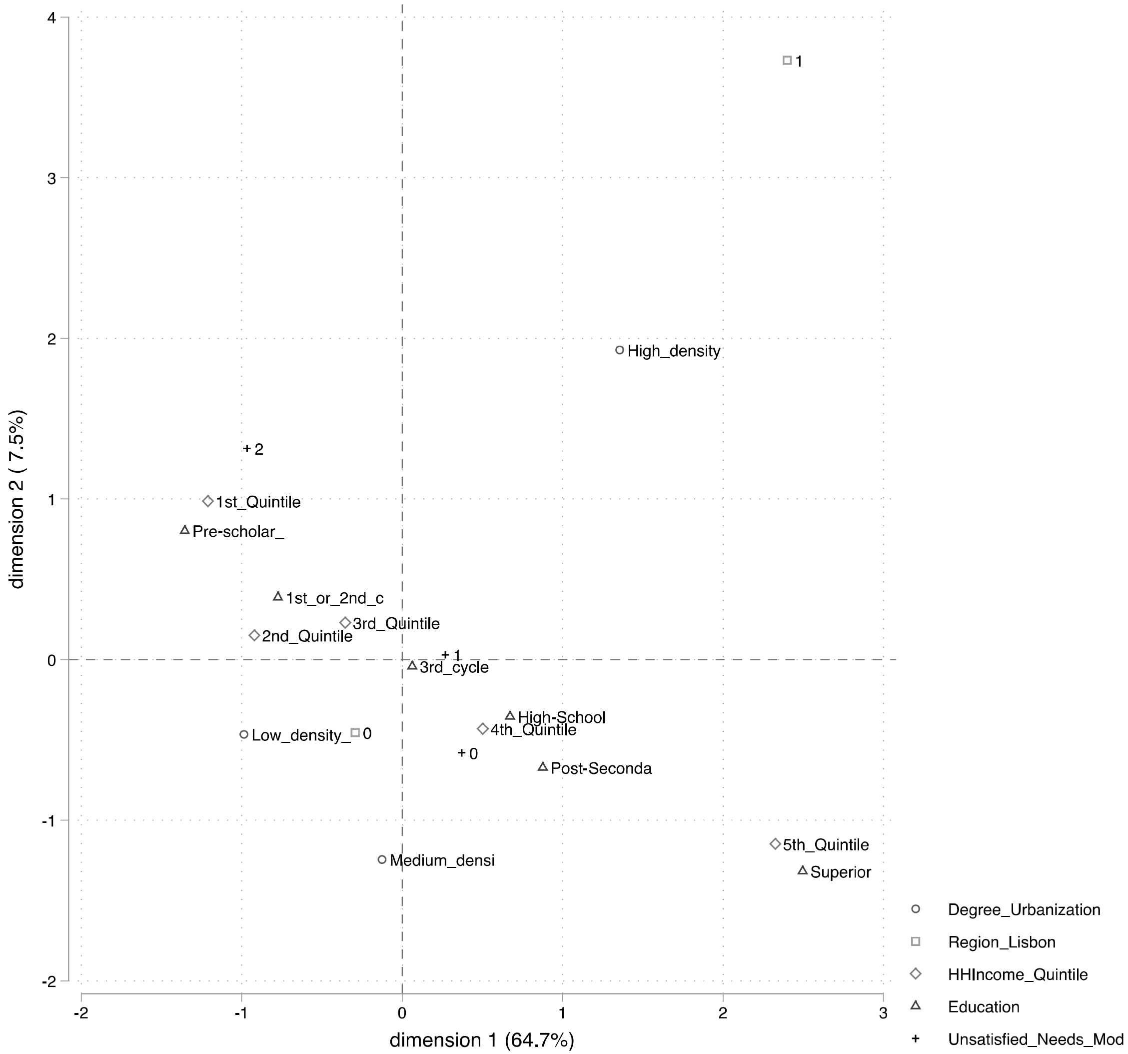


Figure 10 – *MCA's First Dimension Projection Plot*

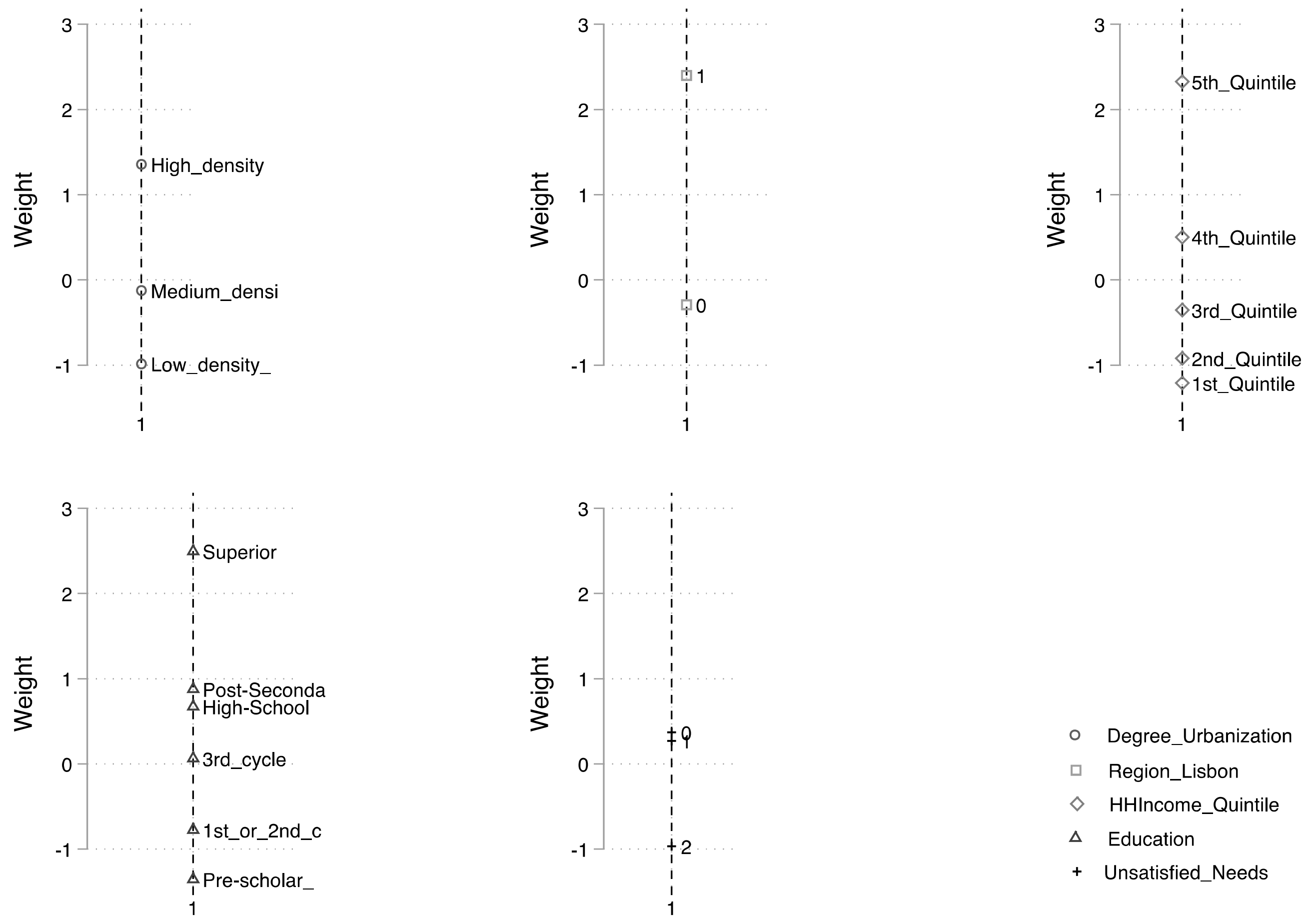


Table 5 - ***MCA Stata Output***

	Weight	Sq. Corr.
Degree of Urbanization		
High density	1.356	0.589
Medium density	-0.126	0.043
Low density	-0.985	0.800
Region_Lisbon		
Non-Lisbon	-0.292	0.560
Lisbon	2.400	0.560
HHI Quintile		
1st	-1.211	0.667
2nd	-0.921	0.747
3rd	-0.354	0.353
4th	0.503	0.415
5th	2.326	0.707
Education		
Preschool	-1.355	0.627
1st or 2nd cycle	-0.774	0.749
3rd cycle	0.063	0.028
High-School	0.673	0.575
Post-Secondary	0.877	0.417
Higher Education	2.496	0.702
Unmet Needs		
No unmet needs	0.371	0.657
No needs	0.270	0.282
At least one unmet need	-0.967	0.746

Note that whenever the weight given to a particular category is positive, it means that this category is a positive contributor to the index. The idea is that categories which are usually constituted by individuals from higher status in the socioeconomic scale should increase the final value, and vice-versa. Further explanation about this interpretation is presented in *Appendix C. Multiple Correspondence Analysis Procedure*.

The implementation of the procedure in Stata was done through the Burt matrix[6], which is adjusted by default. This is done because "*the inertia explained by the first dimension [in the unadjusted case] is severely underestimated*" (Khangar and Kamalja, 2017). As such, the main modification that using the adjusted Burt matrix entails is the change in the value of the

[6] There are different ways to implement MCA. Burt matrix, Indicator matrix or Joint Correspondence Analysis are some examples. In the end, the results of using each method might differ, except for Burt and Indicator method, which will lead to the same final weights. Their difference is mainly to which initial matrix is Correspondence Analysis applied. These technical details are not relevant to this work's main analysis. But, as a curiosity, we can mention that Table 6 is an example of an Indicator Matrix. Burt Matrix (C), in turn, is obtained through the Indicator Matrix (Z), through the following transformation: $C=Z^TZ$. For additional information, see Nenadic and Greenacre (2011).
Burt matrix is defined as being the most suitable method when using MCA as an analogue of PCA for categorical data.

dimensions' inertias and their predicting power, as one can understand by the differences between Table C1. 3 and Table C1. 4, in the appendix. Inertias, in this case, describe the part of the variation that each one of the dimensions explains from the total inertia (0.61601, in the unadjusted case), i.e., each inertia presents the variability explained by the corresponding dimension.[7]

The objective is then to obtain the SES value of each respondent. As Kohn (2012) explains, this should be done through the Stata command *predict*, whereby individuals are assigned their coordinate of the first dimension of the MCA method, based on the categories to which they belong in each of the intermediate variables. Inherent to this command is equation 1. It means that to obtain the coordinate of each individual's SES, we have to sum the weights of the category of each variable to which each individual belongs (those will have I^k_{i,j_k} equal to 1, whereas the value will be 0 when they do not belong) and divide the result by approximately 3.0425. This last result is based on the principal inertia of dimension 1 of the unadjusted Burt matrix, present in Table C1. 3 and its computation is explained in equation 2.

$$C_i = \frac{\sum_{k=1}^{K} \sum_{j_k}^{k} W^k_{j_k} I^k_{i,j_k}}{A}, \text{ where} \quad \textbf{(1)}$$

$\boldsymbol{C_i}$ is the coordinate estimate;
$\boldsymbol{K}$ the number of intermediate variables;
$\boldsymbol{W^k_{j_k}}$ the weight of each category, $\boldsymbol{j_k}$, of each intermediate variable $\boldsymbol{k}$.
$\boldsymbol{I^k_{i,j_k}}$, a binary variable that takes the value of 1 when the respondent belongs to category $\boldsymbol{j_k}$ and 0 otherwise.

$$A = K \times \sqrt[4]{principal\ inertia\ of\ dimension 1\ of\ the\ Unadjusted\ Burt\ Matrix} = 5 \times \sqrt[4]{\mathbf{.1371093}} \approx 3.0425 \quad \textbf{(2)}$$

$$C_i \times A = (\alpha_1 high_{den} + \alpha_2 med_{den} + \alpha_3 low_{den}) + (\beta_1 non_{Reg_{Lis}} + \beta_2 reg_{Lis}) + (\varphi_1 quin_1 + \varphi_2 quin_2 + \varphi_3 quin_3 + \varphi_4 quin_4 + \varphi_5 quin_5) + (\delta_1 pre + \delta_2 cycle_2 + \delta_3 cycle_3 + \delta_4 high\ Sc + \delta_5 post\_sec + \delta_6 superior) + (\lambda_1 non\_un + \lambda_2 no\_needs + \lambda_3 sats\_need) \quad \textbf{(3)}$$

[7] In this specific case, the first dimension explained 22.26% of the total variability in the unadjusted case, whereas the same number increased to 64.65% when using the adjusted matrix (Table C1. 3 and Table C1. 4).

Equation 3, in turn, is a different way of presenting equation 1. Each category corresponds to a binary variable that will take the value of 1 if the individual belongs to the category, and 0 otherwise. The parameters are, thus, the weights used to compute the coordinates of each respondent, which are present in Table 5.

Now, the methodology is easily explained through an example. Consider the following Indicator matrix, in Table 6, where 5 example observations and their categorizations are presented in matrix form. Note that the legend of the categorical variables and corresponding categories is introduced in Table C1. 1.

Table 6 - *Indicator Matrix of 5 Example Observations*

	Categorical Variables																		
	A					B			C		D			E					
Obs.	1	2	3	4	5	1	2	3	1	2	1	2	3	1	2	3	4	5	6
#1	0	0	0	0	1	0	0	1	1	0	0	0	1	0	0	1	0	0	0
#2	0	0	1	0	0	0	1	0	1	0	1	0	0	0	0	0	0	1	0
#3	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	0	0
#4	0	0	0	1	0	1	0	0	1	0	0	0	1	1	0	0	0	0	0
#5	1	0	0	0	0	0	0	1	1	0	1	0	0	0	0	1	0	0	0

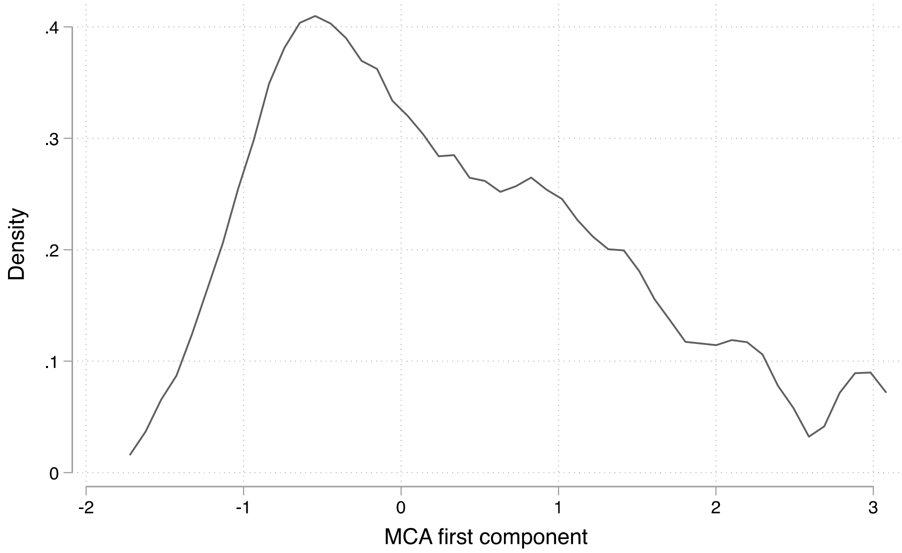
Consider observation #1. The respondent is someone from the 5th quintile, who has at least one unmet need, does not live in Lisbon, but still lives in a high-density area and is qualified with the 3rd cycle. This said, and remembering that the weights of each category are presented in Table 5 and Table C1. 2, the numerator should sum the weights given to someone who belongs to the 5th quintile (2.326), has at least one unmet need (-0.967), does not live in Lisbon (-0.292), lives in a high-density region (1.356), and completed the 3rd cycle (0.063).

So, the created index for this individual, based on equation 1, should be the following:

$$C_1 = \frac{1 \times 2.326 + 1 \times \ (-0.967) + 1 \times (-0.292) + 1 \times (1.356) + 1 \times 0.063}{5 \times \sqrt[4]{.1371093}} \approx \frac{2.486}{3.0425} \approx 0.8171$$

Now, since the values of the first component have been computed for all the observations, one can represent the Kernel Density estimates plot for our sample's MCA's first component (Figure 11). The distribution is, obviously, right-skewed just as income distributions are, which is an indication that this is, most likely, a valid variable. Remember that this is the usual pattern for the modern societies' household equivalent income distribution.

Figure 11 - ***Kernel Density Estimates for MCA's First Component***



4.1.4. Transformation and final SES measure

For now, this is what most approximates our final SES variable. However, it is still not adequate to use it in the computation of concentration indices, since it can take both negative and positive values (as its range is approximately from -2 to 3). This will be a problem in the computation of concentration indices, and especially SES's Gini, since it assumes SES is a positive value, as is the case of income. The procedure adopted was to add the absolute value of the lowest SES value to all the observations, which resulted in an SES variable with a zero-lower bound.

$$C_i^N = C_i + |\min(C_i)| \qquad \textbf{(4)}$$

At the last stage, in order to understand whether this is already a suitable variable for this analysis, one can compute its Gini Index. If the results are similar to the Portuguese income distribution Gini available from other sources, then this variable most likely encompasses the most critical elements of individual's SES and can be accepted. The Gini Index in Portugal, in 2014, was 0.34 (INE and PORDATA, 2020). This final SES variable's Gini is approximately 0.33 (Table 8). Moreover, as previously explained, the fact that its distribution is right-skewed increases this variable's plausibility as a good proxy for SES. As such, there is enough evidence to corroborate the use of MCA's first component after the transformation.

Figure 12 shows the final distribution of the SES variable. In Table 7, the summary statistics of the final SES variable are presented. Interestingly, by analyzing the differences between the group of respondents with none of the four studied chronic conditions and the group with at least one, it is obvious that the SES mean and median are lower for the second group, which is an indication of the possible negative correlation between SES and chronic disease prevalence.

Figure 12 - *Kernel Density Estimates for SES Final Variable*

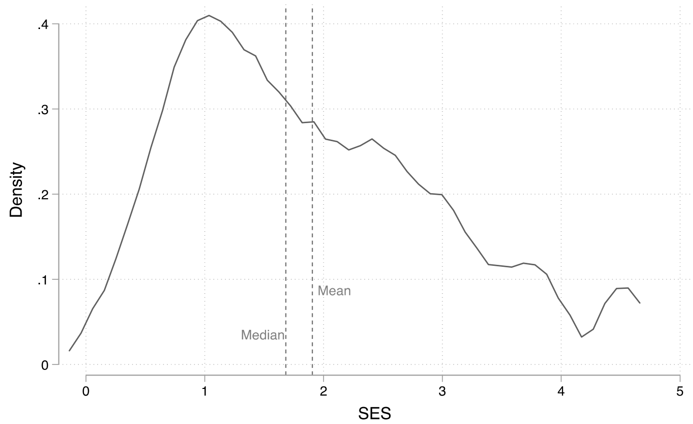


Table 7 - *SES Variable Statistics*

	SES variable			
	Mean	Median	Min	Max
All population	1.90577	1.681593	0	4.522052
Non-Chronic Patients	2.070872	1.94025	0	4.522052
Chronic Patients	1.661888	1.462069	0	4.522052

4.2. On the Measurement of Health Inequality

4.2.1. Concentration Indices

The methodologies that will be used to evaluate the prevalence of these diseases according to the SES, measured by the variable created above, will be the concentration indices and the correspondent concentration curves. Additionally, to obtain these relevant concentration indices, confidence values and respective confidence intervals, the approach will be to use the Bootstrapping technique, which will offer robust results.

In simple terms, bootstrapping is a computer-intensive statistical tool that uses resampling with replacement. This is, from the initial sample, a certain number of different samples is generated, whereby different estimates will be obtained. It is through them that standard deviations, confidence intervals and hypothesis tests are attained, even in cases without the usual assumptions (e.g., normality, sufficiently high number of observations), which is usually the case with complex survey data. Hesterberg (2011) succinctly explains:

> *"Bootstrap methods can be remarkably more accurate than classical inferences based on Normal or t distributions. The bootstrap uses the same basic procedure regardless of the statistic being calculated, without requiring the use of application-specific formulae." (Hesterberg, 2011)*

Our case is one of these complex survey data examples. There are different sampling weights, different information in the dataset which was recorded in different situations, and it is also not possible to guarantee the *a priori* distribution of the parameters. As such, the concentration indices' standard errors presented throughout this work will be based on this resampling technique.

Concentration indices measure the inequality of the chosen variable (which, in this case, will be the diseases' prevalence) based on the chosen ordering variable, which in our case will be the previously created SES measure. A common way of presenting it is through the concentration curve, where, on the x-axis, we have the accumulated share of the population ranked from lowest to highest, by the socioeconomic position, and, on the y-axis, we have the corresponding accumulated share of the health variable. A deeper explanation of these curves will be developed in the next subsection.

The concentration index was first introduced as a health inequality measure by Wagstaff et al. (1989). The main idea appeared in Kakwani (1977). Several different papers have used concentration indices to study inequality in different health-related studies. Some examples are Mangalore et al. (2007) in the study of mental health problems, van Doorslaer et al. (2004) focusing on inequalities in healthcare utilization or Wagstaff and van Doorslaer (2000) in the delivery of healthcare. Other studies focused on cross-country analysis, such as inequality in child mortality (Wagstaff, 2000). Notice that the well-known Gini Index is a special case of a concentration index.

To compute these concentration indices the general formula is the following (Kakwani, 1980):[8]

[8] The limits reported neglect the possibility of asymmetric sample weights, i.e. they assume all observations have the same weight.

$$CI = \frac{2}{n\bar{y}}\sum_{i=1}^{n}(y_i - \bar{y})(H_i - \bar{H}) = \frac{2}{y_m}cov(y_i, H_i) \quad \textbf{(5)}$$

where y is the socioeconomic status measure (previously referred to as C_i^N), and H the disease's prevalence, n the number of observations and $\bar{y}$ and $\bar{H}$ being, respectively, their means.

However, note that in this case we have weighted data, as such, Lerman and Yitzaki (1989) state that the concentration index can be obtained as the following:

$$CI = \frac{2}{n\bar{y}}\sum_{i=1}^{n} w_i(y_i - \bar{y})\big(\hat{H}(y_i) - \bar{H}\big), \text{ where} \quad \textbf{(6)}$$

$\sum_{i=1}^{n} w_i = 1$, $\bar{y} = \sum_{i=1}^{n} w_i y_i$ and $\hat{H}(y_i)$ is obtained through:

$$\hat{F}(y_i) = \sum_{j=0}^{i-1} w_j + \frac{w_i}{2}, \text{ where } w_0 = 0 \quad \textbf{(7)}$$

These were used to construct a program in Stata 15.1., that effectively delivered concentration index estimates with the corresponding bootstrapped standard error.

Alternative ways of attaining the concentration index value can be presented. For instance, we can achieve the same concentration index by using a convenient regression.

$$\frac{2\sigma_R^2}{y_m} y_i = \alpha + \beta R_i + \varepsilon_{1,i} \quad \textbf{(8)}$$

Where σ_R^2 is the variance of the disease prevalence and β the estimated index.

Additionally, in the case of equally-weighted observations, we can also mention that these indices can take values between $\frac{1-n}{n}$ and $\frac{n-1}{n}$, where the first number is negative, whereas the second is positive. When the number is negative, it means that there is a higher concentration of the prevalence of the health problem amongst the lower socioeconomic levels of the population. On the contrary, when the number is positive, there is a higher concentration of the health problem among the higher socioeconomic levels of the population.

Notice also that, when the index takes exactly the minimum value ($\frac{1-n}{n}$), we are in a maximum inequality scenario, meaning that all the disease is accumulated in the lowest SES observation. When the value is equal to its upper bound ($\frac{n-1}{n}$), health is again maximally unequally distributed, but now, with the disease concentrated only in the highest SES member of the population. It is also worth mentioning that when a variable is equally distributed amongst the population, the index value is evidently zero, because it corresponds to a zero inequality situation.

4.2.2. Concentration Curves

The use of concentration curves comes as a result of the relevance of concentration indices. They visually display the degree of concentration computed through the concentration index, so as to help getting a complete picture. As previously mentioned, they exhibit the population, organized from the lowest to the highest ordering variable, in the x-axis, which, in our case, will be the SES variable. In turn, in the y-axis, the objective is to plot the cumulative percentage of the study variable of each x% of the population, ordered by the SES variable. In this case, the goal will be to plot the prevalence of the diseases studied per each cumulative share of the population ordered by increasing SES. This big group of concentration curves accounts for special cases, like the well-known Lorenz Curve, which studies the share of income each percentage of the ordered population, according to incomes, owns.

The basic idea is to visually represent the difference between the degree of concentration of the variable being analyzed and the equality line, which corresponds to the 45º line. This line represents a completely equally distributed study variable. Remember that this also corresponds to the case where the concentration index equals zero.

As such, one will plot the concentration curves of each of the four diseases and compare them with the equality line. The concentration index corresponds to twice the area between the concentration curve and the equity line.

The concentration curve tends to be above the equality line when the concentration index is negative. On the other hand, when the value is positive, one should expect the corresponding curve to be below the reference line. Furthermore, the main idea is that the higher the curve is in the graph, the lower the value of the concentration index. In absolute terms, the analysis is different. The further away from the reference line, the higher is the concentration index, in absolute value.

In this specific case, the hypothesis we are trying to test is that there is a higher concentration of health problems amongst the lower SES individuals. This means that one should expect to have a significant and negative concentration index, with concentration curves above the equality line.

5. Results

5.1. Concentration Indices and Curves Results

In this section, the results of the concentration indices for the prevalence of the four chronic conditions and for the SES will be presented. These will also be displayed in the corresponding concentration curves.

First, one should compute the Gini (which is the concentration index) of our SES variable. This is 0.3258 (Table 8). Besides the Gini, there is also information about the deciles and interpreting at least one of these is useful to further understand the dispersion pattern of this variable. Focusing on decile 5, for instance, one can say that the 50% of the population with the lowest SES, i.e., the bottom half, accounts only for approximately 26% of the total SES measure. This means that if one was to add up the values of the SES for all the individuals, this lower half would only have 26% of that total. Clearly, this means that the top half would have approximately 74%. A similar analysis can be done for any of the other deciles.

Table 8 - *SES Concentration Index (Gini) and Concentration Curve Deciles*[9]

		SES			
		Observed Coef.	Bootstrap Std. Err.	Normal-based [95% Conf. Interval]	
Concentration Index		**.3258**	.0021	.3218	.3298
Deciles	1	.0204	.0004	.0196	.0212
	2	.0615	.0008	.0600	.0630
	3	.1151	.0011	.1130	.1172
	4	.1819	.0014	.1792	.1846
	5	.2633	.0016	.2602	.2664
	6	.3620	.0018	.3585	.3655
	7	.4806	.0019	.4769	.4843
	8	.6203	.0017	.6169	.6236
	9	.7862	.0012	.7839	.7885
	10	1	.	.	.

Afterward, the goal is to use the constructed variable to analyze the degree of concentration of the chronic diseases. Now is the time to answer the main question of this research: What is the

[9] Note: All of the observed coefficients, in the SES variable and in the four chronic diseases are statistically significant with all their p-values closel to zero. The complete tables for these variables can be found in the tables in *Appendix E.1. Unstandardized Results.*

relationship between the distributions of chronic diseases and the distribution of the SES levels, in Portugal?

In a first analysis, it is expected that the distributions of chronic diseases are more concentrated among the lower status individuals due to the negative values estimated for the concentration indices, as can be seen in Table 9.

In this particular case, COPD has a concentration index of -0.2299677, chronic pain -0.1339939, depression -0.1491089 and diabetes -0.2350278 (Table 9). Crudely analyzing, this means that diabetes is the condition with the most unequal SES-related prevalence pattern. All of these have p-values close to zero and are, thus, statistically significant. As such, we confirm that chronic diseases are unequally prevalent in the population regarding SES and that there are relatively common patterns for their unequal prevalence and correlation between SES and their incidence.

However, as one might imagine, further details are helpful to better interpret these measures. To do so, additional data is presented with the deciles for each of these conditions. Therefore, throughout this analysis, the idea will always be to focus on the 50% share threshold, i.e., decile 5, as it facilitates interpretation and comparison. However, as previously mentioned, what is analyzed at the 50% level is also valid for any other decile.

First, from Table 9, in the case of diabetes, it can be found that the lowest 50% of SES individuals have approximately 67% of the total diabetes prevalence. Regarding depression, the same 50% of the population has, in turn, 59.5% of the disease. As for COPD, the same 50% account for 65.8% of the prevalence. Lastly, chronic pain is distributed more equally, with the lowest 50% gathering 58.2% of the pain prevalence. Figure 13 displays the concentration curve for the four conditions and the SES variable. Remember that we should expect the concentration curve of the diseases to be above the equality line and higher, when the concentration index value increases (in absolute value). On the other hand, SES's curve should be below the equality line, as is typical of Lorenz curves for income, consumption or wealth.

Figure 14 corroborates the hypothesis of a negative correlation between SES and disease prevalence, with relatively similar patterns for the four diseases. Figure 15 is a graphical representation, using depression and SES as examples, of the concentration curves and respective confidence intervals, proving that our estimates for concentration indices and concentration curves are precise, due to the low confidence intervals displayed.

Table 9 - *Chronic Conditions Concentration Indices and Concentration Curve Deciles*

		Diabetes		Depression		COPD		Chronic Pain	
		Observed Coef.	Bootstrap Std. Err.	Observed Coef.	Bootstrap Std. Err.	Observed Coef.	Bootstrap Std. Err.	Observed Coef.	Bootstrap Std. Err.
Concentration Index		**-.2350278**	.017634	**-.1491089**	.0167972	**-.2299677**	.0244931	**-.1339939**	.0098587
Deciles	1	.1871703	.0110981	.1603365	.0090328	.2036733	.0142659	.1618131	.0048117
	2	.3246876	.0137006	.2787655	.0112281	.3451053	.0180789	.2863588	.0065579
	3	.4511807	.0150821	.3995794	.0132796	.4798639	.0203044	.3973275	.0079252
	4	.5547655	.0157146	.5037183	.014708	.5705507	.0205884	.4977572	.0084879
	5	**.6673498**	.0157442	**.5944156**	.0148444	**.6547002**	.0208692	**.5856574**	.00839
	6	.7534064	.0152052	.6871211	.0148059	.7288603	.0203295	.6740368	.0086417
	7	.8429813	.0126233	.7924419	.0133908	.8024219	.018896	.768678	.0078455
	8	.9133513	.0109417	.8803537	.0112115	.9014426	.0157423	.8588971	.0068495
	9	.9655601	.0065369	.9456624	.0079061	.9541757	.0105578	.9316588	.0052886
	10	1	.	1	.	1	.	1	.

Note: Standard errors estimated by bootstrapping.

Figure 13 - ***Concentration Curves: Chronic Diseases and SES Variable***

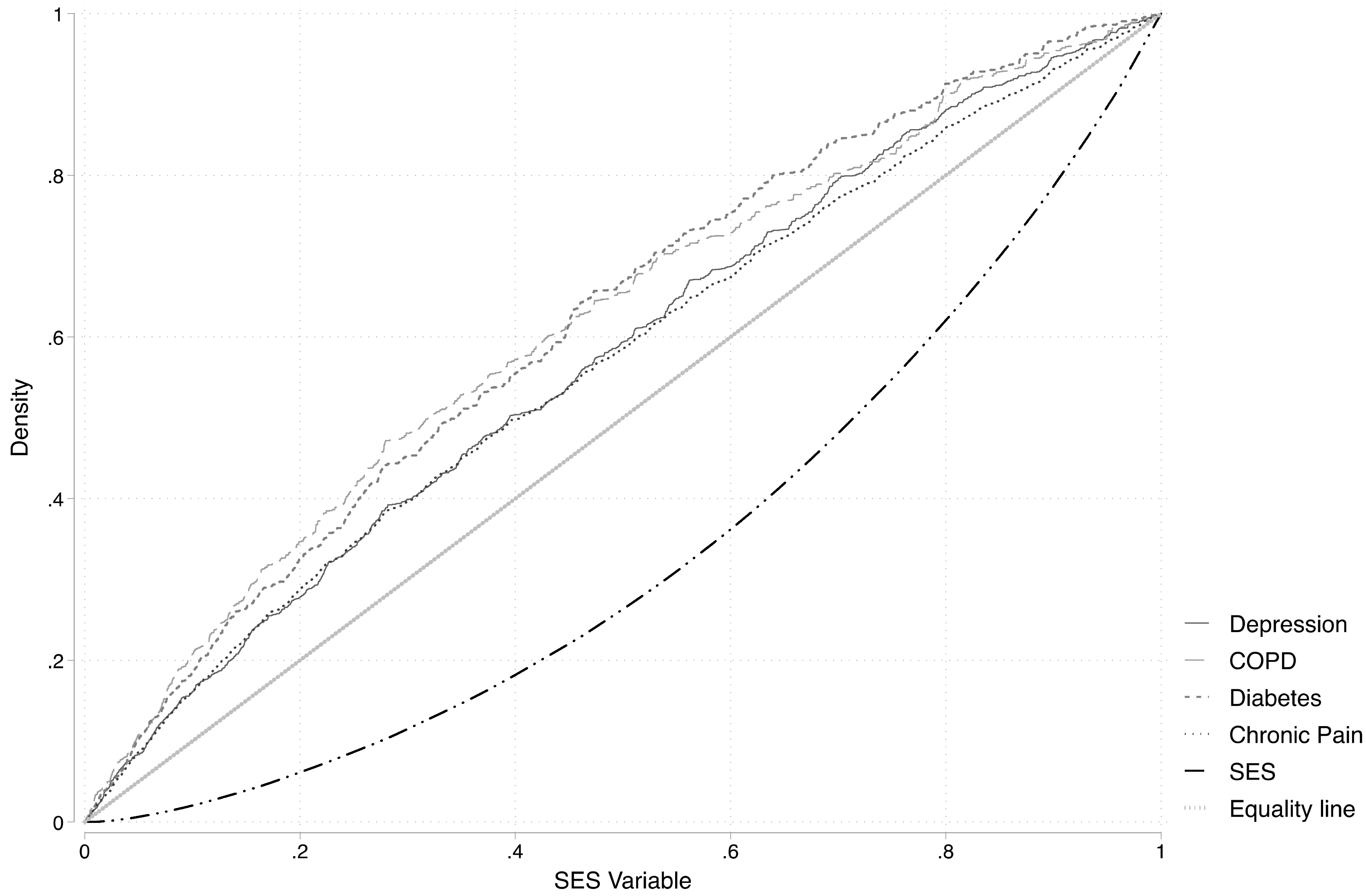
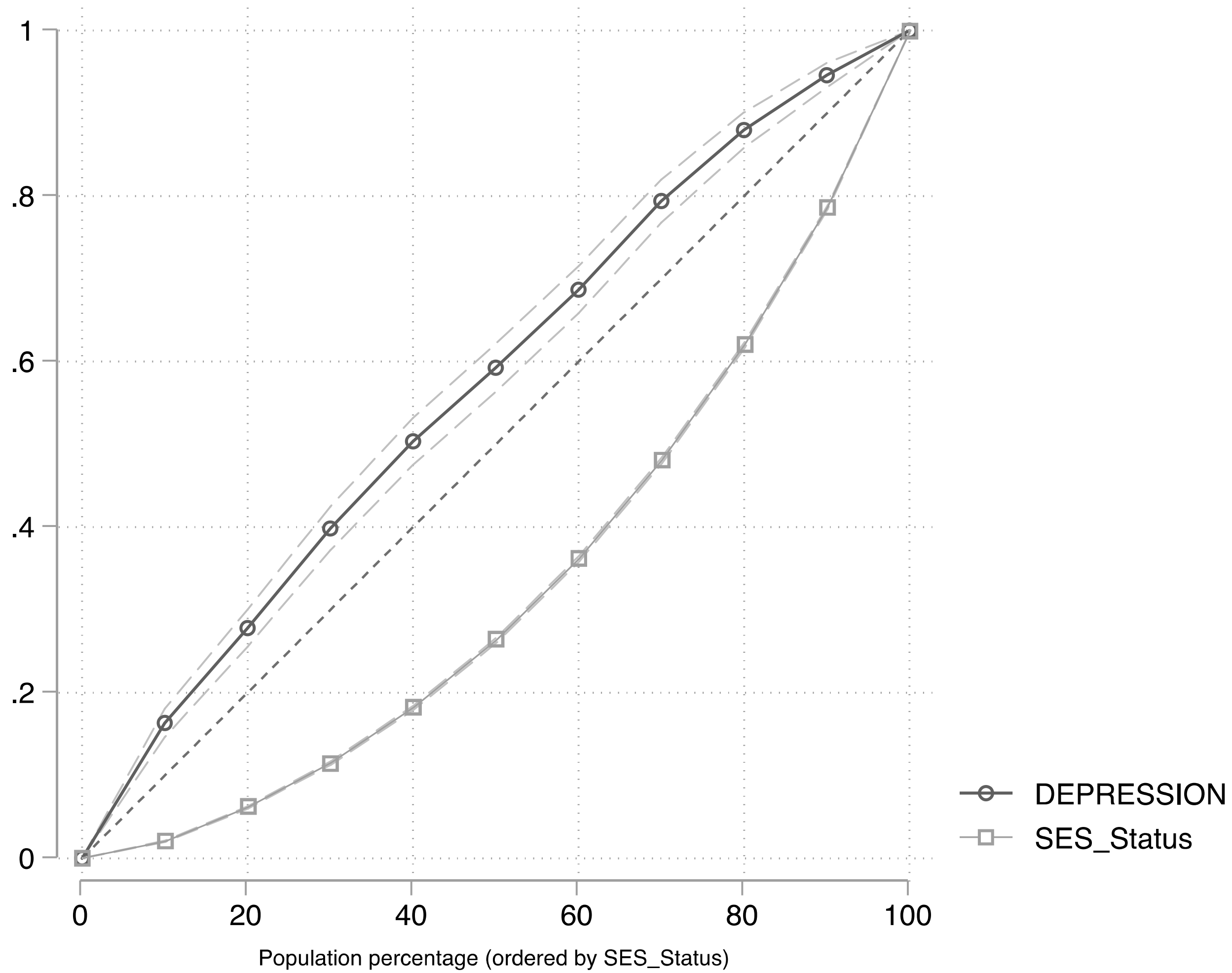


Figure 14 - ***Depression and SES Concentration Curves with Respective Confidence Intervals***



Note: Confidence intervals for SES are so narrow, that we are hardly able to perceive them at this scale.

5.2. The Case of Evitable SES-related Inequality in Health

5.2.1. Motivation

What has been done so far is enough to answer this work's main research question. Indeed, the goal was to understand whether there is any evidence of SES-related inequality in the prevalence of the studied diseases. The answer is straightforward: yes.

Nonetheless, there is a vast amount of literature focusing on the idea of whether this inequality can be considered fair or unfair. In other words, part of this lack of equity may be due to inevitable characteristics, possibly not amenable to correction by public policies. For example, from *Section 3*, there is evidence that older individuals are more prone to have these chronic diseases and if, additionally, one suspects that they are also more likely to belong to lower SES, this directly means that part of the inequality we had considered in the previous analysis was inevitable, in a sense. Moreover, this is likely not to be considered unfair, in the framework of this analysis, since almost all individuals will reach that point in their lives.

In simplified terms, the objective is to distinguish between policy-relevant or irrelevant characteristics. These are differentiated through the possibility of policy implementation influencing them or not. The general consensus in the literature is that the most notorious case is the one of demographic characteristics, in particular, age and gender, since these are unavoidable and, thus, should not be accounted for when discussing unfair or evitable inequity (e.g., Hong et al., 2011; Mangalore et al., 2007).

Note that, in our case, defining demographics as policy irrelevant is slightly more doubtful. Even if there is evidence of having a larger share of older individuals or of one of the genders with these diseases and, additionally, belonging to lower socioeconomic levels, it does not mean public policy implementation cannot influence this outcome. This is indeed possible. For this reason, throughout this section, the choice is not to name these demographic characteristics as policy irrelevant, but rather to distinguish between "evitable" and "inevitable" shares of inequality in health, which can still be a slightly doubtful definition, again because of the argument that perhaps these demographic inequalities can be addressed by public policy.

That is why this section is only an addendum to the previous analysis, because **all** the previously computed inequality indices are relevant for our policy implementation purposes. This follows closely what Gravelle (2003) mentions in his research work: "*if the direct effect of age on health* *can be altered or income redistributed by age, the degree of income-related health inequality can be changed.*"

The methodology presented to account for this distinction is known as demographic standardization. Its main goal is to level out the effect of standardizing variables. There are two methods in the literature: direct and indirect standardization. In this work, the second method was used, as it is the most recommended in previous works, due to its advantages of not requiring the use of grouped data (e.g., O'Donnell et al., 2007; Mangalore et al., 2007; Wagstaff and van Doorslaer, 2000).

The main objective of such a methodology is to obtain "*the distribution that would be observed if all individuals had their own age but the same mean age-gender effect as the entire population*" (Mangalore et al., 2007). We are trying to equalize the population, so as to only capture the effect of SES, setting age and gender constant.

The detailed procedure of the adopted methodology can be found in *Appendix D. Indirect Standardization Procedure* and follows the work of Gravelle (2003). Note also that, in Figure 15, we can find the odds ratio for each one of the age groups, for each of the diseases. These were based on the logistic regressions used in the Indirect Standardization Procedure (Table D1. 1 and Table D1. 2), where health outcomes were regressed on dummies for gender and for *n-1* age groups were used[10]. This is extremely useful to confirm that indeed there is an unequal displaying pattern of each one of the diseases, based on the respondents' age. In general, as age increases, individuals exhibit a general increasing odds ratio pattern.

5.2.2. Results

The results in Table 10 show that there is indeed a considerable share of inevitable inequality, based on gender and age. As such, results for the share of avoidable inequality, though still consistent with the hypothesis that lower SES is coincident with the prevalence of each one of the chronic conditions, display lower concentration indices. This immediately shows that a large share of the total inequality presented before is due to inequalities in age and gender. Now, it is COPD that presents a more unequal pattern, after deleting the effect of age and gender and

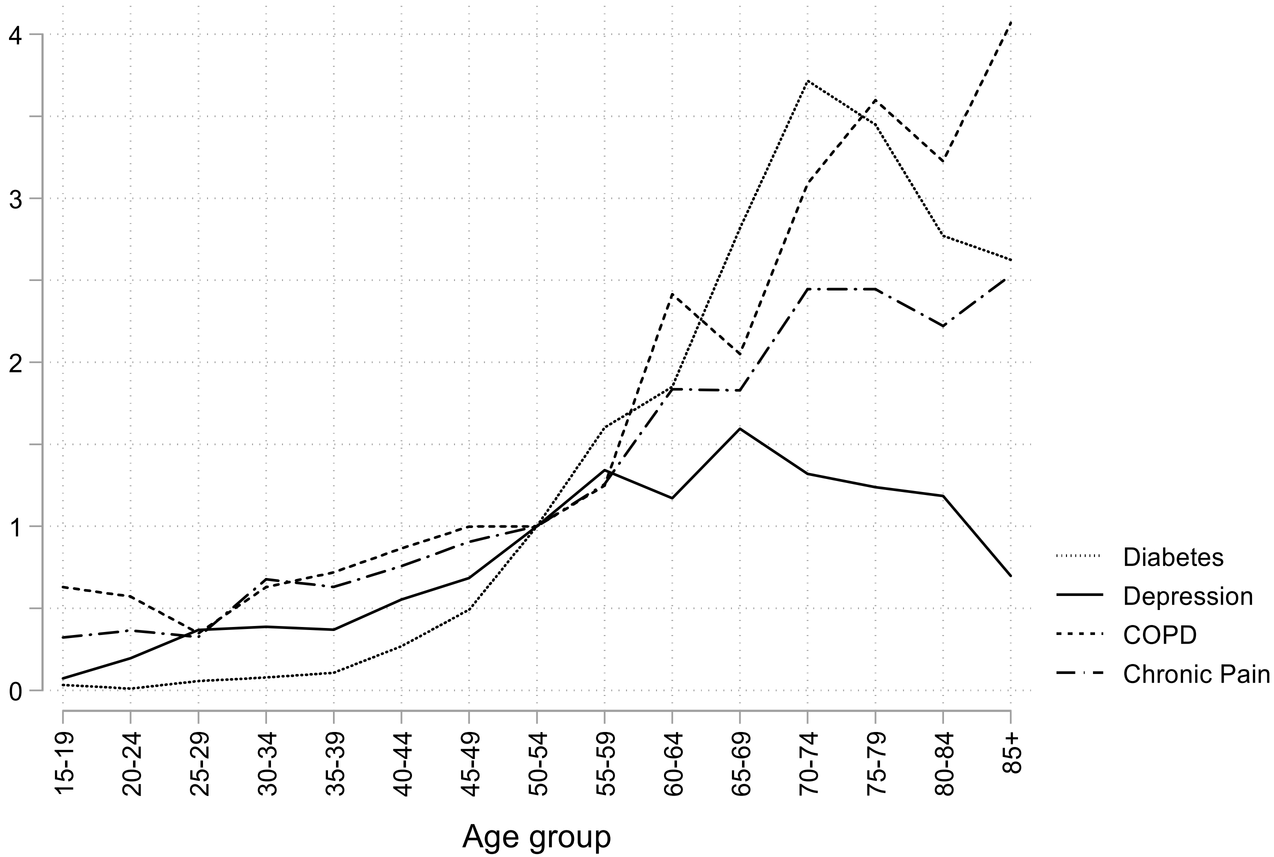
[10] Further details about these regressions are also outlined in *Appendix D. Indirect Standardization Procedure.*

depression is now the least unequally distributed. The complete concentration indices results are presented in *Appendix E.2. Standardized Results.*

Nonetheless, one should not deviate from the initial focus of this research work. As previously explained, the result of this thesis' main objective was presented in *Section 5.1*. Nothing in *Section 5.2* discredits the previous numbers and conclusions. This section acts only as a complement to what was previously done, rather than a substitute. This is useful to better disentangle the possible different effects playing a role in the bigger picture of inequality in chronic disease prevalence.

That is, the goal was always to understand whether lower levels of SES displayed higher levels of diseases prevalence, regardless of that prevalence being due to age, gender or any other factor inequalities.

Figure 15 - *Odds Ratio for Age Groups*



Note: Baseline is age group 50-54. Remember that the interpretation goes as follows: if age 65-69 for depression presents an odds ratio of approximately 1.55, it means that this group of people has 55% higher chances of displaying depression, than the baseline group (50-54). If instead, for COPD, the age group 20-24 presents approximately 0.55, it immediately means that these are 45% less likely to display COPD.

Table 10 - ***Standardized Chronic Conditions Concentration Indices and Concentration Curve Deciles***

		Diabetes		Depression		COPD		Chronic Pain	
		Observed Coef.	Bootstrap Std. Err.	Observed Coef.	Bootstrap Std. Err.	Observed Coef.	Bootstrap Std. Err.	Observed Coef.	Bootstrap Std. Err.
Concentration Index		**-.0377736**	.0053214	**-.0336145**	.006049	**-.0492651**	.0079267	**-.0388855**	.0044311
Deciles	1	.110402	.0033729	.1100687	.0034623	.1190358	.0046807	.1153285	.0020798
	2	.2107701	.0042277	.2109929	.0042574	.2265833	.0057145	.2214026	.0028304
	3	.3193981	.0047092	.3177739	.0049942	.3385138	.0063939	.324993	.0033626
	4	.4235988	.0047754	.421939	.0053977	.4389989	.0067395	.4289541	.0037315
	5	.5307434	.0047495	.5202708	.0053264	.5352784	.0069543	.5248068	.003953
	6	.6289522	.0044528	.6219245	.0052866	.6284638	.006754	.6226996	.0038759
	7	.728356	.0038927	.7265086	.0048749	.7207816	.0062131	.7232789	.0037798
	8	.8208852	.003189	.8242777	.0040819	.8230501	.0052308	.8200885	.0031939
	9	.9148853	.0022435	.9158344	.0031582	.9140447	.0036449	.9127356	.0025706
	10	1	.	1	.	1	.	1	.

Note: Standard errors estimated by bootstrapping.

6. Analysis

6.1. Discussion

The evidence presented throughout this work confirms the existence of a chronic disease equity gap in Portugal. The burden of chronic conditions is more concentrated amongst the lower SES individuals, i.e., there is evidence that supports the idea that being from a lower SES increases people's probability of reporting some of these diseases. There is more than one possible interpretation for the findings, based on different causal relationships.

A first hypothesis might be that there is a problem with access to health care. Perhaps, the individuals with a lower SES are more likely to suffer from a lack of medical care, including prevention. This may increase the prevalence of chronic diseases. This causal pathway may also be due to worse health behaviors adopted at the lower levels of the socioeconomic continuum.

However, the causality may be reversed. It may be the case that lower levels of health lead to lower productivity of human capital or even lower labor market presence, generating situations characterized by low SES.

Nevertheless, remember that the SES created variable does not include income measures only. It also concerns social status indicators and, thus, these should not, in principle, be influenced by changes in income. Education is a clear example. This means that our summary measure is an advantage, in this case. Using solely income as an SES measure increases the probability of being subject to the possibility of lower SES being caused by the chronic condition and not the opposite. This also prevents us from entering a spiral, where, for instance, an individual from a lower SES is diagnosed with diabetes and, then, its productivity decreases with the course of the disease, meaning that his disposable income might decrease even further.

In short, the above evidence suggests that there is a need to strengthen the implementation of chronic disease prevention strategies, due to the conditions' higher prevalence in lower SES. However, further research should be carried out in order to fully characterize the most probable cause-consequence associations in this field. If these policies are effectively implemented, and knowing that chronic patients do display lower levels of life satisfaction and health status, it is possible to foster their quality of life.

6.2. Public Policy Implementation

The conclusions presented above are deeply relevant, as having more information on the prevalence of chronic diseases will most likely lead to the improvement of public policies. Campos-Matos et al. (2016), citing previous works, state that it is probably the lack of studies displaying evidence of Portuguese health inequalities that causes them to remain unaddressed by public policy. This paradigm has been changing, with an increasing number of reports providing evidence in this matter. This is where our hope of having a strong evidence-based study enters.

Note that Portugal is one of the developed countries displaying higher health inequalities (e.g., Mackenbach et al, 2008). WHO (2010) has made several remarks about the urge to reduce these inequalities as Portugal has not been giving them their due importance. However, data from previous NHS editions shows that, in previous years, these inequalities were already evident (e.g., Pereira and Pedro, 2004). Moreover, they have increased between 2000 and 2014, as reported by Dimitrovová et al. (2017).

This is somewhat puzzling due to the high commitment the National Health Sector makes in its description: "*it should guarantee the provision of universal, general, and tendentiously free health care services*"[11] (Nunes and Ferreira, 2018).

Nonetheless, this evidence shows that this universal National Health Service is clearly not enough. Campos-Matos et al. (2016) argue that this indication may be a result of, among other factors, "*the engrained belief that the National Health Service, as a universal and (relatively) inexpensive service at point of care, is enough to face these inequalities.*"

Furtado and Pereira (2010) point out some of the possible reasons for these inequalities. They advocate that some factors might condition public service utilization, such as the inequalities in the probability of attaining a consultation or in the probability of performing preventive exams. Other factors were also presented as a justification for the astonishing health inequality in Portugal. Van Doorslaer et al. (2004) have also uncovered evidence of inequalities in health service utilization in several countries. Many other explanations can be found for these inequalities.

---

[11] Note that all Portuguese residents can access the National Health Service and that these are mainly financed through taxation, with the exception of a case where a copayment, called moderating fees, occur.

Be that as it may, the absolute truth that this data presents us is that inequality is there, and thus, either a redesign or reimplementation should be completed, in order to better manage these policies and to reduce the burden of chronic diseases in the lower levels of the socioeconomic scale. To do so, additional research should be fostered in order to better encounter the primary causes of these results.

7. Conclusion

Persistent socioeconomic health inequalities are clearly demonstrated in health literature, with Portugal being one of the most unequal countries. This study examines the prevalence of several chronic diseases in Portugal and their possible relation with the distribution of SES. This studies' results are aligned with these main conclusions in the literature, as we have corroborated that, in Portugal, there are relevant socioeconomic inequalities in the prevalence of chronic diseases. Additionally, these show to have a similar pattern regardless of the disease and its nature.

As such, they are a much needed boost to prove the need for policy redesign, in order to better assist vulnerable population groups. It is indeed a fact that both low SES and chronic health problems prevalence are positively correlated. As such, further research, investment and effort need to be carried out to develop this information, to understand the source of these inequalities and what are the best policies to manage them.

An advantageous innovation in this research comes from the use of a new SES variable. This SES index was created using MCA to summarize a set of variables in the NHS of 2014 and it displayed the desirable characteristics that an SES measure should present.

Nonetheless, one must keep in mind that this study also entails some limitations. The most important bears with the fact that one is dealing with self-reported health data, which can involve several types of bias, such as recall and social desirability bias.[12] Moreover, a second limitation may be concerned with our focus on a specific time span, meaning that we cannot be sure about the robustness of the results based on 2014 alone. However, throughout this research, no evidence pointed to the direction that this would significantly impact our results.

In summary, we believe this is a comprehensive study, focusing on four main diseases and enabling a deeper knowledge of their distribution characteristics. The evidence is clear and it is a wake-up call for policies to mitigate this socioeconomic gap and to reduce the burden of diabetes, depression, COPD and chronic pain, particularly on the lower levels of the socioeconomic scale. For that, additional research should be implemented to better design causal-pathways for the results found.

---

[12] Recall bias is related to all the inaccurate responses of individuals, due to the time lag between the occurrence and the reporting of an event. Desirability bias, in turn, concerns the desire of responding to what is perceived as good.

8. References

Abdi, H., & Valentin, D. (2007). Multiple Correspondence Analysis. *Encyclopedia of Measurement and Statistics*.

Andersen, I., Thielen, K., Nygaard, E., & Diderichsen, F. (2009). Social inequality in the prevalence of depressive disorders. *Journal of Epidemiology and Community Health*. https://doi.org/10.1136/jech.2008.082719

Anderson, G., & Horvath, J. (2004). The growing burden of chronic disease in America. *Public Health Reports*. https://doi.org/10.1016/j.phr.2004.04.005

Asselin, L.-M. (2002). Composite Indicator of Multidimensional Poverty. *Analysis of Multidimensional Poverty Theory and Case Studies*.

Azevedo, L. F., Costa-Pereira, A., Mendonça, L., Dias, C. C., & Castro-Lopes, J. M. (2016). The economic impact of chronic pain: a nationwide population-based cost-of-illness study in Portugal. *European Journal of Health Economics*. https://doi.org/10.1007/s10198-014-0659-4

Azevedo, L. F., Costa-Pereira, A., Mendonça, L., Dias, C. C., & Castro-Lopes, J. M. (2012). Epidemiology of chronic pain: A population-based nationwide study on its prevalence, characteristics and associated disability in Portugal. *Journal of Pain*. https://doi.org/10.1016/j.jpain.2012.05.012

Beagley, J., Guariguata, L., Weil, C., & Motala, A. A. (2014). Global estimates of undiagnosed diabetes in adults. *Diabetes Research and Clinical Practice*. https://doi.org/10.1016/j.diabres.2013.11.001

Blazer, D. G., & Hybels, C. F. (2005). Origins of depression in later life. In *Psychological Medicine*. https://doi.org/10.1017/S0033291705004411

Blumer, D., & Heilbronn, M. (1982). Chronic pain as a variant of depressive disease the pain-prone disorder. *Journal of Nervous and Mental Disease*. https://doi.org/10.1097/00005053-198207000-00001

Brown, K., Nevitte, A., Szeto, B., & Nandi, A. (2015). Growing social inequality in the prevalence of type 2 diabetes in Canada, 2004–2012. *Canadian Journal of Public Health*. https://doi.org/10.17269/CJPH.106.4769

Burney, P. G. J., Patel, J., Newson, R., Minelli, C., & Naghavi, M. (2015). Global and regional trends in COPD mortality, 1990-2010. In *European Respiratory Journal*. https://doi.org/10.1183/09031936.00142414

Campos-Matos, I., Russo, G., & Perelman, J. (2016). Connecting the dots on health inequalities - A systematic review on the social determinants of health in Portugal. In *International Journal for Equity in Health*. https://doi.org/10.1186/s12939-016-0314-z

Centers for Disease Control and Prevention. (2020). Definition of Chronic Condition. Accessed on 13th September 2020 on: https://www.cdc.gov/chronicdisease/about/index.htm

Chapmann, K. R., Mannino, D. M., Soriano, J. B., Vermeire, P. A., Buist, A. S., Thun, M. J., Connell, C., Lee, T. A., Miravitlles, M., Aldington, S., & Beasley, R. (2006). Epidemiology and costs of chronic obstructive pulmonary disease. In *European Respiratory Journal*. https://doi.org/10.1183/09031936.06.00024505

Cortinovis, I., Vella, V., & Ndiku, J. (1993). Construction of a socio-economic index to facilitate analysis of health data in developing countries. *Social Science and Medicine*. https://doi.org/10.1016/0277-9536(93)90127-P

Costa, P. S., Santos, N. C., Cunha, P., Cotter, J., & Sousa, N. (2013). The use of multiple correspondence analysis to explore associations between categories of qualitative variables in healthy ageing. *Journal of Aging Research*. https://doi.org/10.1155/2013/302163

de Sousa, R. D., Rodrigues, A. M., Gregório, M. J., Branco, J. D. C., Gouveia, M. J., Canhão, H., & Dias, S. S. (2017). Anxiety and depression in the portuguese older adults: Prevalence and associated factors. *Frontiers in Medicine*. https://doi.org/10.3389/fmed.2017.00196

Dimitrovová, K., Costa, C., Santana, P., & Perelman, J. (2017). Evolution and financial cost of socioeconomic inequalities in ambulatory care sensitive conditions: An ecological study for Portugal, 2000-2014. *International Journal for Equity in Health*. https://doi.org/10.1186/s12939-017-0642-7

Elliott, A. M., Smith, B. H., Penny, K. I., Smith, W. C., & Chambers, W. A. (1999). The epidemiology of chronic pain in the community. *Lancet*. https://doi.org/10.1016/S0140-6736(99)03057-3

Engels, T., Baglione, Q., Audibert, M., Viallefont, A., Mourji, F., Faris, M. E. A., Yahyaoui, M., Slassi, I., & Aidi, S. (2014). Socioeconomic status and stroke prevalence in morocco:results from the Rabat-Casablanca study. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0089271

Espelt, A., Borrell, C., Roskam, A. J., Rodríguez-Sanz, M., Stirbu, I., Dalmau-Bueno, A., Regidor, E., Bopp, M., Martikainen, P., Leinsalu, M., Artnik, B., Rychtarikova, J., Kalediene, R., Dzurova, D., Mackenbach, J., & Kunst, A. E. (2008). Socioeconomic inequalities in diabetes mellitus across Europe at the beginning of the 21st century. *Diabetologia*. https://doi.org/10.1007/s00125-008-1146-1

Everson, S. A., Maty, S. C., Lynch, J. W., & Kaplan, G. A. (2002). Epidemiologic evidence for the relation between socioeconomic status and depression, obesity, and diabetes. *Journal of Psychosomatic Research*. https://doi.org/10.1016/S0022-3999(02)00303-3

Ezzrari, A., & Verme, P. (2013). A multiple correspondence analysis approach to the measurement of multidimensional poverty in Morocco 2001–2007. In *Economic Studies in Inequality, Social Exclusion and Well-Being*. https://doi.org/10.1007/978-1-4614-5263-8_7

Falcão, I. M., Pinto, C., Santos, J., Fernandes, M. D. L., Ramalho, L., Paixão, E., & Falcão, J. M. (2008). Estudo da prevalência da diabetes e das suas complicações numa coorte de diabéticos portugueses: Um estudo na rede médicos-sentinela. *Revista Portuguesa de Clínica Geral*. https://doi.org/10.32385/rpmgf.v24i6.10566

Furtado, C., & Pereira, J. (2010). Equidade e Acesso aos Cuidados de Saúde. *Ensp-Unl*, 52. http://1nj5ms2lli5hdggbe3mm7ms5.wpengine.netdna-cdn.com/files/2010/08/EA1.pdf

Gardete-Correia, L., Boavida, J. M., Raposo, J. F., Mesquita, A. C., Fona, C., Carvalho, R., & Massano-Cardoso, S. (2010). First diabetes prevalence study in Portugal: PREVADIAB study. *Diabetic Medicine*. https://doi.org/10.1111/j.1464-5491.2010.03017.x

Gershon, A. S., Dolmage, T. E., Stephenson, A., & Jackson, B. (2012). Chronic obstructive pulmonary disease and socioeconomic status: A systematic review. In *COPD: Journal of Chronic Obstructive Pulmonary Disease*. https://doi.org/10.3109/15412555.2011.648030

Global Initiative for Chronic Obstructive Lung Disease, Inc.. (2019). Pocket guide to COPD diagnosis, management, and prevention: A guide for health care professionals.

Gouveia, M., & Augusto, M. (2011). Indirect costs of chronic pain. *Revista Portuguesa de Saude Publica*. https://doi.org/10.1016/S0870-9025(11)70013-X

Gouveia, N., Rodrigues, A., Eusébio, M., Ramiro, S., Machado, P., Canhão, H., & Branco, J. C. (2016). Prevalence and social burden of active chronic low back pain in the adult Portuguese population: results from a national survey. *Rheumatology International*. https://doi.org/10.1007/s00296-015-3398-7

Gravelle, H. (2003). Measuring income related inequality in health: Standardisation and the partial concentration index. *Health Economics*. https://doi.org/10.1002/hec.813

Großschädl, F., Stolz, E., Mayerl, H., Rásky, É., Freidl, W., & Stronegger, W. (2016). Educational inequality as a predictor of rising back pain prevalence in Austria - Sex differences. *European Journal of Public Health*. https://doi.org/10.1093/eurpub/ckv163

Hesterberg, T. (2011). Bootstrap. *Wiley Interdisciplinary Reviews: Computational Statistics*. https://doi.org/10.1002/wics.182

Hong, J., Knapp, M., & Mcguire, A. (2011). Income-related inequalities in the prevalence of depression and suicidal behaviour: A 10-year trend following economic crisis. *World Psychiatry*. https://doi.org/10.1002/j.2051-5545.2011.tb00012.x

Hoy, D., March, L., Brooks, P., Blyth, F., Woolf, A., Bain, C., Williams, G., Smith, E., Vos, T., Barendregt, J., Murray, C., Burstein, R., & Buchbinder, R. (2014). The global burden of low back pain: Estimates from the Global Burden of Disease 2010 study. *Annals of the Rheumatic Diseases*. https://doi.org/10.1136/annrheumdis-2013-204428

Instituto Nacional de Estatística. (2014). *Inquérito Nacional de Saúde.* Accessed on 10th September on: https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_publicacoes&PUBLICACOES pub_boui=263714091&PUBLICACOESmodo=2&xlang=pt

Instituto Nacional de Estatística & PORDATA. (2014). Gini Index. Accessed on 15th October 2020 on: https://www.pordata.pt/en/Portugal/Gini+index+(percentage)-2166

Instituto Nacional de Estatística & PORDATA. (2014). Per capita relative PPP. Accessed on 13th October 2020 on: https://www.pordata.pt/Municipios/Poder+de+compra+per+capita-118

Instituto Nacional de Estatística & PORDATA. (2014). Portuguese resident population by age group. Accessed on 17th October 2020 on: https://www.pordata.pt/Portugal/População+residente++estimativas+a+31+de+Dezembro+total+e+por+grupo+etário-7

Instituto Nacional de Estatística & PORDATA. (2014). Portuguese population by educational abilities. Accessed on 17th October 2020 on: https://www.pordata.pt/Portugal/Popula%c3%a7%c3%a3o+residente+com+15+e+mais+anos+por+n%c3%advel+de+escolaridade+completo+mais+elevado+(percentagem)-884

Instituto Nacional de Estatística & PORDATA. (2014). Portuguese resident population by gender. Accessed on 17th October 2020 on: https://www.pordata.pt/Portugal/Popula%C3%A7%C3%A3o+residente++m%C3%A9dia+anual+total+e+por+sexo-6

Kakwani, N.C. (1980), Income Inequality and Poverty: Methods of Estimation and Policy Applications., New York: Oxford University Press

Kakwani, N. C. (1977). Measurement of Tax Progressivity: An International Comparison. *The Economic Journal*. https://doi.org/10.2307/2231833

Kakwani, N., Wagstaff, A., & Van Doorslaer, E. (1997). Socioeconomic inequalities in health: Measurement, computation, and statistical inference. *Journal of Econometrics*. https://doi.org/10.1016/S0304-4076(96)01807-6

Kessler, R. C., McGonagle, K. A., Swartz, M., Blazer, D. G., & Nelson, C. B. (1993). Sex and depression in the National Comorbidity Survey I: Lifetime prevalence, chronicity and recurrence. *Journal of Affective Disorders*. https://doi.org/10.1016/0165-0327(93)90026-G

Khangar, N. V., & Kamalja, K. K. (2017). Multiple Correspondence Analysis and its applications. *Electronic Journal of Applied Statistical Analysis*. https://doi.org/10.1285/i20705948v10n2p432

Kohn, J. L. (2012). What is health? A multiple correspondence health index. *Eastern Economic Journal*. https://doi.org/10.1057/eej.2011.5

Lerman, R. I., & Yitzhaki, S. (1989). Improving the accuracy of estimates of Gini coefficients. *Journal of Econometrics*. https://doi.org/10.1016/0304-4076(89)90074-2

Lin, P. J., Shaya, F. T., & Scharf, S. M. (2010). Economic implications of comorbid conditions among Medicaid beneficiaries with COPD. *Respiratory Medicine*. https://doi.org/10.1016/j.rmed.2009.11.009

Lõpez-Campos, J. L., Tan, W., & Soriano, J. B. (2016). Global burden of COPD. In *Respirology*. https://doi.org/10.1111/resp.12660

Lorant, V., Deliège, D., Eaton, W., Robert, A., Philippot, P., & Ansseau, M. (2003). Socioeconomic inequalities in depression: A meta-analysis. *American Journal of Epidemiology*. https://doi.org/10.1093/aje/kwf182

Mackenbach, J. P., Stirbu, I., Roskam, A.-J. R., Schaap, M. M., Menvielle, G., Leinsalu, M., & Kunst, A. E. (2008). Socioeconomic Inequalities in Health in 22 European Countries. *New England Journal of Medicine*. https://doi.org/10.1056/nejmsa0707519

Mangalore, R., Knapp, M., & Jenkins, R. (2007). Income-related inequality in mental health in Britain: The concentration index approach. *Psychological Medicine*. https://doi.org/10.1017/S003329170600969X

Mannino, D. M., Higuchi, K., Yu, T. C., Zhou, H., Li, Y., Tian, H., & Suh, K. (2015). Economic burden of COPD in the presence of comorbidities. *Chest*. https://doi.org/10.1378/chest.14-2434

Mielck, A., Reisig, V., & Rathmann, W. (2005). Health inequalities among persons with type 2 diabetes: The example of intermittent claudication. *Gesundheitswesen*. https://doi.org/10.1055/s-2005-858230

Miravitlles, M., Naberan, K., Cantoni, J., & Azpeitia, A. (2011). Socioeconomic status and health-related quality of life of patients with chronic obstructive pulmonary disease. *Respiration*. https://doi.org/10.1159/000328766

Mutyambizi, C., Booysen, F., Stokes, A., Pavlova, M., & Groot, W. (2019). Lifestyle and socio-economic inequalities in diabetes prevalence in South Africa: A decomposition analysis. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0211208

Nenadic, O., & Greenacre, M. (2011). Computation of Multiple Correspondence Analysis, with Code in R. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.847698

Nicholas, M., Vlaeyen, J. W. S., Rief, W., Barke, A., Aziz, Q., Benoliel, R., Cohen, M., Evers, S., Giamberardino, M. A., Goebel, A., Korwisi, B., Perrot, S., Svensson, P., Wang, S. J., & Treede, R. D. (2019). The IASP classification of chronic pain for ICD-11: Chronic primary pain. In *Pain*. https://doi.org/10.1097/j.pain.0000000000001390

Nunes, A. M., & Ferreira, D. C. (2019). The health care reform in Portugal: Outcomes from both the New Public Management and the economic crisis. *International Journal of Health Planning and Management*. https://doi.org/10.1002/hpm.2613

O'Donnell, O., van Doorslaer, E., Wagstaff, A., & Lindelow, M. (2007). Analyzing Health Equity Using Household Survey Data. In *Analyzing Health Equity Using Household Survey Data*. https://doi.org/10.1596/978-0-8213-6933-3

Pereira J, Pedro J. (2004) Education, income and health inequalities in Portugal. *Applied Health Economics and Health Policy*

Prescott, E., Lange, P., & Vestbo, J. (1999). Socioeconomic status, lung function and admission to hospital for COPD: Results from the Copenhagen City Heart Study. *European Respiratory Journal*. https://doi.org/10.1034/j.1399-3003.1999.13e28.x

Robbins, J. M., Vaccarino, V., Zhang, H., & Kasl, S. V. (2005). Socioeconomic status and diagnosed diabetes incidence. *Diabetes Research and Clinical Practice*. https://doi.org/10.1016/j.diabres.2004.09.007

Rodrigues, A. P., Sousa-Uva, M., Fonseca, R., Marques, S., Pina, N., & Matias-Dias, C. (2017). Depression and unemployment incidence rate evolution in Portugal, 1995-2013: General Practitioner Sentinel Network data. *Revista de Saude Publica*. https://doi.org/10.11606/S1518-8787.2017051006675

Rustøen, T., Wahl, A. K., Hanestad, B. R., Lerdal, A., Paul, S., & Miaskowski, C. (2005). Age and the experience of chronic pain: Differences in health and quality of life among younger, middle-aged, and older adults. *Clinical Journal of Pain*. https://doi.org/10.1097/01.ajp.0000146217.31780.ef

Saastamoinen, P., Leino-Arjas, P., Laaksonen, M., & Lahelma, E. (2005). Socio-economic differences in the prevalence of acute, chronic and disabling chronic pain among ageing employees. *Pain*. https://doi.org/10.1016/j.pain.2004.12.033

Santos, J., Kislaya, I., Antunes, L., Santos, A. J., Rodrigues, A. P., Neto, M., & Dias, C. M. (2017). Diabetes: Socioeconomic inequalities in the Portuguese population in 2014 | Diabetes: Desigualdades socioeconómicas na população Portuguesa em 2014. *Acta Medica Portuguesa*. https://doi.org/10.20344/amp.8235

Simon, G. E., & Vonkorff, M. (1998). Suicide mortality among patients treated for depression in an insured population. *American Journal of Epidemiology*. https://doi.org/10.1093/oxfordjournals.aje.a009428

Sociedade Portuguesa de Diabetologia. (2016). Diabetes: factos e números 2015 − Relatório Anual do Observatório Nacional de Diabetes. In *Lisboa: Sociedade Portuguesa da Diabetologia*.

Sociedade Portuguesa de Pneumologia. (2020). Portuguese information about COPD. Accessed on 30th September 2020 on: https://www.sppneumologia.pt/doentes/dpoc

Sommer, I., Griebler, U., Mahlknecht, P., Thaler, K., Bouskill, K., Gartlehner, G., & Mendis, S. (2015). Socioeconomic inequalities in non-communicable diseases and their risk factors: An overview of systematic reviews. *BMC Public Health*. https://doi.org/10.1186/s12889-015-2227-y

Sortsø, C., Lauridsen, J., Emneus, M., Green, A., & Jensen, P. B. (2017). Socioeconomic inequality of diabetes patients' health care utilization in Denmark. *Health Economics Review*. https://doi.org/10.1186/s13561-017-0155-5

Teixeira, C. de O. (2018). *Equity and the prevalence of chronic pain in Portugal*.

Todd, A., McNamara, C. L., Balaj, M., Huijts, T., Akhter, N., Thomson, K., Kasim, A., Eikemo, T. A., & Bambra, C. (2019). The European epidemic: Pain prevalence and socioeconomic inequalities in pain across 19 European countries. *European Journal of Pain (United Kingdom)*. https://doi.org/10.1002/ejp.1409

van Doorslaer, E., Koolman, X., & Jones, A. M. (2004). Explaining income-related inequalities in doctor utilisation in Europe. *Health Economics*. https://doi.org/10.1002/hec.919

Van Doorslaer, E., Wagstaff, A., Bleichrodt, H., Calonge, S., Gerdtham, U. G., Gerfin, M., Geurts, J., Gross, L., Häkkinen, U., Leu, R. E., O'Donnell, O., Propper, C., Puffer, F., Rodríguez, M., Sundberg, G., & Winkelhake, O. (1997). Income-related inequalities in health: Some international comparisons. *Journal of Health Economics*. https://doi.org/10.1016/S0167-6296(96)00532-2

Varela, A. V. (2020). Surge of Inequality : How Different Neighborhoods React to Flooding, mimeo.

Viegi, G., Scognamiglio, A., Baldacci, S., Pistelli, F., & Carrozzi, L. (2001). Epidemiology of Chronic Obstructive Pulmonary Disease (COPD). In *Respiration*. https://doi.org/10.1159/000050456

Von Korff, M., Wagner, E. H., Dworkin, S. F., & Saunders, K. W. (1991). Chronic pain and use of ambulatory health care. *Psychosomatic Medicine*. https://doi.org/10.1097/00006842-199101000-00006

Vyas, S., & Kumaranayake, L. (2006). Constructing socio-economic status indices: How to use principal components analysis. *Health Policy and Planning*. https://doi.org/10.1093/heapol/czl029

Wagstaff, A., & Doorslaer, E. V. (2000). Measuring and Testing for Inequity in the Delivery of Health Care. *Journal of Human Resources*. https://doi.org/10.2307/146369

Wagstaff, A. (2000). Socioeconomic inequalities in child mortality: Comparisons across nine developing countries. In *Bulletin of the World Health Organization*. https://doi.org/10.1590/S0042-96862000000100003

Wagstaff, A., Van Doorslaer, E., & Paci, P. (1989). Equity in the finance and delivery of health care: Some tentative cross-country comparisons. *Oxford Review of Economic Policy*. https://doi.org/10.1093/oxrep/5.1.89

World Health Organization. (2020). Information about Chronic Conditions. Accessed on 13th September 2020 on: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases

World Health Organization. (2020). Information about COPD. Accessed on 30th September 2020 on: https://www.who.int/news-room/fact-sheets/detail/chronic-obstructive-pulmonary-disease-(copd)

World Health Organization. (2020). Information about Depression. Accessed on 30th September 2020 on: https://www.who.int/health-topics/depression#tab=tab_1

World Health Organization. (2020). Information about Diabetes. Accessed on 29th September 2020 on: https://www.who.int/news-room/fact-sheets/detail/diabetes

World Health Organization. (2010). Portugal Health System; Performance Assessment. *WHO*.

World Health Organization - Regional Office for Europe. (2010). WHO evaluation of the national health plan of Portugal (2004–2010)

World Health Organization - Regional Office for Europe. (2005). WHO European Ministerial Conference on Mental Health: Facing The Challenges, Building Solutions. *Mental Health and Working Life*.

9. Appendix

A. Additional Data Analysis

A.1. NHS Sample as a good representation of the Portuguese population

In this subsection of Appendix A, the objective is to illustrate that, as mentioned in *Section 3*, our sample is a good fit for the entire Portuguese population. Remember that the calibration method was applied so that the NHS sample would fit the population almost perfectly, in terms of gender, age, education and region, among others. This said, by easily comparing some statistics for both the sample and the population, the goal will be to demonstrate that.

In terms of gender, in 2014, the Portuguese population was composed of 47.5% female individuals, and 52.5% male (INE and PORDATA, 2020), while the NHS sample presented 46.82% women and 53.18% men (Table A1. **1**).

Table A1. 1 – *Population and Sample Gender Division*

	Population	Sample
Female	47.5%	46.82%
Male	52.5%	53.18%
Total	100%	100%

Source: Data for population gender division was obtained from *Instituto Nacional de Estatística* and PORDATA.

In what regards region division based on NUTS II (Table A1. 2), the highest share of individuals lives in the North Region (35.1%). According to INE and PORDATA (2020), in 2014, 34.9% of the entire Portuguese population also resides in the North. If we observe the remaining areas, we understand that they also display similar patterns to those of the entire Portuguese population.

Table A1. 2 - *Sample and Population Region Division*

	Population	Sample
North	34.9%	35.1%
Algarve	4.3%	4.2%
Center	21.9%	22.2%
Lisbon	27%	26.6%
Alentejo	7.1%	7.2%
Azores	2.4%	2.3%
Madeira	2.5%	2.5%
Total	100%	100%

Source: Data from population region division was obtained from *Instituto Nacional de Estatística* and PORDATA.

Focusing on age, in Table A1. 3, all age groups present the exact same frequency, both in the sample and population, data from INE and PORDATA (2020) shows.

Table A1. 3 - *Population and Sample Age Group Division*

	Population	Sample
15-19	6.22%	6.22%
20-24	6.23%	6.23%
25-29	6.38%	6.38%
30-34	7.53%	7.53%
35-39	8.86%	8.86%
40-44	8.97%	8.97%
45-49	8.52%	8.52%
50-54	8.51%	8.51%
55-59	7.82%	7.82%
60-64	7.28%	7.28%
65-69	6.60%	6.60%
70-74	5.46%	5.46%
75-79	4.93%	4.93%
80-84	3.76%	3.76%
85+	2.94%	2.94%
Total	100.00%	100.00%

Source: Data for population age division was obtained from *Instituto Nacional de Estatística* and PORDATA.

Concerning education, although some categories are slightly different for the sample and the population, the analysis is straightforward as the divisions are extremely similar (Table A1. 4).

Table A1. 4 - *Population and Sample Education Division*

	Population	Sample
Preschool	8.9%	9.14%
1st or 2nd cycle	35.00%	34.6%
3rd cycle	20.5%	19.5%
High-School	19.2%	18.6%
Post-Secondary	-	1.07%
Higher Education	16.5%	17.1%
Total	100.00%	100.00%

Source: Data for population education division was obtained from *Instituto Nacional de Estatística* and PORDATA.

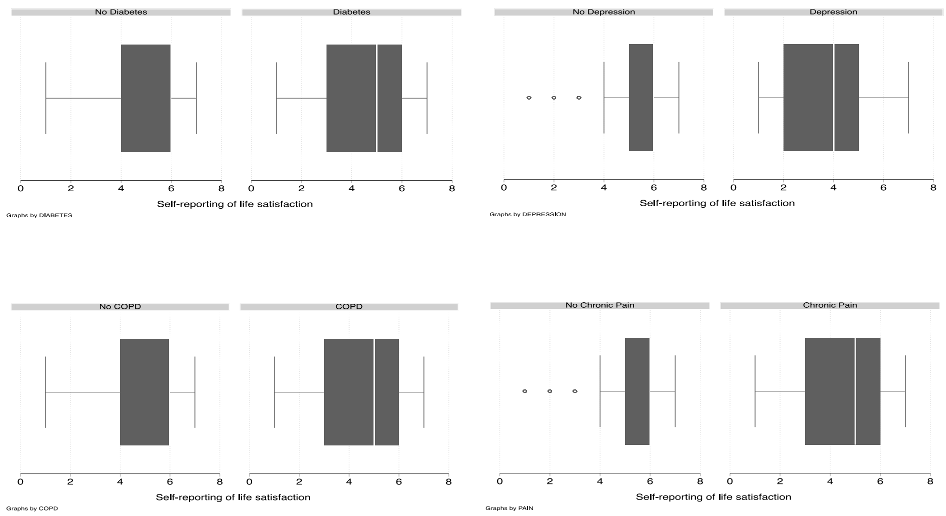
As such, all of these results corroborate the calibration used to ensure a more representative sample overall, and specifically, in terms of education, gender, region and age. Therefore, this sample is representative of the Portuguese population in 2014, and so, in principle, the results at which one arrives should be valid to use for inference and to conduct this analysis.

A.2. Additional Results, Facts and Characteristics about the NHS Sample and Chronic Patients

In this section, additional characteristics about our sample's chronic patients and their main differences when compared to non-patients will be presented. Also, complementary tables for some of the figures presented in *Section 3* will be presented, as is the case of Table A2. 1, concerning age groups and the Intensity of Unmet Needs division.

First off, chronic patients present lower life satisfaction levels (which are measured on a scale from 1 to 7, with the latter representing higher satisfaction), as it can be seen from the boxplots in Figure A2.1. In the presence of the diseases, dispersion in the results to lower values is higher. This will understandably decrease their life satisfaction means, and, thus, this is evidence that the quality of life of these patients is usually lower than that of the remaining population. Obviously, these same patients are also more likely to be found in the pools of individuals who consider their health status as being worse (Table A2. 1).

Figure A2. 1 - *Life Satisfaction Self-reporting by Diseases' Prevalence*



Additional facts about these patients can be mentioned:

1. Either by hospital stay or ambulatory healthcare, the percentage of chronic patients is higher when one analyzes the universe of respondents who recently had used these services (Table A2. 2).

Table A2. 1 - *Chronic Diseases Prevalence by Age Group, Intensity of Unmet Needs and Health Status Self-appreciation*

	Diabetes		Depression		COPD		Chronic Pain	
	No	Yes	No	Yes	No	Yes	No	Yes
Age Group								
15-19	99.68%	0.32%	98.75%	1.25%	97.37%	2.63%	88.43%	11.57%
20-24	99.90%	0.10%	96.68%	3.32%	97.60%	2.40%	87.01%	12.99%
25-29	99.44%	0.56%	93.95%	6.05%	98.52%	1.48%	88.22%	11.78%
30-34	99.23%	0.77%	93.55%	6.45%	97.36%	2.64%	78.32%	21.68%
35-39	98.96%	1.04%	93.80%	6.20%	96.98%	3.02%	79.42%	20.58%
40-44	97.41%	2.59%	91.05%	8.95%	96.39%	3.61%	76.33%	23.67%
45-49	95.36%	4.64%	89.21%	10.79%	95.86%	4.14%	73.04%	26.96%
50-54	91.00%	9.00%	85.10%	14.90%	95.85%	4.15%	71.07%	23.93%
55-59	86.34%	13.66%	81.09%	18.91%	94.87%	5.13%	66.25%	33.75%
60-64	84.55%	15.45%	82.93%	17.07%	90.52%	9.48%	57.63%	42.37%
65-69	78.28%	21.72%	78.12%	21.88%	91.81%	8.19%	57.46%	42.54%
70-74	73.26%	26.74%	80.79%	19.21%	88.10%	11.90%	50.20%	49.80%
75-79	74.76%	25.24%	81.37%	18.63%	86.34%	13.66%	49.76%	50.24%
80-84	78.75%	21.25%	81.45%	18.55%	87.48%	12.52%	51.37%	48.63%
85+	79.82%	20.18%	87.31%	12.69%	84.46%	15.54%	46.88%	53.23%
Intensity of Unmet Needs								
0	90.73%	9.27%	90.13%	9.87%	94.81%	5.19%	71.64%	28.36%
1	89.33%	10.67%	84.20%	15.80%	92.49%	7.51%	64.34%	35.66%
2	85.84%	14.16%	77.13%	22.87%	90.13%	9.87%	52.72%	47.28%
3	86.88%	13.12%	66.44%	33.56%	89.53%	10.47%	41.11%	58.89%
4	86.24%	13.76%	36.35%	63.65%	882.68%	17.32%	27.13%	72.87%
Health Status self-appreciation								
Very good	99.23%	0.77%	98.80%	1.20%	98.30%	1.70%	91.32%	8.68%
Good	99.64%	0.36%	96.35%	3.65%	97.89%	2.11%	84.92%	14.08%
Reasonable	86.04%	13.96%	83.20%	16.80%	93.02%	6.98%	61.94%	38.06%
Bad	73.78%	26.22%	68.75%	31.25%	84.60%	15.40%	32.02%	67.98%
Very bad	74.06%	25.94%	61.44%	38.56%	77.53%	22.47%	21.70%	78.30%

Table A2. 2 - *Chronic Disease Prevalence by Health Variables and Self-reporting of Life Satisfaction*

	Diabetes		Depression		COPD		Chronic Pain	
	No	Yes	No	Yes	No	Yes	No	Yes
Hospital Stay								
No	91.46%	8.54%	89.15%	10.85%	95.03%	4.97%	72.01%	27.99%
Yes	82.98%	17.02%	77.95%	22.05%	86.13%	13.87%	53.12%	46.88%
Visits to the hospital, excluding hospital stay								
No	92.56%	7.44%	91.36%	8.64%	95.84%	4.16%	77.86%	22.14%
Yes	87.88%	12.12%	83.29%	16.71%	91.80%	8.20%	59.05%	40.95%
Health Insurance								
No	89.28%	10.72%	87.06%	12.94%	93.59%	6.41%	68.47%	31.53%
Yes	96.15%	3.85%	92.20%	7.80%	96.69%	3.31%	77.51%	22.49%
Long Time Limitation								
No	99.64%	0.36%	98.37%	1.63%	99.70%	0.30%	88.12%	11.88%
Yes	83.75%	16.25%	80.20%	19.80%	89.97%	10.03%	56.49%	43.51%
Degree of Limitation								
Not limited	94.58%	5.42%	94.16%	5.84%	97.23%	2.77%	83.14%	16.86%
Limited but not severely	83.68%	16.32%	77.31%	22.69%	89.34%	10.66%	46.94%	53.06%
Severely limited	78.23%	21.77%	68.87%	31.13%	83.13%	16.87%	30.81%	69.19%
Missed work due to health problems								
No	96.12%	3.88%	95.52%	4.48%	97.75%	2.25%	82.28%	17.72%
Yes	94.78%	5.22%	84.48%	15.52%	95.39%	4.61%	62.30%	37.70%

2. Chronic patients are more likely found in the group of people who do not have health insurance. This uncovers the possibility of a lower than average insurance coverage in chronic patients (Table A2. 2).
3. These are also more likely to belong to the group of patients with a long-lasting health condition (16.35% versus 0.36% for the case of diabetes). These long-term diseases are, most likely, the chronic diseases we are analyzing (Table A2. 2).
4. Besides, they are much more prone to being partially or completely limited in general daily activities. The probability of finding someone with these diseases increases as we increase the degree of limitation (Table A2. 2).
5. They are also more likely to miss working activities, due to health issues (Table A2. 2).

Table A2. 3 - ***CP Prevalence by Pain Interference***

	Chronic Pain	
	No	Yes
Pain Interference		
None	92.56%	7.44%
Reduced	51.10%	48.90%
Moderately	24.00%	76.00%
High	4.24%	95.76%
Extremely	1.11%	98.89%

6. Individuals with chronic pain are obviously more likely to belong to categories that display higher levels of pain interference in their daily levels (Table A2. 3).
7. As previously mentioned in the literature review section, it is a fact that both diabetes and COPD can frequently carry other conditions, such as heart-related diseases, for instance. It is more probable to select someone who suffers from either diabetes or COPD, from the group of people who suffered a heart attack, stroke or coronary heart diseases, than otherwise (Table A2. 4).
8. As hinted in the literature review, one of the most known causes for COPD is smoking. Interestingly, the group where we can, more easily, find someone with COPD is the one of people who have once smoked. This number is higher than the one of the group of people who at the time of the inquiry smoked. We may hypothesize that, most likely, people who were diagnosed with COPD stopped smoking once diagnosed with the disease. As such, it still seems plausible that smoking is one of the disease's causes (Table A2. 5).

9. Depressed individuals are more often found in groups that more frequently display disinterest in performing regular activities, a depressed mood or feeling of worthlessness, which are symptoms consistent with their depression self-reporting (Table A2. 6).

Table A2. 4 - *Diabetes and COPD Prevalence by Heart Conditions*

	Diabetes			COPD		
	No	Yes	Total	No	Yes	Total
Heart Attack						
No	91.11%	8.89%	100%	94.45%	5.55%	100%
Yes	66.33%	33.67%	100%	81.18%	18.82%	100%
Stroke						
No	90.99%	9.01%	100%	94.43%	5.57%	100%
Yes	74.86%	25.14%	100%	83.28%	16.72%	100%
Coronary Heart Disease						
No	91.33%	8.67%	100%	94.89%	5.11%	100%
Yes	76.26%	23.74%	100%	79.53%	20.47%	100%

Table A2. 5 - *COPD Prevalence by Smoking Habits*

	COPD		
	No	Yes	Total
Smokes daily	94.80%	5.20%	100%
Smokes occasionally	97.33%	2.67%	100%
Has smoked, but not anymore	93.67%	6.33%	100%
Does not smoke	94.07%	5.93%	100%

10. Individuals suffering from any of these diseases are more likely to self-report depressed mood, fatigue or disinterest in doing usual things. Consequently, there is a clear indication that the probability of finding any of the studied diseases is higher when dealing with a depressed individual, with the clearest difference being the case of chronic pain. This high interaction between depression and the other conditions is possibly due to the high physical and emotional burden they carry. This corroborates the hypothesis introduced in the literature review, whereby some studies conclude that chronic pain and depression usually come hand-in-hand (Table A2. 7).

Table A2. 6 - *Depression Incidence by Depressive Symptoms*

	Disinterest doing usual things		Depressed mood		Guilt or feeling of uselessness	
Depression	No	Yes	No	Yes	No	Yes
Never	95.31%	4.69%	96.86%	3.14%	93.93%	6.07%
Several days	78.95%	21.05%	77.44%	22.56%	67.31%	32.69%
More than half of the days	60.96%	39.04%	57.10%	42.90%	55.03%	44.97%
Almost every day	50.85%	49.15%	41.08%	58.92%	35.93%	64.07%
Total	100%	100%	100%	100%	100%	100%

Table A2. 7 - *Diabetes, COPD and CP Incidence by Depressive Symptoms*

		Diabetes		COPD		Chronic Pain	
		No	Yes	No	Yes	No	Yes
Disinterest in doing usual things							
	Never	92.34%	7.66%	95.75%	4.25%	79.28%	20.72%
	Some days	88.87%	11.13%	91.85%	8.15%	55.50%	44.50%
	More than half of the days	81.75%	18.25%	88.67%	11.33%	40.11%	59.89%
	Almost every day	81.96%	17.04%	88.19%	11.81%	33.82%	66.18%
Depressed mood							
	Never	92.69%	7.31%	96.17%	3.83%	80.35%	19.65%
	Some days	87.64%	12.36%	91.63%	8.37%	53.38%	46.62%
	More than half of the days	81.42%	18.58%	86.01%	13.99%	37.73%	62.27%
	Almost every day	84.11%	15.89%			34.74%	65.26%
Fatigue							
	Never	92.84%	7.16%	96.66%	3.34%	84.91%	15.09%
	Some days	90.58%	9.42%	93.42%	6.58%	63.05%	36.95%
	More than half of the days	84.45%	15.55%	88.49%	11.51%	40.76%	59.24%
	Almost every day	82.13%	17.87%	86.44%	13.56%	28.87%	71.13%
Depression							
	No	91.55%	8.45%	95.11%	4.89%	74.45%	25.55%
	Yes	84.26%	15.74%	87.52%	12.48%	39.24%	60.76%

B. Variable Creation Procedures

To create the **Intensity of Unmet Needs** variable, 4 dataset variables were used. These are the following "In the past 12 months, due to financial problems, was there a need in any of the four presented areas that was not met:

I. Medical appointments or treatments;
II. Dental appointments, exams or treatments;
III. Prescription medicines;
IV. Psychiatric, psychology or psychotherapy appointment, or other mental health treatment."

To answer these questions, respondents had 3 hypotheses:

a) There was not a need to acquire that service/medicine;
b) There was a need and it was met;
c) There was a need and it was not satisfied due to financial problems.

To each respondent, one of 5 possible values was assigned. These could be 0, 1, 2, 3 or 4.[13]

0. If the individual faced at least one of the needs and was able to address it. However, the individual must not have any unmet need because of financial problems. [14]
1. If the individual was uncapable of meeting one of those needs, when necessary, independently of having satisfied the other three.
2. If the individual was uncapable of meeting two of those needs, when necessary, independently of having satisfied the other two.
3. If the individual was uncapable of meeting three of those needs, when necessary, independently of having satisfied the other one.
4. If the individual was uncapable of meeting four of those needs, when necessary.

To clarify the methodology used in the creation of this variable, imagine one individual that, to question I answered a), to question II answered c) and to questions III and IV answered b). This means that this individual will have a value of 1 in the Intensity of Unmet Needs variable.

[13] Besides, the value was recorded as missing if the respondent did not face any of the abovementioned needs.
[14] This is, all needs which an individual presents must be satisfied. If he just presents one need, he just has to be able to satisfy that one, and so on.

Although he was indeed able to address some of his needs, the fact that he was unable to address one (and only one), means he will belong to category 1.

In what regards **Unmet Needs**, this variable is only allowed to take one of 3 values - 0,1, or 2:

0. Faced at least one of the needs and was able to address it. However, the individual must not have any unmet need due to financial problems.[15]
1. Did not face any of the four needs.
2. Was uncapable of meeting at least one of those needs, when necessary, independently of having met the other three.

There are essentially two major differences between this variable and the previously created Intensity of Unmet Needs. The major reason for these is that, whereas the first variable was only used to better display some descriptive statistics of the population, the second was included in the MCA procedure. These differences will be explored in the next two paragraphs.

The first is the undeniable decrease in the number of categories. Implicitly, the Intensity of Unmet Needs assumes that all of the unmet needs counted equally. This means that, for instance, not being able to pay for a psychiatric appointment, was considered the same as not being able to acquire medication, on the socioeconomic scale. This may not be true. Note that, usually, psychiatric appointments are perceived more as a superior good, rather than a necessity, as is in the case of medication. For this reason, the newly created Unmet Needs variable only entails information of whether there was at least one unmet need and does not distinguish between respondents with a different number of unmet needs, since these can have different degrees of importance.

The second is related to the categorization of missing values. Remember that, following the work of Intensity of Unmet Needs, individuals who did not face any need should be recorded as missing values. Here, this will not be the case, because MCA formulation would entail some problems if there were missing values. Previous works advise the creation of a category to encompass all the observations that were recorded as missing. Here, since we do not know if the individual would belong to category 0 or 2 if he faced any need, the decision was to assume a mean value, acting as a benchmark, and, thus, the choice of the value 1.

[15] This is, all needs which an individual presents must be satisfied. If he just presents one need, he just has to be able to satisfy that one, and so on.

C. Multiple Correspondence Analysis Procedure

In this appendix, the objective will be to provide further details about the procedure of the index creation, through MCA. Below, in Table C1. 1, is the detailed information about the 5 categorical variables (i.e., our intermediate variables), that will be used throughout. Each one has a different number of categories, where the respondent can be allocated. Remember that each respondent must be allocated to one and only one category in each and every intermediate variable.

Table C1. 1 - *Categorical Variables Used in MCA*

MCA Categorical variables
Income (2 indicators)
A. HHI Quintile
1. 1st Quintile
2. 2nd Quintile
3. 3rd Quintile
4. 4th Quintile
5. 5th Quintile
B. Unmet Needs
1. No unmet needs
2. No needs
3. At least one unmet need
Region (2 indicators)
C. Region_Lisbon
1. Non-Lisbon
2. Lisbon
D. Degree of Urbanization
1. High density
2. Medium density
3. Low density
Education (1 indicator)
E. Years of Schooling
1. Preschool
2. 1st or 2nd cycle
3. 3rd cycle
4. High-School
5. Post-Secondary
6. Higher Education

The choice to include all these variables in the index creation, through MCA, must be a trade-off between the presence of more information and the decreasing predicting power. This is because the higher the number of variables, the lower the predicting power of the first component – the created index. As the number of variables increases, it becomes increasingly difficult to gather all the information in only one component. In this case, as the option was to use all the intermediate variables presented, the SES summary variable has a predictive power of 64.65%. The choice was based on the fact that all the variables mentioned above represented different and necessary features of an individual socioeconomic position. This predicted power corresponds to the percentage of variability in the data explained by this component (Table C1. 4).

Notice that, as mentioned in *Section 4*, the weights of each category of the intermediate variables represent the contribution that belonging to each group entails, and thus, should be positive when these belong to categories which are usually associated with higher SES, and vice-versa.

From the interpretation of the output, in Table C1. 2, this is clearly the case. Note that the weights from the different categories in each one of the variables appear either in ascending or descending order, as seen in the weight column. If the category presented first is usually associated with higher SES, the order is descending and, if otherwise, it is obviously because the first category of the intermediate variable is more prone to encompass people from lower SES. The interpretation is, undoubtedly, intuitive. Note that categories are ordered and thus, for instance, belonging to the 4^{th} quintile of the household equivalent income distribution must not attain a higher contribution to the socioeconomic scale, than being from the 5^{th}, for instance.

Table C1. 2 - *MCA Complete Stata Output*

	Overall			Dimension 1		
	Mass	Quality	%inert	Weight	Sq. Corr.	Contribution
Degree of Urbanization						
High density	0.06	0.727	0.121	**1.356**	0.589	0.111
Medium density	0.065	0.536	0.015	-0.126	0.043	0.001
Low density	0.074	0.821	0.058	-0.985	0.8	0.072
Region_Lisbon						
Non-Lisbon	0.178	0.716	0.018	-0.292	0.56	0.015
Lisbon	0.022	0.716	0.144	**2.4**	0.56	0.125
HHI Quintile						
1st	0.044	0.719	0.063	-1.211	0.667	0.065
2nd	0.041	0.749	0.03	-0.921	0.747	0.035
3rd	0.04	0.37	0.009	-0.354	0.353	0.005
4th	0.038	0.45	0.015	**0.503**	0.415	0.01
5th	0.037	0.726	0.184	**2.326**	0.707	0.201
Education						
Preschool	0.026	0.653	0.049	-1.355	0.627	0.047
1st or 2nd cycle	0.079	0.77	0.041	-0.774	0.749	0.048
3rd cycle	0.034	0.03	0.003	**0.063**	0.028	0
High-School	0.03	0.593	0.015	**0.673**	0.575	0.014
Post-Secondary	0.002	0.445	0.002	**0.877**	0.417	0.001
Higher Education	0.029	0.724	0.167	**2.496**	0.702	0.181
Unmet Needs						
No unmet needs	0.123	0.843	0.017	**0.371**	0.657	0.017
No needs	0.024	0.283	0.004	**0.270**	0.282	0.002
At least one unmet need	0.054	0.905	0.044	-0.967	0.746	0.05

Table C1. 3 - ***Principal Inertias of Unadjusted Burt Matrix***

Unadjusted Burt Method			
Dimensions	Principal Inertia	Explained percentage	Cumulative percentage
1	**.1371093**	22.26	22.26
2	.0665026	10.79	33.05
3	.0541352	8.79	41.84
4	.0446015	7.24	49.08
5	.0430243	6.98	56.06
6	.0416733	6.76	62.83
7	.0399697	6.49	69.31
8	.0388178	6.30	75.61
9	.0380842	6.18	81.80
10	.0365728	5.94	87.73
11	.0291639	4.73	92.47
12	.0259959	4.22	96.69
13	.0121923	1.98	98.67
14	.0082172	1.33	100.00
Total	.6160598	100.00	

Table C1. 4 - ***Principal Inertias of Adjusted Burt Matrix***

Adjusted Burt Method			
Dimensions	Principal Inertia	Explained percentage	Cumulative percentage
1	.0453066	64.65	64.65
2	.0052347	7.47	72.12
3	.0016677	2.38	74.50
4	.0001957	0.28	74.78
5	.0000861	0.12	74.91
6	.0000268	0.04	74.94
Total	.0700748	100.00	

D. Indirect Standardization Procedure

This process implementation follows closely the work of Gravelle (2003), where the author presents the steps to follow in an indirect standardization procedure. Note that, in the same work, some criticism to this process is also presented, due to the possible inconsistency of the estimators, which will be explained in the next paragraph. Nonetheless, this process will be used due to the general consensus in its advantages when compared to the direct case, as previously stated.

The **first step** is to choose the variables for which we want to standardize. These, as stated, will be age and gender. The goal, then, is to design a regression, where we regress our health variable on the chosen variables, to get an estimate of the effects that the standardizing variables have on health outcomes. Gravelle (2003) recalls that by doing this and if SES is also correlated with age and gender, we may be removing some of its effect by removing the effect of age and gender and that, thus, estimators may be inconsistent. This is an additional reason that justifies the fact that this methodology was only presented as an addition to the rest of the work.

One way of trying to overcome this problem could be to include some of the variables of the SES index in the regression, nonetheless, Gravelle (2003) also explains that this does not completely eliminate the problem. As such, the simple procedure was adopted to easily compute the avoidable share of inequality.

A logistic regression was designed, using robust standard errors, to account for the possibility of heteroskedasticity, since our dependent variables are binary.

$$\begin{aligned} h_i = \alpha + \varphi female_i + \beta_1 age_{1_i} + \beta_2 age_{2_i} + \beta_3 age_{3_i} + \beta_4 age_{4_i} + \beta_5 age_{5_i} + \beta_6 age_{6_i} \\ + \beta_7 age_{7_i} + \beta_9 age_{9_i} + \beta_{10} age_{10_i} + \beta_{11} age_{11_i} + \beta_{12} age_{12_i} + \beta_{13} age_{13_i} \\ + \beta_{14} age_{14_i} + \beta_{15} age_{15_i} + \varepsilon_i \end{aligned} \quad \textbf{(9)}$$

Where age_j is respectively one dummy variable for each one of the age groups[16], from *j=1* to *j=7 and j=9* to *j=15* [17], that takes the value of 1 when the individual belongs to the group, and 0 otherwise. $female$ is a dummy variable, that takes the value of 1 if the individual is a female,

[16] These groups can be seen in Table A1. 3.

[17] Note that there are 15 age groups, but the procedure, when there exist n groups, is to create n-1 dummies, to avoid multicollinearity. Here, the choice was to set as a baseline the 8th group.

and 0 otherwise.[18] After regressing, the corresponding value for each individual was predicted and corresponded to $\hat{h}_i$.

Note that we will have 4 logistic regressions, one for each condition and, thus, 4 different groups of estimators, one for each case. Both the second and the third steps will also have to be repeated for each one of the diseases.

Then, at a **second stage**, the objective is to compute the indirectly standardized health outcome. As such, the initial "real" health outcome that could take values of 0 or 1, will be substituted by an indirectly computed one. This means that the indirectly standardized health outcome is the expected health outcome, if one were to use as explanatory variables only gender and age. So, the goal is to subtract this value ($\hat{h}_i$) to the "real" self-reported health outcome (h_i^R). After that, the sample mean of the "real" health outcome is added ($\overline{h^R}$) to uniformize the newly constructed variable.

$$h_i^N = h_i^R - \hat{h}_i + \overline{h^R} \qquad \textbf{(10)}$$

However, some transformations are needed, before moving forward to the indirectly standardized concentration index computation. Note that now it is likely that some of the individual's standardized health outcomes are negative, while others remain positive. As already mentioned, throughout this research, when computing concentration indices, one should only use positive numbers, since the index is sensitive to the absolute values of the observation, and not just to their relative difference.

As such, to maintain the initial interval that "real" health could take, the **third step** will be to apply two transformations to the previous h_i^N. The first consisted of summing the minimum value the new standardized health outcome could take to every observation, in order to avoid having negative observations.[19]

$$h_i^{t1} = h_i^N + |\min(h^N)| \qquad \textbf{(11)}$$

[18] Notice that the choice to use this model with these variables was based on a complete procedure, where other hypotheses (e.g., using interactive terms with both age and gender or having a variable concerning age and age squared) were considered and this was the model that displayed the best representation.
[19] Remember that this procedure had already been applied when constructing the final SES measure, with the objective of dragging the entire distribution to the positive spectrum, so it could be used in the computation of concentration indices.

Secondly, this already transformed variable was again altered so as to have as an upper bound the value of 1, as in the original health outcome variable. Therefore, the maximum value of h_i^{t1} was used as a divisor in all the observations.

$$h_i^{t2} = \frac{h_i^{t1}}{\max\,(h^{t1})} \qquad \textbf{(12)}$$

Thus, the final transformed variable has a lower bound of 0 and an upper bound of 1, exactly like the "real" health outcome.

Then, the final objective is to compute the indirectly standardized concentration index, which is solely the concentration index of the indirectly standardized health. Therefore, it is just a matter of applying the concentration index equation, but using the indirectly standardized health instead of the "real" health outcome. These values are presented in the next section.

Note that all logistic regression results are presented in the following tables (Table D1. 1 and Table D1. 2).

Table D1. 1 - ***Auxiliary Logistic Regression for Diabetes and Depression***

	Diabetes						Depression					
	Coefficient	Robust Std. Error	z	P>z	[95% Conf. Interval]		Coefficient	Robust Std. Error	z	P>z	[95% Conf. Interval]	
Female	-.1654639	.0737239	-2.24	0.025	-.30996	-.0209678	1.185794	.0775566	15.29	0.000	1.033786	1.337803
Age_group												
15-19	-3.438	.8289552	-4.15	0.000	-5.062723	-1.813278	-2.629683	.4488022	-5.86	0.000	-3.509319	-1.750046
20-24	-4.612039	.6344425	-7.27	0.000	-5.855523	-3.368554	-1.634073	.3049959	-5.36	0.000	-2.231854	-1.036292
25-29	-2.874994	.7392641	-3.89	0.000	-4.323925	-1.426063	-1.001386	.2782474	-3.60	0.000	-1.546741	-.4560312
30-34	-2.547803	.3738392	-6.82	0.000	-3.280515	-1.815092	-.9482459	.2130722	-4.45	0.000	-1.36586	-.5306321
35-39	-2.240773	.3724555	-6.02	0.000	-2.970773	-1.510774	-.9955038	.1788889	-5.56	0.000	-1.34612	-.644888
40-44	-1.314367	.2790464	-4.71	0.000	-1.861288	-.767446	-.5924091	.1660784	-3.57	0.000	-.9179168	-.2669014
45-49	-.7102471	.2347554	-3.03	0.002	-1.170359	-.250135	-.3799253	.1636445	-2.32	0.020	-.7006627	-.0591879
55-59	.4708201	.1768446	2.66	0.008	.1242111	.8174291	.2943942	.1479514	1.99	0.047	.0044148	.5843736
60-64	.6157557	.1715917	3.59	0.000	.2794421	.9520693	.1588426	.1446978	1.10	0.272	-.1247598	.442445
65-69	1.035692	.1654786	6.26	0.000	.7113603	1.360024	.4667007	.142528 3	.27	0.001	.187351	.7460505
70-74	1.31275	.1701373	7.72	0.000	.9792873	1.646213	.2773319	.1513293	1.83	0.067	-.0192681	.5739319
75-79	1.23808	.1686788	7.34	0.000	.9074757	1.568685	.2130512	.1512957	1.41	0.159	-.083483	.5095853
80-84	1.019358	.1911603	5.33	0.000	.6446902	1.394025	.1700672	.169421 1	.00	0.315	-.1619917	.5021262
85+	.9651692	.2055084	4.70	0.000	.5623802	1.367958	-.3602157	.2248982	-1.60	0.109	-.801008	.0805766
_cons	-2.229367	.1436717	-15.52	0.000	-2.510958	-1.947775	-2.482401	.1252534	-19.82	0.000	-2.727893	-2.236909

Note: Baseline age group is group 8th, with ages 50-54.

Table D1. 2 - ***Auxiliary Logistic Regression for COPD and Chronic Pain***

	COPD						Chronic Pain					
	Coefficient.	Robust Std. Error	z	P>z	[95% Conf. Interval]		Coefficient.	Robust Std. Error	z	P>z	[95% Conf. Interval]	
Female	.2897221	.0912663	3.17	0.002	.1108434	.4686008	.8298185	.052081	15.93	0.000	.7277415	.9318954
Age_cod												
15-19	-.4634592	.3632042	-1.28	0.202	-1.175326	.248408	-1.135923	.1853562	-6.13	0.000	-1.499214	-.772631
20-24	-.5593187	.3560756	-1.57	0.116	-1.257214	.1385766	-1.007793	.1704658	-5.91	0.000	-1.3419	.673686
25-29	-1.055538	.3929835	-2.69	0.007	-1.825771	-.2853041	-1.124522	.1951118	-5.76	0.000	-1.506934	-.7421096
30-34	-.464215	.3098083	-1.50	0.134	-1.071428	.1429981	-.389955	.1352013	-2.88	0.004	-.6549447	-.1249654
35-39	-.3291218	.2803085	-1.17	0.240	-.8785163	.2202727	-.4616704	.1238851	-3.73	0.000	-.7044807	-.2188602
40-44	-.1454158	.2653152	-0.55	0.584	-.6654239	.3745924	-.2785423	.1204583	-2.31	0.021	-.5146361	-.0424484
45-49	-.0016527	.2696725	-0.01	0.995	-.5302011	.5268958	-.0988299	.1222695	-0.81	0.419	-.3384738	.1408139
55-59	.2216705	.2408664	0.92	0.357	-.2504189	.6937599	.2290442	.1182482	1.94	0.053	-.002718	.4608064
60-64	.8816894	.2197967	4.01	0.000	.4508958	1.312483	.6072648	.1130133	5.37	0.000	.3857628	.8287669
65-69	.7179416	.2298533	3.12	0.002	.2674375	1.168446	.6036418	.1140678	5.29	0.000	.380073	.8272106
70-74	1.128267	.2184967	5.16	0.000	.7000212	1.556513	.8940987	.1171317	7.63	0.000	.6645248	1.123673
75-79	1.280491	.2156818	5.94	0.000	.8577625	1.70322	.8943684	.1179018	7.59	0.000	.6632851	1.125452
80-84	1.171103	.2262236	5.18	0.000	.7277132	1.614493	.7982003	.1262381	6.32	0.000	.5507782	1.045622
85+	1.403162	.2567064	5.47	0.000	.9000265	1.906297	.928532	.1558945	5.96	0.000	.6229844	1.23408
_cons	-3.300764	.1918026	-17.21	0.000	-3.67669	-2.92484	-1.369449	.0896403	-15.28	0.000	-1.545141	-1.193757

Note: Baseline age group is group 8th, with ages 50-54.

E. Additional Tables of Results[20]

E.1. Unstandardized Results

Table E1. 1 – ***Concentration Index and Concentration Curve Deciles for the SES variable***

		SES					
		Observed Coef.	Bootstrap Std. Err.	z	P>z	Normal-based [95% Conf. Interval]	
Concentration Index		.3258083	.002054	158.62	0.000	.3217825	.3298342
Deciles	1	.0203849	.0004222	48.28	0.000	.0195574	.0212124
	2	.0615125	.000752	81.80	0.000	.0600387	.0629864
	3	.1151078	.0010919	105.42	0.000	.1129677	.117248
	4	.1819016	.0013536	134.39	0.000	.1792487	.1845545
	5	.2632855	.0015657	168.16	0.000	.2602168	.2663542
	6	.3619958	.0017964	201.52	0.000	.3584749	.3655166
	7	.48059	.0018929	253.89	0.000	.4768799	.4843
	8	.6202542	.0017294	358.66	0.000	.6168647	.6236437
	9	.7861716	.0011668	673.79	0.000	.7838847	.7884584
	10	1	.	.	.	.	.

.

Table E1. 2 – ***Concentration Indices and Concentration Curve Deciles for Diabetes***

		Diabetes					
		Observed Coef.	Bootstrap Std. Err.	z	P>z	Normal-based [95% Conf. Interval]	
Concentration Index		-.2350278	.017634	-13.33	0.000	-.2695897	-.2004659
Deciles	1	.1871703	.0110981	16.87	0.000	.1654185	.2089222
	2	.3246876	.0137006	23.70	0.000	.2978348	.3515403
	3	.4511807	.0150821	29.92	0.000	.4216204	.480741
	4	.5547655	.0157146	35.30	0.000	.5239654	.5855657
	5	.6673498	.0157442	42.39	0.000	.6364918	.6982078
	6	.7534064	.0152052	49.55	0.000	.7236047	.7832081
	7	.8429813	.0126233	66.78	0.000	.81824	.8677226
	8	.9133513	.0109417	83.47	0.000	.891906	.9347966
	9	.9655601	.0065369	147.71	0.000	.9527481	.9783722
	10	1	.	.	.	.	.

[20] All standard errors on this Appendix Section were estimated by bootstrapping.

Table E1. 3 – ***Concentration Indices and Concentration Curve Deciles for Depression***

		Depression					
		Observed Coef.	Bootstrap Std. Err.	z	P>z	Normal-based [95% Conf. Interval]	
Concentration Index		-.1491089	.0167972	-8.88	0.000	-.1820308	-.116187
Deciles	1	.1603365	.0090328	17.75	0.000	.1426326	.1780405
	2	.2787655	.0112281	24.83	0.000	.2567587	.3007722
	3	.3995794	.0132796	30.09	0.000	.3735519	.425607
	4	.5037183	.014708	34.25	0.000	.474891	.5325455
	5	.5944156	.0148444	40.04	0.000	.5653211	.6235102
	6	.6871211	.0148059	46.41	0.000	.6581021	.7161401
	7	.7924419	.0133908	59.18	0.000	.7661965	.8186873
	8	.8803537	.0112115	78.52	0.000	.8583795	.9023278
	9	.9456624	.0079061	119.61	0.000	.9301666	.9611581
	10	1	.	.	.	.	.

Table E1. 4 – ***Concentration Indices and Concentration Curve Deciles for COPD***

		COPD					
		Observed Coef.	Bootstrap Std. Err.	z	P>z	Normal-based [95% Conf. Interval]	
Concentration Index		-.2299677	.0244931	-9.39	0.000	-.2779733	-.1819621
Deciles	1	.2036733	.0142659	14.28	0.000	.1757127	.231634
	2	.3451053	.0180789	19.09	0.000	.3096712	.3805394
	3	.4798639	.0203044	23.63	0.000	.4400681	.5196597
	4	.5705507	.0205884	27.71	0.000	.5301983	.6109031
	5	.6547002	.0208692	31.37	0.000	.6137973	.6956031
	6	.7288603	.0203295	35.85	0.000	.6890152	.7687054
	7	.8024219	.018896	42.47	0.000	.7653864	.8394573
	8	.9014426	.0157423	57.26	0.000	.8705884	.9322969
	9	.9541757	.0105578	90.38	0.000	.9334829	.9748685
	10	1	.	.	.	.	.

Table E1. 5 – *Concentration Indices and Concentration Curve Deciles for Chronic Pain*

		Chronic Pain					
		Observed Coef.	Bootstrap Std. Err.	z	P>z	Normal-based [95% Conf. Interval]	
Concentration Index		-.1339939	.0098587	-13.59	0.000	-.1533166	-.1146711
Deciles	1	.1618131	.0048117	33.63	0.000	.1523824	.1712438
	2	.2863588	.0065579	43.67	0.000	.2735056	.299212
	3	.3973275	.0079252	50.13	0.000	.3817943	.4128607
	4	.4977572	.0084879	58.64	0.000	.4811212	.5143932
	5	.5856574	.00839	69.80	0.000	.5692133	.6021015
	6	.6740368	.0086417	78.00	0.000	.6570995	.6909741
	7	.768678	.0078455	97.98	0.000	.7533012	.7840548
	8	.8588971	.0068495	125.40	0.000	.8454723	.872322
	9	.9316588	.0052886	176.16	0.000	.9212934	.9420242
	10	1	.	.	.	.	.

E.2. Standardized Results

Table E2. 1 - *Standardized Results for Diabetes*

		Diabetes					
		Observed Coef.	Bootstrap Std. Err.	z	P>z	Normal-based [95% Conf. Interval]	
Concentration Index		-.0377736	.0053214	-7.10	0.000	-.0482033	-.0273439
Deciles	1	.110402	.0033729	32.73	0.000	.1037912	.1170128
	2	.2107701	.0042277	49.85	0.000	.2024839	.2190562
	3	.3193981	.0047092	67.82	0.000	.3101682	.328628
	4	.4235988	.0047754	88.70	0.000	.4142392	.4329584
	5	.5307434	.0047495	111.75	0.000	.5214346	.5400523
	6	.6289522	.0044528	141.25	0.000	.6202248	.6376796
	7	.728356	.0038927	187.11	0.000	.7207264	.7359856
	8	.8208852	.003189	257.41	0.000	.8146348	.8271356
	9	.9148853	.0022435	407.80	0.000	.9104882	.9192824
	10	1	.	.	.	.	.

Table E2. 2 - *Standardized Results for Depression*

		Depression					
		Observed Coef.	Bootstrap Std. Err.	z	P>z	Normal-based [95% Conf. Interval]	
Concentration Index		-.0336145	.006049	-5.56	0.000	-.0454703	-.0217588
Deciles	1	.1100687	.0034623	31.79	0.000	.1032828	.1168546
	2	.2109929	.0042574	49.56	0.000	.2026484	.2193373
	3	.3177739	.0049942	63.63	0.000	.3079854	.3275623
	4	.421939	.0053977	78.17	0.000	.4113598	.4325183
	5	.5202708	.0053264	97.68	0.000	.5098312	.5307104
	6	.6219245	.0052866	117.64	0.000	.611563	.6322861
	7	.7265086	.0048749	149.03	0.000	.7169539	.7360632
	8	.8242777	.0040819	201.93	0.000	.8162773	.8322781
	9	.9158344	.0031582	289.98	0.000	.9096444	.9220244
	10	1	.	.	.	.	.

Table E2. 3 - *Standardized Results for COPD*

		COPD					
		Observed Coef.	Bootstrap Std. Err.	z	P>z	Normal-based [95% Conf. Interval]	
Concentration Index		-.0492651	.0079267	-6.22	0.000	-.0648011	-.0337292
Deciles	1	.1190358	.0046807	25.43	0.000	.1098618	.1282099
	2	.2265833	.0057145	39.65	0.000	.2153832	.2377835
	3	.3385138	.0063939	52.94	0.000	.3259819	.3510456
	4	.4389989	.0067395	65.14	0.000	.4257896	.4522081
	5	.5352784	.0069543	76.97	0.000	.5216483	.5489086
	6	.6284638	.006754	93.05	0.000	.6152262	.6417014
	7	.7207816	.0062131	116.01	0.000	.7086042	.732959
	8	.8230501	.0052308	157.35	0.000	.812798	.8333023
	9	.9140447	.0036449	250.77	0.000	.9069008	.9211887
	10	1	.	.	.	.	.

Table E2. 4 - *Standardized Results for Chronic Pain*

		Chronic Pain					
		Observed Coef.	Bootstrap Std. Err.	z	P>z	Normal-based [95% Conf. Interval]	
Concentration Index		-.0388855	.0044311	-8.78	0.000	-.0475703	-.0302007
Deciles	1	.1153285	.0020798	55.45	0.000	.1112522	.1194048
	2	.2214026	.0028304	78.22	0.000	.215855	.2269501
	3	.324993	.0033626	96.65	0.000	.3184025	.3315836
	4	.4289541	.0037315	114.95	0.000	.4216404	.4362678
	5	.5248068	.003953	132.76	0.000	.517059	.5325546
	6	.6226996	.0038759	160.66	0.000	.6151029	.6302963
	7	.7232789	.0037798	191.36	0.000	.7158707	.7306872
	8	.8200885	.0031939	256.77	0.000	.8138287	.8263483
	9	.9127356	.0025706	355.07	0.000	.9076973	.9177738
	10	1	.	.	.	.	.