CHARLES UNIVERSITY
FACULTY OF SOCIAL SCIENCES
Institute of Economic Studies

The effect of the smoking ban in bars and restaurants on the health of the non-smoking population in the Czech Republic

Bachelor's thesis

Author: Adéla Bláhová
Study program: Economics and Finance
Supervisor: PhDr. Jana Votápková, Ph.D.
Year of defense: 2024

Declaration of Authorship

The author hereby declares that he or she compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain any other academic title.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis in whole or in part and agrees with the thesis being used for study and scientific purposes.

During the preparation of this thesis, the author used OpenAI´s ChatGPT to assist in developing the R-code, and refining the writing style. After using this tool, the author reviewed and edited the content as necessary and takes full responsibility for the content of the publication.

Prague, April 29, 2024

Adela Blahova

Abstract

In 2017, the Czech Republic became one of many European countries to adopt a comprehensive smoking ban prohibiting smoking in various public places including bars and restaurants. This thesis estimates the effect of this policy on the health of Czech non-smokers. To evaluate the impact of the smoking ban, we apply the difference-in-differences method to compare our treatment group, i.e. Czech non-smoking population, with the control group represented by non-smokers from the Slovak Republic, where such a policy was not implemented. The study used micro-level data from the European Health Interview Survey in 3 subsequent waves: 2009, 2014, and 2019. Two econometric models were employed on two distinct dependent variables, that were chosen as proxies of health. Firstly, the presence of longstanding health problem is modeled utilizing logistic regression. Subsequently, the number of nights spent in the hospital in the past 12 months is estimated by a Zero-inflated negative binomial model. We found a significant decrease in both the probability of experiencing a longstanding health problem and the expected number of nights spent in the hospital, suggesting that the smoking ban positively affected the health of the Czech non-smoking population.

Keywords	smoking ban, health, passive smoking, EHIS, Czech Republic
Title	The effect of the smoking ban in bars and restaurants on the health of the non-smoking population in the Czech Republic
Author's e-mail	84249192@fsv.cuni.cz
Supervisor's e-mail	jana.votapkova@fsv.cuni.cz

Abstrakt

V roce 2017 se Česká republika stala jednou z mnoha evropských zemí, které zavedly zákon zakazující kouření na mnoha veřejných místech včetně barů a restaurací. Tato práce zkoumá efekt zákazu kouření na zdraví českých nekuřáků. K odhadu tohoto efektu požíváme metodu difference-in-differences, která porovnává treatment skupinu, tedy české nekuřáky, s control skupinou, která se skládá z nekuřáků ze Slovenské Republiky, kde takové nařízení nebylo zavedeno. Studie používá data z šetření domácností EHIS ze tří po sobě jdoucích vln: 2009, 2014, a 2019. Dvě různé závislé proměnné byly vybrány jako proxy zdraví: přítomnost dlouhodobého zdravotního problému a počet nocí strávených v nemocnici za posledních 12 měsíců. Byly použity dva různé ekonometrické modely, Logit model a Zero-infalted negative binomial model. Výsledky prokázaly, že zavedení zákazu kouření v barech a v restauracích vedlo ke snížení jak pravděpodobnosti výskytu dlouhodobého zdravotního problému, tak i očekávaného počtu nocí strávených v nemocnici. Tyto výsledky naznačují, že zákaz kouření měl pozitivní vliv na zdraví českých nekuřáků.

Klíčová slova	zákaz kouření, zdraví, pasivní kouření, Česká republika
Název práce	Vliv zákazu kouření v restauracích na zdraví nekuřáků v ČR

E-mail autora	84249192@fsv.cuni.cz
E-mail vedoucího práce	jana.votapkova@fsv.cuni.cz

Acknowledgments

I would like to express my gratitude to PhDr. Jana Votápková, Ph.D., for her guidance, feedback, and support throughout the time of writing this thesis. Her insights were invaluable in shaping my bachelor's thesis into the final version. Additionally, I am grateful to my colleagues and classmates at the university, without whom I could not have imagined graduating at all.

Typeset in FSV LaTeX template with great thanks to prof. Zuzana Havrankova and prof. Tomas Havranek of Institute of Economic Studies, Faculty of Social Sciences, Charles University.

Bibliographic Record
Blahova, Adela: *The effect of the smoking ban in bars and restaurants on the health of the non-smoking population in the Czech Republic*. Bachelor's thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies, Prague. 2024, pages 54. Advisor: PhDr. Jana Votápková, Ph.D.

Contents

List of Tables

List of Figures

Acronyms

SHS Second-hand smoke

DiD Difference-in-Differences

ZINB Zero-inflated Negative Binomial

NB Negative Binomial

EHIS European Helath Interview Survey

LHP Longstanding Health Problem

MLE Maximum Likelihood Estimator

LPM Linear Probability Model

LR Likelihood Ratio

AIC Akeike Information Criterion

BIC Bayesian Information Criterion

VIF Variation Inflation Factor

CZ Czech Republic

SK Slovak Republic

WHO World Health Organization

ÚZIS ČR Ústav zdravotnických informací a statistiky ČR

Chapter 1

Introduction

The use of tobacco along with the consumption of alcohol is one of the most serious yet preventable causes of health problems in the Czech Republic, including cardiovascular, tumor, and respiratory diseases (Sovinová *et al.* 2014). Those diseases caused by smoking represent a substantial economic burden on the health system. In recent years, there has been a growing concern about the effects of passive smoking on non-smokers' health, as not only does smoking itself have severe health consequences, but also exposure to tobacco smoke is harmful (Lazuras *et al.* 2009). Second-hand smoke (SHS) contains over 400 chemicals, more than 50 carcinogens, and many toxic substances (NCI 1999). Tobacco use leads to premature deaths of over 8 million people annually, with approximately 1.3 million of them being non-smokers exposed to SHS (IHME 2019). There is no safe level of SHS exposure, even a brief exposure can result in adverse health effects (WHO 2023). In children SHS exposure can lead to respiratory infections, ear infections, and asthma attacks, in infants it can even cause sudden infant death syndrome, an unexplained and unexpected death within the first year of life (IHME 2019). Therefore, exposure to second-hand smoke is a serious public health concern, which should not be neglected.

To address those issues World Health Organization (WHO) adopted the Framework Convention on Tobacco Control (FCTC) in 2003. Being the first global public health treaty against smoking it requires the member states to adopt tobacco control policies to reduce the smoking prevalence and mitigate SHS exposure (WHO 2023). While policies such as imposing taxes on cigarettes, advertisement bans, or repelling cigarette packaging aim at the consumer, i.e. the smoker, smoke-free policies prohibiting smoking in various public places are intended to protect non-smokers from passive smoking. On May 31, 2017, the

Czech Republic adopted legislation to protect the health of its citizens from the harmful effects of addictive substances. This policy, among other things, prohibits smoking in various public places including bars and restaurants.

The objective of this thesis is to examine the effectiveness of this particular smoking ban on the Czech non-smoking population. In the analysis, we will focus on non-smokers as their health should be affected by this smoking ban, given that they are expected to experience less exposure to SHS. The study employs the data from the European Helath Interview Survey (EHIS) from 3 consequent waves: 2009, 2014, and 2019. Two dependent variables were chosen as indicators of health. Firstly, we utilize the logistic regression on a binary response variable representing the presence of a longstanding health problem. Subsequently, we inspect the count variable for the number of nights spent in the hospital during the past 12 months, for which we apply the Zero-inflated Negative Binomial (ZINB). Therefore, the two hypotheses studied in the thesis are:

1. The smoking ban positively influenced, i.e. decreased, the probability of experiencing the longstanding health problem among the non-smoking population.

2. The smoking ban led to a reduction in the expected number of nights spent in the hospital for Czech non-smokers.

To capture the effect of the smoking ban we employ the Difference-in-Differences (DiD) method often used for evaluating the impact of policy changes. The DiD method compares the treatment group that is influenced by the policy to the control group unaffected by the policy change. The non-smoking population of the Slovak Republic will serve as our control group. In the Slovak Republic, no such policy was implemented during the examined period, and we assume no Slovak citizens moved to the Czech Republic due to the smoking ban. Therefore, the treatment is exogenous and there is no selection bias present in our data.

The results of the thesis revealed that the smoking ban led to a significant improvement in the health of Czech non-smokers. Following the implementation, there was a substantial decrease in the probability of an individual experiencing a longstanding health issue, along with a reduction in the expected number of hospital nights over the previous 12 months.

The thesis is structured as follows: Chapter 2 includes statistics on the prevalence of tobacco consumption and other related measures as well as a brief history of smoking and its regulation, Chapter 3 summarizes the existing research on the topic. In Chapter 4, we explain the methodology used for the analysis. Chapter 5 focuses on the data we work with, descriptive statistics are presented, and dependent as well as independent variables are introduced. Regression results are examined in Chapter 6, and finally, in Chapter 7 the results and limitations of the study are discussed along with th motivation for further research.

Chapter 2

Smoking: the harmful sensation

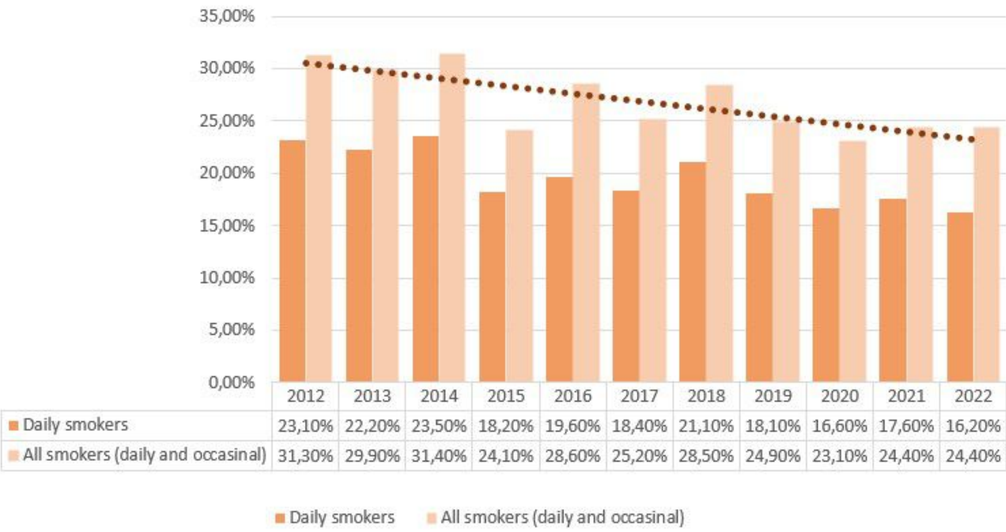
This chapter covers statistics on tobacco consumption in the Czech Republic as well as a brief history of smoking and its regulation.

2.1 Statistics on the prevalence of tobacco use

The National Survey of Tobacco and Alcohol Use conducted in 2022 showed that 24.4% of the Czech population aged 15 and above are smokers. Two-thirds of them, 16.2% of the whole population, are daily smokers, and 8.2% represent occasional smokers, who smoke less than daily, but at least once a month. The share of smokers differs significantly by gender, 30% of all men and 18.7% of all women. Over the long term, we can observe a declining trend in tobacco use since 2012 when the prevalence of Czech smokers was 31.3%, with 23.1% being daily smokers (Figure 2.1). Although we have witnessed some fluctuations during this period, the prevalence was the highest in 2014 (31.4%) but rapidly dropped to 24.1% in 2015. On the other hand, the lowest prevalence was recorded in 2020 (23.1%) which can be partly associated with the Covid-19 pandemic. The decreasing trend in prevalence can be partially explained by the increasing popularity of electronic cigarettes, whose usage rose rapidly from 1.1% in 2013 to 10.2% in 2022.

When we inspect the prevalence across different age groups, we can see that in 2012 most smokers were between the ages of 15 and 24 (around 45%). However, this group experienced a significant decrease from 2012 to 2019, and in 2018 highest shares were reported between the ages of 25 and 44 (35.2%). Since 2020, the prevalence rates among age groups of 15-24, 25-44, and 45-64 fluctuate around 25-28%. The age group 65+ stands out as smoking substantially

Figure 2.1: Prevalence of tobacco use



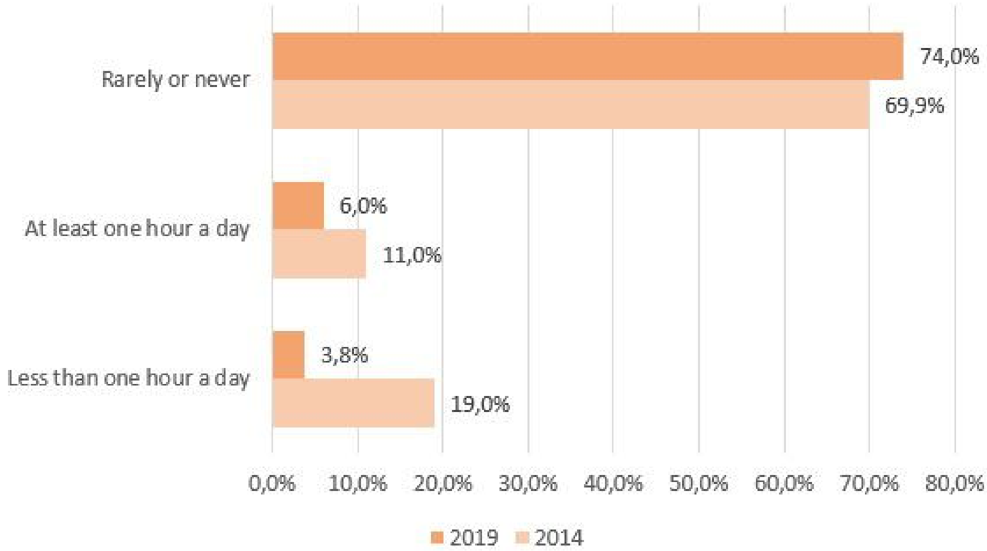
	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Daily smokers	23,10%	22,20%	23,50%	18,20%	19,60%	18,40%	21,10%	18,10%	16,60%	17,60%	16,20%
All smokers (daily and occasinal)	31,30%	29,90%	31,40%	24,10%	28,60%	25,20%	28,50%	24,90%	23,10%	24,40%	24,40%

less. Additionally, individuals with university education tend to have a lower prevalence of tobacco use compared to those with secondary and primary education, while there is no significant difference in smoking rates between urban and rural areas. On average, a daily smoker in 2022 smoked 15-24 cigarettes a day, with men typically smoking more cigarettes per day than women. (Csémy *et al.* 2023)

In 2019, the Czech Republic ranked as the 13th state in the number of daily smokers in the EU with 19.4%. The average among 27 European countries was 18.4%. Bulgaria occupied the top rugs in this ranking with 28.7%, on the other hand, the country with the lowest prevalence was Sweden, where only 6.4% of the population were daily smokers (Eurostat 2019).

Not only first-hand smoking but also exposure to secondhand smoke presents a considerable public health concern. People are usually exposed to secondhand smoke in various enclosed spaces, at their homes, at work, in public places, or in bars and restaurants. In 2019, 30.9% of people aged 15 or more in the EU reported they were exposed to smoke. About half of them (15.4%) experienced this exposure daily, while the rest did so less often. In the Czech Republic, those numbers were below the EU average in 2019. Specifically, 10.7% of Czech citizens reported exposure to tobacco smoke less than once a week, 3.8% for less than 1 hour daily, and 6% were exposed for at least 1 hour every day (Figure 2.2). A total of 74% were exposed to smoke rarely or never. Nevertheless, those statistics were worse in 2014, when 11% reported they were exposed at least 1 hour every day, 19% less than 1 hour every day. In

Figure 2.2: SHS exposure



2014, 69.9% claimed to be rarely or never exposed to smoke (Eurostat 2019).

The greatest exposure to tobacco smoke among the 27 EU countries is in Greece, where 62.4% were exposed to smoke in 2019. On the other hand, countries with the lowest tobacco exposure included Norway with 8.3% or Iceland with 10.9% (Eurostat 2019).

2.2 History of smoking in Europe

The history of tobacco smoking in Europe began in 1492 when two members of Christopher Columbus´s crew, Luis de Torres and Rodrigo de Jerez, became the first Europeans to smoke tobacco. They learned this practice from Cuban natives who dried and smoked the leaves of a plant called Nicotiana tabacum. This plant was widely cultivated throughout the whole American continent where it was used in various ritual practices. (Gilman & Xun 2006)

In 1571 Nicholas Monardes, a doctor from Sevilla, suggested that smoking could ease fatigue and help a person to relax. It was even believed to have certain healing effects and can be used to cure syphilis. The trend of smoking quickly spread from Spain to Portugal, England, and the Netherlands, and was introduced to the French court in the 16th century by Jean Nicot, after whom Nikotin and Nicotina were named. Tobacco possibly found its way to the Czech Republic with the emperor Rudolf II. and his Spanish courtiers. By the end of the 17th century, smoking had become a global phenomenon and a great business (Gilman & Xun 2006). In 1843, the French tobacco industry began producing shredded tobacco in thin papers, giving birth to the *cigarette*. The first cigarette-packing machine was introduced at the 1878 World´s Fair in Paris (Hrych *et al.* 1996).

It was not until the 19th century that scientists started to investigate the effect of tobacco smoking on human health, although the association with a decline in morality persisted. As early as 1761, John Hill linked smoking to cancer of the nasal cavities. Finally, in 1948 the epidemiological studies of lung cancer from Ricard Doll convinced American as well as British authorities that smoking is harmful to our health. All the facts about the harmful effects of smoking were proven and accepted by the medical public in the 1970s.

The history of bans on smoking goes way back to the beginning of the 17th century when King James I banned smoking in England, attributing it to a perceived barbaric custom leading to moral decay. Many attempts at smoking bans followed all over the world, primarily due to safety concerns. However, these efforts proved unsuccessful, as smoking became a habit people were not willing to give up on. (Gilman & Xun 2006)

In 1994 the Ninth World Conference on Tobacco or Health took place in Paris, at which the idea of an international treaty on tobacco control was presented. Nearly a decade later, on May 21st, 2003, the WHO Framework Convention on Tobacco Control was adopted by the World Health Assembly and came into force in 2005 (WHO 2009). The total of 168 countries that signed this treaty shall adopt and implement effective legislative and other measures to help reduce tobacco consumption, nicotine addiction, and exposure to tobacco smoke (WHO 2003). Subsequently, smoking bans have been implemented worldwide in response to this recommendation. The Czech Republic signed and ratified this treaty in 2012 (GGTC 2021).

The primary regulatory framework for tobacco products in the Czech Republic is governed by Act No. 110/1997 Coll. on food and tobacco products. Additional regulations include a ban on the marketing of tobacco products, warning graphics on tobacco products, early and effective prevention for children, supervision of lobbying by tobacco companies, or taxes imposed on tobacco products. Important was the implementation of Act No. 65/2017 Coll. on health protection against the harmful effects of addictive substances, which came into force on the 31st of May 2017 and strictly defines areas where smoking is not allowed (Act 2017). The Czech Republic became the 23rd country to have introduced a complete ban on smoking in restaurants. The first country to do so was Ireland in 2004 (Fišer 2008). In my thesis, I will examine the effectiveness of this particular smoking ban by comparing the data on tobacco consumption and health in the Czech Republic with data from Slovakia where such a ban was not introduced.

Chapter 3

Literature review

Many studies from around the world have demonstrated that smoke-free policies positively influence exposure to secondhand smoke (Fernández *et al.* 2017) (Jankowski *et al.* 2020). For instance, a multi-country study was conducted examining the efficiency of smoke-free laws in reducing exposure to SHS in 7 European countries: France, Greece, Ireland, Italy, Portugal, Turkey, and Scotland (Ward *et al.* 2013). They measured the efficiency of the smoke-free legislation by comparing the data on $PM_{2.5}$, a particulate matter with a diameter of approximately 2.5 micrometers, which has become a widely used marker of SHS (Prignot 2011). Particle concentration levels were recorded in hospitality venues across the countries before and after the implementation of the policy. The results of this study indicate a reduction in $PM_{2.5}$ in all countries. Countries that enforced more fully comprehensive smoke bans (France, Ireland, Italy, Scotland, and Turkey) experienced a greater reduction in SHS exposure than those with only partial smoke-free laws (Greece, Portugal).

Some studies indicate that the impact of smoking bans on SHS varies across different population groups. For example, a study from Scotland (Haw & Gruer 2007) suggests that the reduction in smoke exposure following the prohibition of smoking in all enclosed public spaces and workplaces was greater for non-smokers living in non-smoking households. However, non-smokers living in smoking households continue to experience high levels of secondhand smoke. Sims *et al.* (2012) discovered similar findings in England when examining SHS exposure among groups with different socioeconomic statuses.

The effect of smoking bans on smoking behavior is not as straightforward. Although some studies show that prohibiting smoking in workplaces, transport, and public places leads to a reduction in the smoking prevalence along with the number of cigarettes smoked, and increases the motivation to stop smoking (Kim 2009) (Zablocki *et al.* 2014), most studies do not find a significant change in smoking behavior (Jones *et al.* 2015) (Anger *et al.* 2011) (Lee *et al.* 2011) (Catalano & Gilleskie 2021).

Research on the effect of a smoking ban on the health of the population is scarce. Kuehnle & Wunder (2017) carried out a study examining the effects of smoking bans on the self-assessed health of both smokers and non-smokers in Germany. By employing a difference and differences approach on the data from the Socio-Economic Panel, a nationally representative household survey, they discovered heterogeneous effects for different population subgroups. While the non-smoking population experienced positive and statistically significant improvements in health, smokers reported no or even adverse health effects in response to bans. Other studies emphasize the positive effect of the smoking ban on hospital admissions (Barone-Adesi *et al.* 2006) (Khuder *et al.* 2007). For example, a study from Geneva, Switzerland (Humair *et al.* 2014) showed a significant decrease in the number of hospitalizations for chronic obstructive pulmonary disease as well as for acute coronary syndromes.

In the Czech Republic, the research on the effect of the smoking ban is quite limited. In 2019, a cohort study was conducted by Kulhánek *et al.* (2021) aimed to assess changes in daily cigarette consumption, the ratio of cigarettes smoked in pubs, street, work, and home, and motivation to quit smoking before and post the implementation of the smoking ban in bars and restaurants. However, this study had a small sample size, consisting of only 131 adult smokers, and only measured immediate changes within 2-3 months after the implementation. The findings of this study indicate that there was a statistically significant decrease in the consumption of cigarettes, on average by 1.7 cigarettes per day. The percentage of cigarettes smoked in indoor public places decreased on average by 23.6%, on the other hand, smoking on the streets increased by 19.1%. There was a slight increase in the motivation to quit smoking, although the effect size was relatively small.

This thesis will contribute to the stream of missing research by conducting a difference-in-differences analysis. The common history between Czechia and Slovakia allows us to carry out a quasi-natural experiment in which Czechia will be the treatment country where the smoking ban was introduced in 2017 and Slovakia will be the control country with no regulation on smoking. The thesis will focus on the effect of second-hand smoke.

Chapter 4

Methodology

In this chapter, I will outline the underlying methodology of my thesis: the Difference-in-Differences approach, Logit model, and Zero-inflated Negative Binomial model.

I will examine the health of the Czech population using 2 different variables and employing 2 different models. The first variable I will inspect is the longstanding health problem, a binary response variable equal to 1 if a person is suffering from a longstanding health problem of any kind or 0 otherwise. For this variable, I will use the Logit model. Secondly, I will consider the number of nights spent in hospital over the past 12 months as an indicator of health. Since this is a count variable with a substantial amount of zeros, I will apply the Zero-Inflated Negative Binomial model (ZINB).

4.1 Difference-in-Differences method

The Difference-in-Differences method (DiD), widely used in economics for assessing the effects of policy changes, allows us to conduct a quasi-experiment by comparing the changes in outcomes over time between two groups: the control group and the treatment group. If we did not use the DiD and examined only the treatment group pre-form and post-reform, results would be possibly influenced by trends that are not associated with the smoking ban.

We have available observations from two periods before the policy was implemented in 2017 as well as one period after. The treatment group consists of Czech non-smokers, while Slovak non-smokers form our control group. We assume that any general trends equally influence the control group as they do the treatment group. Thus, our data sample fulfills the criteria that are crucial for adopting the quasi-natural experiment framework (Meyer 1994). Additionally, the selection bias is not present in our data sample as EHIS operates as a randomized household survey.

To find the effect of the smoking ban, we estimate the model

$$health_{it} = \beta_0 + \beta_1 post + \beta_2 treatment + \beta_3 interaction + \beta_4 X_{it} + \epsilon_{it} \tag{4.1}$$

where i´s represent individual and t´s denote time periods. The dependent variable is either the longstanding health problem or the number of nights spent in the hospital for a person i. *Post* is a dummy variable for the post-reform period equal to unity for the year 2019 and equal to 0 for pre-reform years 2008 and 2014, and *treatment* is a dummy variable representing the treatment group. The *interaction* is equal to $post * treatment$, which takes the value 1 for Czech citizens after the implementation of the ban and 0 otherwise. A set of individual characteristics (sex, age, education, household income,BMI, alcohol usage, ...) is represented by vector X_{it}. β_0 denotes the intercept and ϵ_{it} is the error term.

Statistically, the association between the smoking ban implementation and the outcome effect on health is estimated by the interaction term. Therefore, the estimate of β_3 is the parameter of our interest, sometimes referred to as the DiD estimator or the average treatment effect since it measures the effect of the policy on an average outcome of the dependent variable (Wooldridge 2012). If the smoking ban led to an improvement in non-smokers' health, the DiD estimator will be significantly different from zero and negative.

4.2 Logit model

When dealing with a binary dependent variable Linear Probability Model (LPM) is often used for estimation. However, LPM has some important limitations, such as the possibility of fitted probabilities falling below 0 or above 1, and the constant partial effect of any explanatory variable. Fortunately, more sophisticated models for binary response exist to address these drawbacks, including the logit model. Those models take the form

$$P(y = 1|X) = G(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k) \tag{4.2}$$

where $0 < G(z) < 1$, for all $z \in \mathbb{R}$.

In the logit model, G is the logistic function

$$G(z) = \frac{exp(z)}{1 + exp(z)} \tag{4.3}$$

For the estimation of the logit model, OLS is no longer appropriate due to nonlinearity. Instead, the Maximum Likelihood Estimator (MLE) is employed, which is consistent, asymptotically normal, and asymptotically efficient under general conditions. MLE automatically accounts for any heteroskedasticity in $Var(y|X)$ since it is based on the distribution of y given X (Wooldridge 2012).

To interpret the results of a logit model, we inspect the partial effect, also called the marginal effect, on the probability p(x) = P(y = 1 |X), which is expressed as

$$\frac{\partial p(x)}{\partial x_j} = g(\beta_0 + X\beta)\beta_j \tag{4.4}$$

where $g(z) = \frac{\partial G(z)}{\partial z}$ and x_j is roughly continuous variable. However, our explanatory variable of interest (the interaction term) is binary, and thus the partial effect of changing the interaction term from zero to one, representing the post-reform treatment group, is simply

$$G(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 + \cdots + \beta_k x_k) - G(\beta_0 + \beta_+ x_1 + \beta_2 x_2 + \cdots + \beta_k x_k) \tag{4.5}$$

None of the assumptions of linear regression apply to the logit models, however, some other assumptions still have to hold as (Stoltzfus 2011). Firstly, the dependent variable needs to be binary. Secondly, observations must be independent of each other. The third assumption is a large sample size. Our model satisfies those three assumptions. Fourth is the assumption of no perfect multicollinearity among independent variables. To assess the validity of this as-

Table 4.1: Variation inflation factor

Explanatory variable	VIF
post	1.492256
treatment	1.386915
interaction	1.921949
age	1.192066
sex	1.011316
education	1.045475
income	1.064233
members	1.137239

Figure 4.1: Age vs. log odds of LHP

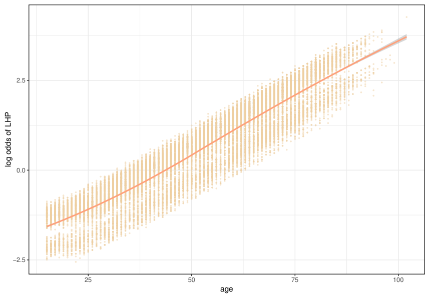
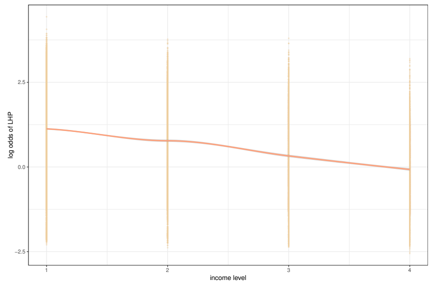


Figure 4.2: Income vs. log odds of LHP



sumption we compute the Variation Inflation Factor (VIF) for each explanatory variable, depicted in Table 4.1, we will employ the generalized version since our models include at least one categorical variable. VIF measures the collinearity among explanatory variables, the higher the VIF the higher the collinearity. A variance inflation factor higher than 5 suggests that there is multicollinearity in our regression. Since all VIFs are lower than 2, we can claim that this assumption is satisfied (Nahhas 2023). The fifth assumption is the linearity between the log odds of the dependent variable and independent variables. This assumption has to be tested only for the continuous independent variables, in our case just for the variables *age* and *income*, which is defined in the Likert scale but we treat it as continuous in the regression. Therefore, we examine the relationships by plotting *age* and *income* against the predicted log odds and adding a smoothing line. The plot is shown in Figure 4.1 and Figure 4.2, we can see that this assumption is satisfied. The last assumption is that there are no extreme values in our sample, which also holds.

To determine which explanatory variables should be included in our model, we conduct the Likelihood Ratio (LR) test and inspect various measures of fit of models. The LR test is based on the difference in the log-likelihood functions for the restricted and unrestricted models. The likelihood ratio statistics used for the LR test is defined as

$$LR = 2(L_{ur} - L_r) \tag{4.6}$$

where L_{ur} is the log-likelihood for the unrestricted model and L_r for the restricted one (Wooldridge 2012). Under the null hypothesis, the unrestricted model does not significantly improve the fit of the model, so if we fail to reject the null hypothesis the restricted model is preferred. Moreover, the pseudo

(McFadden´s) R^2 for binary response will be examined:

$$R^2 = 1 - \frac{L_{ur}}{L_0} \tag{4.7}$$

where L_0 is the log-likelihood function for a model with only an intercept and L_{ur} is the log-likelihood for the estimated model.

Last but not least we investigate the values of the Akeike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) that take into account not just the model fit but also the complexity of the model by penalizing for a large number of predictors. The AIC metric, first introduced by Akaike (1973), is defined as

$$AIC = -2ln(L(\theta)) + 2k \tag{4.8}$$

where k is the number of parameters and $L(\theta)$ is the log-likelihood of the estimated model. Schwarz (1978) defined the BIC as

$$BIC = -2ln(L(\theta)) + kln(n) \tag{4.9}$$

where k and $L(\theta)$ are the same as above, n is the sample size. As we can see, for large samples BIC has a higher penalty for a high number of parameters in the model compared to AIC. To evaluate the fit of the model, we aim to minimize both AIC and BIC (Aho *et al.* 2014). By observing those criteria we compare models with different regressors and choose the preferred model with the best fit.

4.3 Zero-inflated negative binomial model

Poisson distribution is the basis for the Poisson regression model often used for modeling count data, with the probability density function

$$P(Y = y) = \frac{exp(-\mu)\mu^y}{y!} \tag{4.10}$$

where $\mu = exp(x\beta) = E(y)$ (Wooldridge 2012). However, this distribution is often too restrictive, since the Poisson assumption needs to hold: $Var(y) = E(y)$. In our sample Var(y)=43.57 and E(y)=1.5, indicating overdispersion ($Var(y) \neq E(y)$) that needs to be accounted for. Consequently, we will use the Negative Binomial (NB) model with an additional dispersion parameter α.

The probability density function for NB distribution is as follows

$$P(Y = y) = \frac{\Gamma(y + \alpha^{-1})}{y!\Gamma(\alpha^{-1})}(\frac{\alpha^{-1}}{\alpha^{-1} + \mu})^{\alpha^{-1}}(\frac{\mu}{\alpha^{-1} + \mu})^{y} = \phi \tag{4.11}$$

Γ is the gamma function (Long & Freese 2006). Additionally, we will extend the NB model into a zero-inflated alternative, since our data sample contradicts the standard negative binomial prediction of a lower probability of a zero count. The variable *hospital_nights* exhibits an excessive number of zeros, precisely around 86%. On the other hand, the zero-inflated model introduced by Lambert (1992), allows zeros to be generated by 2 distinct processes. Long & Freese (2006) defines two unobserved groups:

1. *Always Zero Group*

    - An individual in this group has an outcome of 0 with a probability of 1.
    - In our sample, those will be the respondents who had experienced a health problem during the past year but were not hospitalized. There may be two distinct reasons for not being hospitalized - either the individual decided not to visit the doctor or the doctor decided not to hospitalize them even though they should have been.
    - Let A=1 if someone is in an *Always Zero Group* and A=0 otherwise. The probability of individual i being in the *Always Zero Group* can be modeled using the logit model with the density function

    $$P(A_i = 1|z_i) = \psi_i \tag{4.12}$$

    where z_i is referred to as the inflation variable since it inflates the number of zeros in the zero-inflated model.

2. *Not Always Zero Group*

    - An Individual in this group might have an outcome equal to zero, but there is a positive probability that she has a nonzero count.
    - This group will include two kinds of respondents: those who DiD not have a health problem and were not hospitalized and those that had a health problem and were hospitalized.

- The probability of an individual i being in the *Not Always Zero Group* is expressed as

$$P(A_i = 0|z_i) = 1 - \psi_i \tag{4.13}$$

The ZINB analysis is then conducted in three steps:

1. The probability of being the *Always Zero Group* is modeled using the logit model.

2. Counts for the *Not Always Zero Group* are modelled, using NB model to indicate the probability of each count.

3. Observed probabilities are computed as a combination of the probabilities associated with the two groups.

The ZINB distribution can be written as

$$P(Y = y) = \begin{cases} \psi(0) + (1 - \psi(0))\phi(0), & \text{if } y = 0 \\ (1 - \psi(0))\phi(y), & \text{if } y > 0 \end{cases} \tag{4.14}$$

And conditional mean

$$E(y|x) = \mu(1 - \psi) \tag{4.15}$$

where ψ (defined in Equation 4.13) is the density function of the binary process (logit) and ϕ (defined in Equation 4.11) is the density of the count process (NB).

Similarly to the logit model, we will use several measures of fit including the pseudo R^2, AIC, BIC as well as the LR test to choose the final model.

Chapter 5

Data

For the analysis, we used the data from the European Health Interview Survey (EHIS), a private household survey that targets the population aged 15 years old and above. The survey focuses on 4 areas of information:

- health status including variables on self-perceived health status or the prevalence of some diseases,
- health care use with variables on the number of visits to a general practitioner or medical specialist, number of hospitalizations, or usage of medicine,
- health determinants variables, such as smoking behavior, alcohol consumption, or eating habits, and
- socio-economic background variables - sex, age, level of education, household income, ...

The first wave of the survey was carried out between 2006 and 2009 in 17 EU countries. Until 2019, the survey took place every 5 years, as of 2019 onwards in 6-year intervals. In the later waves, all 27 EU countries participated in the survey (Eurostat 2024). In the Czech Republic (CZ), the data is collected by

Table 5.1: Number of observations

Year	CZ	SK
2008/2009	1955	4972
2014	6737	5490
2019	7993	5527
Total:	32674	

Table 5.2: Representation of smoking and nonsmoking population

Country	Freq. of nonsmokers	in %	Freq. of smokers	in %	Freq. of NAs
CZ	12426	74.5%	4254	25.5%	5
SK	11561	72.3%	4382	27.4%	46
Total	23987	73.4%	8636	26.4%	51

the Institute of Health Information and Statistics of the Czech Republic (Ústav zdravotnických informací a statistiky ČR (ÚZIS ČR)) with the cooperation of the Czech Statistical Office, in the Slovak Republic (SK) by the Statistical Office of the Slovak Republic. For our study, we employed three waves of the survey that took place in 2009 (in SK the first wave was carried out in 2008), 2014, and 2019. Table 5.1 shows the number of participants in each wave.

The data were provided through contracts with the aforementioned authorities and therefore can not be attached to the thesis. They are available upon request from the author subject to the limitations of the providing institutions. We obtained the data in 6 different data sets, where variables were differently named sometimes even differently defined. Subsequently, all variables were renamed and those essential for the analysis were redefined so that they could be further investigated. Unfortunately, some variables that we anticipated to utilize in the study (such as the prevalence of cancer) were not included in all waves and therefore can not be examined. Moreover, we divided the data into 2 sub-samples: nonsmokers and smokers (which can be further separated into occasional and daily smokers). The representation of each group is shown in Table 5.2. From now on, we will focus only on the non-smoking population. For the purpose of this thesis, we generated a new sub-sample containing only those variables that will be utilized in the models. Any observations carrying missing values were excluded from the sample and a total of $22,717$ observations remained for further analysis. In the next two sections, we will inspect both explained and explanatory variables in more detail, Table 5.3 shows their description.

Table 5.3: Variable description

Variable name	Description	Answer categories
LHP	Longstanding health problem	0 No 1 Yes
hospital nights	Number of nights spent as a patient in a hospital in the past 12 months	number 0 - 365
post	dummy variable representing period post smoking ban	0 observations before the smoking ban 1 observations after the smoking ban
treatment	dummy variable representing the treatment group	0 control group (SK) 1 treatment group (CZ)
interaction	$post * treatment$	1 treatment group post smoking ban 0 other observations
age	Age of respondent	number 15 - 102
sex	Sex of respondent	1 Male 2 Female
education	Highest level of education completed	1 No formal/Primary 2 Lower secondary 3 Upper secondary 4 Post-secondary non-tertiary 5 Bachelors or equivalent level 6 Masters, Doctoral or equivalent level
income	Level of household net monthly income	1 low 2 lower middle 3 upper middle 4 high
members	Number of members in a household	Number 1 - 13

5.1 Dependent variables

5.1.1 Longstanding health problem

Longstanding Health Problem (LHP) is a binary response variable taking the value 1 if an individual is suffering with such a problem and 0 otherwise. The frequency distribution of this variable is shown in Table 5.4, we can see that the larger part of our sample, precisely 62.2% does experience a LHP. In the second

Table 5.4: Londstanding health problem distribution

Longstanding health problem	2009	2014	2019	**Total**
1	2815	6053	5865	**14733**
0	2066	2913	4259	**9238**

Table 5.5: Summary statistics of *hospital nights*

Descriptive statistics	Value
Min	0
Max	90
Mean	1.436
Median	0
Variance	31.045
Standard deviation	5.572
Skewness	7.529
Kurtosis	83.006
Number of observations	22,703

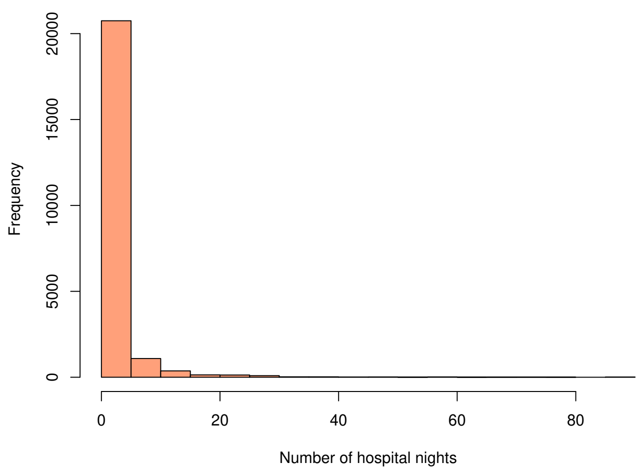
wave, the proportion of people having a health problem was the highest when 67.5% of respondents reported the presence of a longstanding health problem. This value dropped to 57.9% in 2019. In the first wave, the percentage of people suffering from LHP was equal to 57.6%, though this wave had substantially fewer observations compared to the two later waves.

5.1.2 Number of nights spent in hospital over the past 12 months

The variable *hospital nights* is a count variable with a long right tail and a great proportion of zeros (86%). The maximum value of the entire sample is 270, however, only 0.01% of the sample exceeds 90 nights spent in the hospital. Table A.1 displays the detailed frequency distribution of this variable. Some of the high values can be caused by measurement errors, while others may be consequences of unique accidents that are not directly associated with health determinants. Consequently, we exclude those individuals from the sample and employ the ZINB model on 22,703 observations. The distribution plot of *hospital nights* is shown in Figure 5.1. Based on the histogram it is obvious that *hospital nights* have a decreasing tail and does not follow a normal distribution.

Table 5.5 summarizes the descriptive statistics of *hospital nights*. The values we obtained are consistent with the previous statements. The excessive

Figure 5.1: Distribution of *hospital nights*



presence of zeros is captured by both the mean and the median. High variance indicates that our data is spread over a wide range. In a normally distributed dataset, skewness is equal to 0 and kurtosis is typically 3. The value of skewness (7.529) in our sample suggests that data are highly right-skewed. The kurtosis of 83.006 points out that we have more data in the tails than would be expected in a normal distribution, in other words we observe thicker tails.

5.2 Independent variables

In this section, all independent variables used in the analysis are described, and corresponding descriptive statistics are shown in Table 5.6. Variables related to the effect of the smoking ban - *post*, *treatment*, and *interaction* - have already been outlined in Section 4.1.

- **AGE**: The variable *age* denotes the age of an individual and takes on values from 15 to 102. The elderly population prevails in our sample compared to younger individuals, 34.4% are people aged 65 and above, 37.4% belong to the age group between 40 and 65, while only 9.5% individuals in our sample are younger than 25 years old. We assume that older people are of poorer health and therefore the sign of the coefficient should be positive for both logit and ZINB models.

- **SEX**: The variable *sex* is equal to 1 for males and 2 for females. Accord-

Table 5.6: Descriptive statistics of explanatory variables

Variable	Min	Median	Mean	Max	Std.Dev.
post	0	0	0.438	1	0.496
treatment	0	1	0.519	1	0.499
interaction	0	0	0.258	1	0.438
age	15	55	53.22	102	19.4
sex	1	2	1.607	2	0.488
education	1	3	3.441	6	1.228
income	1	2	2.14	4	1.145
members	1	2	2.487	13	1.361

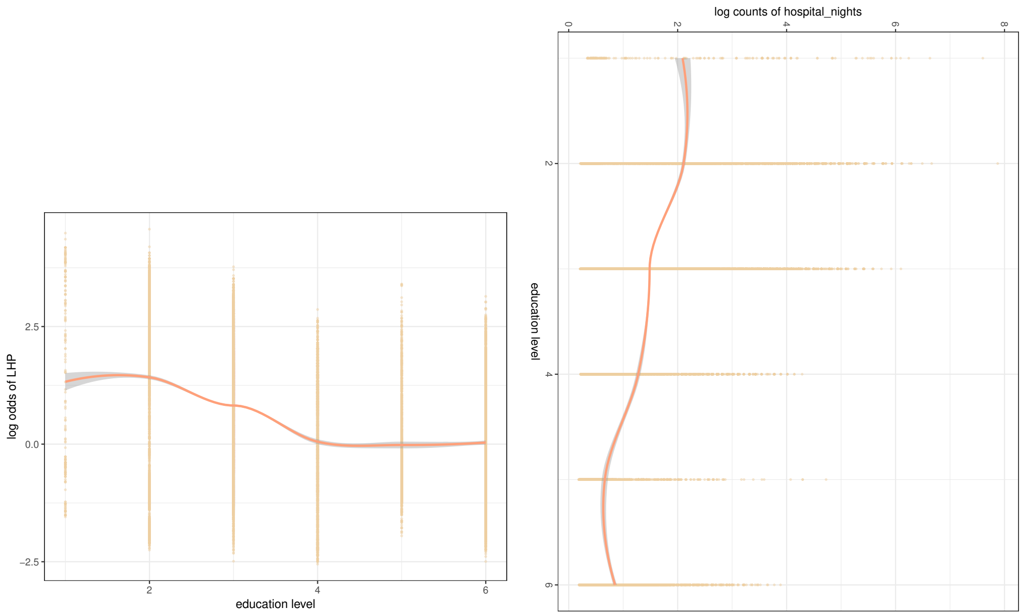
ing to the mean, there are more females in our sample, precisely 60.3%, and the remaining 39.7% are male.

- **EDUCATION**: The highest level of education completed is described by the variable *education*. Individuals with no formal or only primary education are the least represented group in our sample accounting only for approximately 0.8%. The majority of individuals, 55.7%, have completed post-secondary non-tertiary education. Other educational groups make up between 5% and 15% of the sample. We anticipate that a higher level of education is associated with better health, and thus we expect to observe negative coefficients for this variable. This phenomenon along with the effect of income can be described by the Grossman model which is discussed in Chapter 6 in more detail.

- **INCOME**: The variable *income*, defined on a Likert scale ranging from 1 to 4, denotes the corresponding level of the net monthly income of a household. The majority of our respondents (41.5%) belong to low-income households, while the group least represented in our sample are respondents from high-income households, accounting only for 18%. We expect this variable to have a positive relationship with health, as individuals living in wealthier households tend to have healthier lifestyles. The sign of the coefficients should be negative.

- **MEMBERS**: Number of household members is denoted by variable *members* and takes on values between 1 and 13. According to the mean and median, households with fewer members are more frequent in our sample. The effect of household size is expected to have a positive impact on one´s health. As shown in some studies, social integration and social support lead to positive health outcomes(Berkman *et al.* 2000). Additionally, the health of adults is positively influenced by the role of parents, unless they are single parents, in those cases, parenting usually has adverse health effects (Hughes & Waite 2002).

We anticipated including variables *income* and *education* as continuous into the regression in order to make the interpretation easier. While *income* appeared to have a linear relationship with the dependent variables, *education* recorded a non-linear relationship (as illustrated in Figure 5.2). Accordingly, we tried to redefine *education* into 3 categories to accomplish a linear relationship, taking the value 1 for low education, 2 for middle education, and 3 for high education. However, the adjustment led to the estimated coefficient not being statistically significant. Therefore, we decided to include *income* as a continuous variable and *education* as categorical.



(a) Education vs. log odds of LHP

(b) Education vs. log counts of hospital nights

Figure 5.2: Non-linearity of education variable

The correlation matrix is shown in Table A.2. We do not observe any significantly higher correlation between variables that might indicate multicollinearity in our models. The highest correlation was recorded between *interaction* and *post*, precisely 0.6688, which is not surprising since the interaction term is equal to $post * treatment$. The correlation between *interaction* and *treatment* is therefore relatively high as well (0.5674). Additionally, a moderate correlation of 0.5326 is detected between *age* and *members*, nevertheless, it is not sufficiently strong to create any issues in the analysis.

Chapter 6

Results

In this chapter, we will assess the results of our models. The analysis was carried out using the R studio environment.

6.1 Logit model

Firstly, we will inspect various measures of fit as demonstrated in chapter 4 to decide on the specification of the model. Table 6.1 presents those measures for the following 3 models: **model 1** including the following independent variables: *post*, *treatment*, *interaction*, *sex*, *age*, *education*, *income* and *members*, **model 2** with explanatory variables *post*, *treatment*, *interaction*, *sex*, *age*, and **model 3** which is an intercept-only model.

Let us comment on the log-likelihood first, which is the highest for model 1 (-11730.69). We perform LR tests comparing model 1 with model 2 as well as with the intercept-only model. In both cases, we obtain a p-value < 0.00001, suggesting we should utilize model 1 as it significantly increases the accuracy of our model. McFadden´s R^2 supports the results of the LR tests, it is the highest for model 1, exactly 0.2211. Moreover, both AIC and BIC highlight the

Table 6.1: Measures of Fit for Logit model

Measures of fit	model 1	model 2	model 3
Log-Lik	-11706.45	-11815.49	-15061.1
McFadden´s R^2	0.2227362	0.2154964	0
AIC	23438.9	23642.98	30124.21
BIC	23543.3	23691.16	30132.24
LR test for overall significance of model 1		$\chi^2 = 6709.3$	$Prob > \chi^2 = 0.0000$

preferred model. We intend to minimize these criteria, which is again achieved with model 1. The model is overall statistically significant with LR = 6661 and p-value = 0.000.

Finally, we estimate the model by regressing the variable LHP on *post*, *treatment*, *interaction*, *sex*, *age*, *education*, *income* and *members*. Results of the regression are shown in Table 6.2. All explanatory variables are statistically significant at the 0.1 significance level. The coefficient estimates we obtained from the regression are in the log odds scale, which is difficult to interpret. To make the interpretation easier we can obtain the odds ratios as $odds\_ratio = exp(log\_odds\_ratio)$, which are often used to compare the effects among two different groups, i.e. treatment group vs. control group. The odds ratio is defined as the probability of an event occurring divided by the probability of an event not occurring, however, they are often misinterpreted as probability. To interpret the results in the probability scale, we acquire the average marginal effects (AME), which determine the change in the predicted probability of LHP with a unit change in the independent variable, ceteris paribus (Norton *et al.* 2019). AME is equal to the average of individual marginal effects across the sample (Wooldridge 2012), marginal effects were defined in Section 5.2. Both odds ratios and marginal effects are also depicted in Table 6.2.

First of all, we will focus on the variables related to the treatment effect. Variable *post* is statistically significant and has a negative sign. The odds ratio is equal to exp(-0.178051)= 0.8368997 and AME=-0.03974228, so the probability of experiencing a LHP is 3.9% lower after the smoking ban compared to pre-reform periods. A similar impact is recorded for the *treatment* variable, which is also statistically significant although only on the 99% confidence interval with a negative coefficient sign. The odds of having LHP are 0.9149568 times for an individual from the treatment group (CZ) compared to the control group (SK). Yet, the most important variable in the evaluation of the effect of the smoking ban is the *interaction* variable, which is statistically significant at all levels with a p-value < 0.00001. Additionally, the sign of the coefficient is negative and the odds ratio is equal to exp(-0.693162)=0.4999928. The marginal effect is equal to -0.16071357, therefore the smoke-free policy caused the probability of LHP to decrease by 16.1%. Those results indicate that the smoking ban had a substantially positive impact on the health of the Czech non-smoking population.

The effect of *age* is positive as expected, the probability of LHP is 1.4% higher with a marginal increase in *age*. The estimated coefficient for vari-

Table 6.2: Results of the Logistic regression

Variable	Coef. estimate	Std. Error	z value	P>|z|	Odds ratio	Average marginal effect
Intercept	-1.308081	0.226074	-5.786	7.21e-09 ***		
post	-0.178051	0.048128	-3.700	0.000216 ***	0.8368997	-0.03974228
treatment	-0.088878	0.044505	-1.997	0.045820 *	0.9149568	-0.01976121
interaction	-0.693162	0.069657	-9.951	< 2e-16 ***	0.4999928	-0.16071357
age	0.062341	0.001105	56.398	< 2e-16 ***	1.0643252	0.01387035
sex	0.156302	0.032922	4.748	2.06e-06 ***	1.1691794	0.03495539
education*2	-0.581043	0.213131	-2.726	0.006406 **	0.5593150	-0.13620880
education*3	-1.006024	0.208652	-4.822	1.42e-06 ***	0.3656702	-0.21664718
education*4	-1.192796	0.215203	-5.543	2.98e-08 ***	0.3033717	-0.28673694
education*5	-1.064787	0.217314	-4.900	9.59e-07 ***	0.3448014	-0.25736409
education*6	-1.261413	0.211980	-5.951	2.67e-09 ***	0.2832534	-0.30170253
income	-0.081069	0.014651	-5.533	3.14e-08 ***	0.9221300	-0.01803718
members	-0.042247	0.013331	-3.169	0.001529 **	0.9586325	-0.00939971

significance levels: * p<0.1, ** p<0.05, *** p<0.01

able *sex* has a positive sign as well, suggesting that men are healthier than women. The odds ratio is equal to 1.1691794 and AME=0.03495539, in other words, women are 3.5% more likely to suffer from LHP than men, holding all other factors constant. The variable *members* is statistically significant on the 95% confidence interval with a negative sign indicating a positive influence on health. With an increase in household size by one member the predicted probability of a household member experiencing LHP is 0.94% lower. As we assumed both *eductaion* and *income* appear to have a positive effect on health as the coefficients are negative. An individual with a lower secondary education has a 13.62% lower predicted probability of having a LHP compared to an individual with a lower level of education. Similarly, the probability is 21.66% lower for a person with upper secondary education, 28.67% lower for a person with post-secondary non-tertiary education, 25.74% lower for a person with a bachelor´s degree, and 30.17% lower for a person with master´s of doctoral degree compared to a person with no formal or primary education. Moreover, with an increase in household income by one level the predicted probability of LHP is lowered by 1.7%. Those two effects can be explained by the Grossman model. Grossman (2017) views health as a "durable capital stock", which can be increased through investment. The *income* effect appears in two forms. First is the "wage effect" - the higher one´s wage the greater the opportunity cost of being ill. Therefore wealthier people tend to care about their health more. Secondly, wealthier people have more resources for investment in the health stock. The impact of *education* is described by the fact that more educated demand a larger stock of health and are more efficient producers of health, in other words, they produce more health stock from each unit of health investment.

Table 6.3: Measures of Fit for ZINB model

Measures of fit	full model	intercept-only model
Log-Lik	-19084	-19716
R^2 for zero-inflated models	0.471	0
AIC	38222	394337.98
BIC	38438.81	39462.07
LR test for overall significance	$\chi^2 = 1264$	$Prob > \chi^2 = 0.0000$
Vuong test	z=-22.02181	p-value < 2.22e-16

6.2 Zero-Inflated Negative Binomial model

For the zero-inflated negative binomial regression, we choose the same specification as for the logit model. Therefore, we regress *hospital nights* on *post*, *treatment*, *interaction*, *sex*, *age*, *education*, *income* and *members*. We conduct the LR test for the overall significance of the model. The likelihood-ratio statistics is equal to 1259, with a p-value < 0.0001 we can reject the null hypothesis that all of the coefficients in the ZINB model are equal to zero, therefore we can claim that our model is preferred over an empty model. The outcome of the LR test is supported by other measures of fit presented in Table 6.3. To test whether the zero-inflated model is a better fit for our data than the standard negative binomial alternative, we conducted the Vuong test. With z-statistics of approximately -22.14 and one-sided p-value < 0.0001, we reject the null hypothesis and assert that the zero-inflation component in the ZINB model significantly improves the model.

The ZINB model includes two sub-models, as explained in Section 4.3. The first part employs a standard negative binomial regression to determine the count process of the "Not Always Zero group" including respondents that were hospitalized (for those the value of *hospital nights* is a strictly positive integer), and respondents that were not hospitalized since they did not have any health problems (dependent variable *hospital nights* is equal to zero but with non-zero probability). Meanwhile, the second sub-model is a logit model focusing on the likelihood probability of being in the "Always Zero group", taking on value 1 if an individual belongs to this group and 0 otherwise. The "Always Zero group" consists of individuals who were not hospitalized even though they experienced a health problem. Results are depicted in Table 6.4. The $log(\alpha)$ presented in this table is the logarithm of the dispersion parameter for the count process. The p-value lower than 0.0001 suggests that α is significantly different from zero and therefore the utilization of a negative binomial regression is preferred to Poisson due to excess number of zeros.

Firstly, we will focus on the count process. All variables associated with the effect of the smoking ban are statistically significant at all appropriate levels of significance. The sign of the *interaction* variable is negative, indicating an improvement in the health of the Czech population following the implementation of the policy. The DiD estimator is equal to -0.549889 so the log count of nights spent in the hospital is 0.549889 times lower, which is not particularly informative, therefore, we convert them by taking the exponential of the estimated coefficient: exp(-0.549889)= 0.5770138. Now, we can say that the smoking ban reduced the expected number of nights spent in a hospital approximately by a factor of 0.5770138, keeping all other variables constant. Another independent variable that appears to be a significant predictor of the variable *hospital nights* is *age*. The exponential of the log count is equal to 1.0106194, thus with a marginal increase in *age* the number of nights spent hospitalized is expected to increase 1.0106194 times. Other explanatory variables are statistically insignificant.

The zero-inflated part of the model holds less relevance for our analysis since the probability of being in the "Always Zero group" is logically not associated with the effect of the smoking ban. Respondents in this group meet the criteria for being hospitalized and yet are not, therefore they have a 100% probability of a zero outcome. To explain how this happens, we will briefly mention how the Czech and Slovak health care system works. When a person is having a health issue they may decide to see a doctor, who will then assess the health condition of the individual and decide on further treatment. In other words, the physician is the one who decides whether a person should be hospitalized. Consequently, the fact that an observation belongs to the "Always Zero group", and takes on the value of 1 if an individual is not hospitalized despite having a health issue, can be caused by the respondent deciding not to visit the doctor when having a health problem or the doctor deciding not to hospitalize the respondent although (s)he has a problem. The estimated coefficients can be therefore biased since we do not have any information about the physician.

When we inspect the coefficient estimates and corresponding p-values, we can see that variables *age*, *education*, *income*, and *members* appear to have a statistically significant impact on the probability of an always zero outcome. The coefficient of *interaction* is not statistically significant, the smoking ban did not have a significant impact on the odds of being in the "Always Zero group". The same implies for variable *sex*, the gender of the respondent does

Table 6.4: Results of the ZINB regression

Variable	Coef. estimate	Std. Error	z value	P>|z|	exp(coefficient)
count					
Intercept	1.338539	0.197619	6.773	1.26e-11 ***	3.8134661
post	0.321520	0.052235	6.155	7.50e-10 ***	1.3792220
treatment	0.456610	0.047098	9.695	< 2e-16 ***	1.5787131
interaction	-0.549889	0.073107	-7.522	5.41e-14 ***	0.5770138
age	0.010563	0.001205	8.767	< 2e-16 ***	1.0106194
sex	-0.012864	0.036384	-0.354	0.724	0.9872188
education*2	0.027834	0.160654	0.173	0.862	1.0282252
education*3	0.028357	0.157945	0.180	0.858	1.0287624
education*4	0.039682	0.168875	0.235	0.814	1.0404798
education*5	-0.100005	0.184817	-0.541	0.588	0.9048332
education*6	-0.092098	0.167667	-0.549	0.583	0.9120158
income	-0.017401	0.018202	-0.956	0.339	0.9827499
members	0.001989	0.015361	0.129	0.897	1.0019905
$Log(\alpha)$	0.221340	0.038297	5.780	7.49e-09 ***	
zero-inflation					
Intercept	2.639801	0.245549	10.751	< 2e-16 ***	14.0104202
post	0.023454	0.060456	0.388	0.698049	1.0237315
treatment	0.099348	0.054584	1.820	0.068743	1.1044504
interaction	0.023465	0.085654	0.274	0.784127	1.0237420
age	-0.030264	0.001409	-21.476	< 2e-16 ***	0.9701894
sex	0.030464	0.042272	0.721	0.471117	1.0309327
education*2	0.426342	0.205384	2.076	0.037910 *	1.5316439
education*3	0.572226	0.201856	2.835	0.004585 **	1.7722078
education*4	0.692287	0.214365	3.229	0.001240 **	1.9982797
education*5	0.678853	0.227295	2.987	0.002820 **	1.9716149
education*6	0.786684	0.211173	3.725	0.000195 ***	2.1961021
income	0.099542	0.020324	4.898	9.7e-07 ***	1.1046649
members	-0.043530	0.018385	-2.368	0.017898 *	0.9574042

significance levels: * p<0.1, ** p<0.05, *** p<0.01

not significantly change the odds.

- With a marginal increase in *age* the odds of being in the "Always Zero group" decrease 0.9702784 times. To put it differently, it applies that for older individuals the zero outcome is less likely generated by the fact that they did not visit the physician or the physician did not hospitalize them compared to a younger person.

- The odds of being in the "Always Zero group" are 1.5475929 times higher for a person with lower secondary education, 1.7901649 times higher for a person with upper secondary education, 2.0120153 times higher for a person with post-secondary non-tertiary education, 1.9817602 times higher for a person with a bachelor degree or equivalent level of education, and 2.1875401 times higher for someone with master or doctoral degree compared to a person with no formal or primary education. This suggests a non-linear relationship between education and the odds of the always zero outcome, possibly related to increased opportunity cost associated with hospitalization at certain educational levels.

- The odds of being in the "Always Zero group" are 1.1046649 higher with a marginal increase in *income*. In other words, with increasing levels of income, the zero outcome is more likely generated by the fact that the respondent does not visit a doctor or the doctor did not hospitalize them. Those results can be partially explained by the fact that individuals with higher incomes face greater opportunity costs when being hospitalized.

- With a marginal increase in household size the odds of being in the "Always Zero group" decrease 0.9574042 times. This effect could be explained by the influence of the household members on the decision of an individual. For instance, a person may initially opt not to visit a physician but is persuaded to do so by other members of the household.

Chapter 7

Conclusion

This bachelor´s thesis investigated the effectiveness of the smoking ban in bars and restaurants in the Czech Republic. Introducing public smoking bans has become a common practice to decrease second-hand smoke (SHS) exposure and influence smoking behavior. The impact of these bans on SHS exposure is well-documented in the scientific literature, many studies observed a significant reduction in SHS exposure following the implementation of a public smoking ban. On the other hand, the evidence regarding their effect on changes in smoking behavior varies across studies, while some report a statistically significant decline in cigarette consumption, others do not substantiate this finding. Some research focused on the effect of the policy on the health of the population, observing mostly positive outcomes. This thesis estimated the effect of a smoking ban only on the non-smoking population in the Czech Republic.

In the Czech Republic, the smoking ban in bars and restaurants is described by Act No. 65/2017 Coll. on health protection against the harmful effects of addictive substances, which came into force on 31 May 2017. No comprehensive study was carried out in CR evaluating the health outcome associated with the introduction of this smoking regulation. Only a small sample size study was conducted assessing the impact of this particular ban on the smoking behavior of Czech citizens. Therefore this thesis makes a valuable contribution to the existing research on the effect of smoking restrictions on the health of the Czech population.

To examine the effect of the policy on the health of Czech non-smokers, we employ a difference-in-differences method comparing the Czech non-smoking population (treatment group) with the Slovak non-smoking population (control group). In the Slovak Republic, no such comprehensive ban was adopted.

The thesis exploits the EHIS dataset from 2009 to 2019 with two waves prior to the implementation of the Czech smoking ban and one-period after the implementation. We assumed that the treatment and control group are equally influenced by the general trends, which allows us to conduct a quasi-natural experiment and utilize the DiD approach.

We considered two dependent variables as an indicator of health. Firstly, the thesis studied the probability of experiencing a longstanding health problem (LHP) by employing a logistic regression on our data sample. We found a statistically significant change in the probability of LHP after the introduction of the policy. The predicted probability decreased by 16.1% suggesting a substantial improvement in the health of Czech non-smokers. Other explanatory variables that proved to have a significant influence on the probability of LHP include age, sex, level of education, household income and number of members in the household.

Furthermore, we estimated a Zero-inflated negative binomial model with the number of nights spent in the hospital during the past 12 months as a dependent variable. The ZINB regression consists of two submodels that allow the excessive zeros to be generated by two distinct processes: (1) "Not Always Zero group" includes respondents who did not suffer from a LHP and therefore were not hospitalized, (2) "Always Zero group" consists of individuals who experienced a health problem and yet were not hospitalized. The count process of the "Not Always Zero group" was modeled using a negative binomial model, while the second submodel employed logistic regression estimating the probability of an individual being in the "Always Zero group". Those who decided not to consult a physician or cases when the doctor decided not to hospitalize them despite health problems take the value of 1 in the logit submodel. Those who were not hospitalized because they were healthy and had no reason for it, take the value of 0. The results of this analysis revealed a statistically significant reduction in the expected number of nights spent in the hospital following the adoption of the smoking ban. On the other hand, the policy did not have a significant impact on the probability of being in the "Always Zero group", which is expected since the likelihood of being in this group is connected to the decision of an individual rather than their health status and is thus not associated with the effect of the smoking ban. The determinants that appeared to be statistically significant were age, education level, and household income.

To conclude our results suggest that a 2017 smoking ban in bars and restaurants significantly improved the health of the non-smokers in Czechia. Those results are robust across two model specifications using two alternative dependent variables and are consistent with the findings of other researchers (Kvasnicka *et al.* 2018), (Kuehnle & Wunder 2017).

The author of this thesis is aware that the study has some limitations. Certainly, alternative variables might be better suited for this analysis. The longstanding health problem variable that we used includes a wide range of health conditions, investigating solely diseases potentially caused by SHS exposure would be more relevant. Similarly, the variable on the number of nights spent in the hospital encompasses individuals who experienced accidents or underwent operations not related to SHS exposure, potentially leading to biased results of our analysis. The dependent variable on the index of chronic diseases, which can be obtained from the EHIS data, should be investigated. In further research, we will study the effect of the policy on the health of the smoking population and compare the findings with those for the non-smoking population. Furthermore, with subsequent waves of the European Health Interview Survey, there will be an opportunity to assess the effectiveness of the smoking ban over a longer post-implementation period.

Bibliography

(2017): "Protikuřácký zákon." Zákon číslo 65/2017 Sb.

Aho, K., D. Derryberry, & T. Peterson (2014): "Model selection for ecologists: The worldviews of aic and bic." *Ecology* **95(3)**: pp. 631–636.

Akaike, H. (1973): "Maximum likelihood identification of gaussian autoregressive moving average models." *Biometrika* **60(2)**: pp. 254–265.

Anger, S., M. Kvasnicka, & T. Siedler (2011): "One last puff? public smoking bans and smoking behavior." *Journal of health economics* **30(3)**: pp. 591–601.

Barone-Adesi, F., L. Vizzini, F. Merletti, & L. Richiardi (2006): "Short-term effects of italian smoking regulation on rates of hospital admission for acute myocardial infarction." *European heart journal* **27(20)**: pp. 2468–2472.

Berkman, L. F., T. Glass *et al.* (2000): "Social integration, social networks, social support, and health." *Social epidemiology* **1(6)**: pp. 137–173.

Catalano, M. A. & D. B. Gilleskie (2021): "Impacts of local public smoking bans on smoking behaviors and tobacco smoke exposure." *Health economics* **30(8)**: pp. 1719–1744.

Csémy, L., Z. Dvořáková, A. Fialová, M. Kodl, M. Malỳ, & M. Skỳvová (2023): "Užívání tabáku a alkoholu v české republice 2022." *Státní zdravotní ústav* .

Eurostat (2019): "Smoking of tobacco products by sex, age and educational attainment level." online. [cit. 2023-12-08]. Retrieved from:https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Tobacco_consumption_statistics.

Eurostat (2024): "European helath interview survey." online. [cit. 2024-04-09]. Retrieved from: https://ec.europa.eu/eurostat/web/microdata/european-health-interview-survey.

Fernández, E., M. Fu, M. Pérez-Ríos, A. Schiaffino, X. Sureda, & M. J. López (2017): "Changes in secondhand smoke exposure after smoke-free legislation (spain, 2006-2011)." *Nicotine & Tobacco Research* **19(11)**: pp. 1390–1394.

Fišer, M. (2008): *Evropská unie a boj proti kouření.* Master's thesis, Jihočeská univerzita.

GGTC (2021): "Global tobacco industry intergerence index." Online. [cit. 2024-01-15]. Retrieved from: https://globaltobaccoindex.org/country/CZ.

Gilman, S. L. & Z. Xun (2006): *Příběh kouře: člověk a kouření od úsvitu dějin až po současnost.* dybbuk.

Grossman, M. (2017): "On the concept of health capital and the demand for health." In "Determinants of health: an economic perspective," pp. 6–41. Columbia University Press.

Haw, S. J. & L. Gruer (2007): "Changes in the exposure of adult non-smokers to secondhand smoke after implementation of smoke-free legislation in scotland: a national cross-sectional survey." *British Medical Journal* .

Hrych, E. *et al.* (1996): *Slavná historie tabákového dỳmu.* Praha: Forma.

Hughes, M. E. & L. J. Waite (2002): "Health in household context: Living arrangements and health in late middle age." *Journal of health and social behavior* **43(1)**.

Humair, J.-P., N. Garin, E. Gerstel, S. Carballo, D. Carballo, P.-F. Keller, & I. Guessous (2014): "Acute respiratory and cardiovascular admissions after a public smoking ban in geneva, switzerland." *PloS one* **9(3)**.

IHME (2019): "Global burden of disease." *Technical report*, Institute of Health Metrics and Evaluation.

Jankowski, M., V. Rees, W. S. Zgliczyński, D. Kaleta, M. Gujski, & J. Pinkas (2020): "Self-reported secondhand smoke exposure following the adoption of a national smoke-free policy in poland: analysis of serial, cross-sectional, representative surveys." *British Medical Journal* **10(9)**.

Jones, A. M., A. Laporte, N. Rice, & E. Zucchelli (2015): "Do public smoking bans have an impact on active smoking? evidence from the uk." *Health economics* **24(2)**: pp. 175–192.

Khuder, S. A., S. Milz, T. Jordan, J. Price, K. Silvestri, & P. Butler (2007): "The impact of a smoking ban on hospital admissions for coronary heart disease." *Preventive medicine* **45(1)**: pp. 3–8.

Kim, B.-S. (2009): "Workplace smoking ban policy and smoking behavior." *Journal of preventive medicine and public health* **42(5)**: pp. 293–297.

Kuehnle, D. & C. Wunder (2017): "The effects of smoking bans on self-assessed health: Evidence from germany." *Health Economics* **26(3)**: pp. 321–337.

Kulhánek, A., K. Lukavská, I. Švancarová, H. Fidesová, & R. Gabrhelík (2021): "Changes in tobacco use patterns and motivation to quit related to the new smoke-free legislation in the czech republic." *Journal of Public Health* **43(2)**: pp. 348–354.

Kvasnicka, M., T. Siedler, & N. R. Ziebarth (2018): "The health effects of smoking bans: Evidence from g erman hospitalization data." *Health Economics* **27(11)**: pp. 1738–1753.

Lambert, D. (1992): "Zero-inflated poisson regression, with an application to defects in manufacturing." *Technometrics* **34(1)**: pp. 1–14.

Lazuras, L., A. Rodafinos, D. B. Panagiotakos, J. R. Thyrian, U. John, & E. Polychronopoulos (2009): "Support for smoke-free policies in a pro-smoking culture: findings from the european survey on tobacco control attitudes and knowledge." *International Journal of Public Health* **54**: pp. 403–408.

Lee, J. T., S. A. Glantz, & C. Millett (2011): "Effect of smoke-free legislation on adult smoking behavior in england in the 18 months following implementation." *PloS one* .

LONG, J. S. & J. FREESE (2006): *Regression Models for Categotical Dependent Variables Using Stata*, volume 7. Stata press.

MEYER, B. D. (1994): "Natural and quasi-experiments in economics." *Journal of Business & Economic Statistics* **13(2)**: pp. 151–161.

NAHHAS, R. (2023): "Introduction to regression methods for public health using r." *Dayton: Creative Common* .

NCI (1999): *Health effects of exposure toenvironmental tobacco smoke: the report of the CaliforniaEnvironmental Protection Agency.* National Cancer Institute.

NORTON, E. C., B. E. DOWD, & M. L. MACIEJEWSKI (2019): "Marginal effectsâ€"quantifying the effect of changes in risk factors in logistic regression models." *Jama* **321(13)**: pp. 1304–1305.

PRIGNOT, J. J. (2011): "Recent contributions of air- and biomarkers to the control of second-hand smoke (shs): a review." *International Journal of Environmental Research and Public Health* .

SCHWARZ, G. (1978): "Estimating the dimension of a model." *The Annals of Statistics* pp. 461–464.

SIMS, M., J. S. MINDELL, M. J. JARVIS, C. FEYERABEND, H. WARDLE, & A. GILMORE (2012): "Did smokefree legislation in england reduce exposure to secondhand smoke among nonsmoking adults? cotinine analysis from the health survey for england." *Environmental health perspectives* **120(3)**: pp. 425–430.

SOVINOVÁ, H., L. CSÉMY, & V. KERNOVÁ (2014): *Užívání tabáku a alkoholu v České republice: Zpráva o situaci za období posledních deseti let.* Státní zdravotní ústav Praha.

STOLTZFUS, J. C. (2011): "Logistic regression: a brief primer." *Academic emergency medicine* **18(10)**: pp. 1099–1104.

WARD, M., L. M. CURRIE, Z. KABIR, & L. CLANCY (2013): "The efficacy of different models of smoke-free laws in reducing exposure to second-hand smoke: a multi-country comparison." *Health Policy* **110(2-3)**: pp. 207–213.

WHO (2003): "Who framework convention on tobacco control." *WHO FCTC* World Health Organization.

WHO (2009): "History of the who framework convention on tobacco control." *WHO FCTC* World Health Organization.

WHO (2023): online. World Health Orhanization [cit. 2024-04-17]. Retrieved from: https://www.who.int/news-room/fact-sheets/detail/tobacco.

Wooldridge, J. M. (2012): *Introductory Econometrics: A Modern Approach*. 5 edition.

Zablocki, R. W., S. D. Edland, M. G. Myers, D. R. Strong, C. R. Hofstetter, & W. K. Al-Delaimy (2014): "Smoking ban policies and their influence on smoking behaviors among current california smokers: a population-based study." *Preventive medicine* **59**: pp. 73–78.

Appendix A

Appendix

Table A.1: Frequency distribution of variable hospital nights

Number of nights spend in hospital	Frequency	Percentage	Cumulative percentage
0	19855	85.9%	85.9%
1	170	0.7%	86.6%
2	261	1.1%	87.7%
3	302	1.3%	89%
4	264	1.1%	90.1%
5	274	1.2%	91.3%
6	206	0.9%	92.2%
7	336	1.5%	93.7%
8	220	1%	94.7%
9	104	0.4%	95.1%
10	235	1%	96.1%
11	73	0.3%	96.4%
12	60	0.3%	96.7%
13	26	0.1%	96.8%
14	158	0.7%	97.5%
15	59	0.3%	97.8%
16	18	0.08%	97.9%
17	13	0.06%	97.9%
18	18	0.08%	98%
19	7	0.03%	98%
20	87	0.4%	98.4%
21	97	0.4%	98.8%

Table A.1: Frequency distribution of variable hospital nights

Number of nights spend in hospital	Frequency	Percentage	Cumulative percentage
22	8	0.03%	98.8%
24	7	0.03%	98.9%
25	19	0.08%	98.9%
26	10	0.04%	99%
27	2	0.008%	99%
28	18	0.08%	99.1%
29	1	0.004%	99.1%
30	61	0.3%	99.4%
31	3	0.01%	99.4%
32	4	0.02%	99.4%
33	4	0.02%	99.4%
34	1	0.004%	99.4%
35	14	0.06%	99.5%
36	2	0.008%	99.5%
37	4	0.02%	99.5%
38	2	0.008%	99.5%
39	1	0.004%	99.5%
40	16	0.07%	99.6%
41	1	0.004%	99.6%
42	6	0.03%	99.6%
43	2	0.008%	99.6%
44	3	0.01%	99.6%
45	3	0.01%	99.6%
46	3	0.01%	99.7%
48	1	0.004%	99.7%
49	1	0.004%	99.7%
50	13	0.06%	99.7%
53	3	0.01%	99.7%
55	1	0.004%	99.7%
56	3	0.01%	99.8%
57	1	0.004%	99.8%
58	1	0.004%	99.8%
60	15	0.06%	99.8%
63	1	0.004%	99.8%

Table A.1: Frequency distribution of variable hospital nights

Number of nights spend in hospital	Frequency	Percentage	Cumulative percentage
65	1	0.004%	99.8%
68	2	0.008%	99.8%
69	1	0.004%	99.8%
70	2	0.008%	99.8%
71	1	0.004%	99.9%
75	5	0.02%	99.9%
77	1	0.004%	99.8%
78	1	0.004%	99.8%
80	2	0.008%	99.8%
90	13	0.06%	99.9%
92	1	0.004%	99.9%
97	1	0.004%	99.9%
100	2	0.008%	99.9%
119	1	0.004%	99.9%
120	1	0.004%	99.9%
125	1	0.004%	99.9%
141	1	0.004%	99.9%
150	3	0.01%	99.9%
155	1	0.004%	99.9%
180	1	0.004%	99.9%
270	1	0.004%	100%

Table A.2: Correlation matrix

	LHP	hospital nights	post	treatment	interaction	sex	age	education	income	members
LHP	1.0000									
hospital nights	0.1398	1.0000								
post	-0.0740	0.0027	1.0000							
treatment	-0.0176	0.0473	0.1245	1.0000						
interaction	-0.1063	0.0021	0.6688	0.5674	1.0000					
age	0.4856	0.1529	0.0780	0.1527	0.0874	1.0000				
sex	0.0788	0.0169	0.0042	-0.0104	-0.0133	0.0872	1.0000			
education	-0.1492	-0.0608	0.1251	0.0336	0.1270	-0.1369	-0.0614	1.0000		
income	-0.1645	-0.0707	-0.0604	-0.0439	-0.0950	-0.2573	-0.1522	0.2763	1.0000	
members	-0.2702	-0.0850	-0.0457	-0.1382	-0.0709	-0.5326	-0.0802	0.0489	0.1629	1.0000