Charles University

Faculty of Social Sciences

Institute of Economic Studies

Who is more prone to depression?

Analysis of micro-level data of patients with cancer.

Bachelor's thesis

Author: Julie Kristýna Balážová

Study program: Economics and Finance

Supervisor: PhDr. Mgr. Jana Votápková, Ph.D.

Year of defense: 2024

Declaration of Authorship

The author hereby declares that 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.

Prague, April 29, 2024

Julie Kristyna Balazova

Abstract

Depression is very common in cancer patients, affecting about 1 in 5 people with such disease. This thesis uncovers determinants potentially contributing to the development of depression in patients with cancer diagnosis. Special emphasis is placed on how human mental health has responded to the COVID-19 pandemic. The source of our cross-sectional dataset of U.S. population aged 18+ is NHIS. Two dependent variables are examined. The probability of taking medication for depression or anxiety is analyzed using logistic regression, and severity of depression symptoms is estimated by multinomial logistic regression. Several robustness checks are implemented. Additionally to coronavirus symptoms, other regressors include sociodemographic characteristics, household composition, educational attainment, health status and life satisfaction. The results show that women are more prone to depression, regardless of cancer diagnosis. The coronavirus symptoms significantly affect depression among people without cancer, but play no role for people diagnosed with cancer. Older people with cancer are less likely to develop depression, and household composition has vital impact on mental health of all respondents, with the exception of cancer survivors. Education is insignificant for patients in cancer treatment.

Keywords	depression determinants, cancer patients, COVID-19, logistic regression, multinomial logistic regression
Title	Who is more prone to depression? Analysis of micro-level data of patients with cancer.
Author's e-mail	65363587@fsv.cuni.cz
Supervisor's e-mail	jana.votapkova@fsv.cuni.cz

Abstrakt

Deprese je u pacientů s rakovinou velmi častá a postihuje asi 1 z 5 lidí s tímto onemocněním. Tato práce odkrývá determinanty potenciálně přispívající k rozvoji deprese u pacientů s diagnózou rakoviny. Zvláštní důraz je kladen na to, jak pandemie COVID-19 ovlivnila lidské mentální zdraví. Zdrojem našeho souboru dat o populaci USA ve věku 18+ je NHIS. Jsou zkoumány dvě závislé proměnné. Pravděpodobnost užívání léků na depresi nebo úzkosti analyzujeme pomocí logistické regrese, a závažnost příznaků deprese pomocí multinomické logistické regrese. Je použito také několik testů robustnosti. Mezi další nezávislé proměnné, mimo koronavirové symptomy, patří sociodemografické charakteristiky, uspořádání domácnosti, dosažené vzdělání, zdravotní stav a celková spokojenost se životem. Výsledky ukazují, že ženy jsou více náchylnější k depresím, bez ohledu na diagnózu rakoviny. Příznaky koronaviru významně ovlivňují depresi u lidí bez rakoviny, ale nehrají žádnou roli u lidí s diagnostikovanou rakovinou. Starší lidé s rakovinou jsou méně náchylní k rozvoji deprese, a uspořádání domácnosti má zásadní dopad na všechny respondenty s výjimkou pacientů, kteří se z rakoviny už vyléčili. Vzdělání je pro pacienty v onkologické léčbě nepodstatné.

Klíčová slova	determinanty deprese, pacienti s rakovinou, COVID-19, logistická regrese, multinomická logistická regrese
Název práce	Determinanty deprese u pacientů s rakovinou
E-mail autora	65363587@fsv.cuni.cz
E-mail vedoucího práce	jana.votapkova@fsv.cuni.cz

Acknowledgements

I would like to express my gratitude to my supervisor, PhDr. Mgr. Jana Votápková, Ph.D. for her time, patience, valuable suggestions and guidance throughout the time of writing this thesis.

This output was supported by the NPO "Systemic Risk Institute" number LX22NPO5101, funded by European Union - Next Generation EU (Ministry of Education, Youth and Sports, NPO: EXCELES).

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

Bibliographic Record

Balážová, Julie Kristýna: *Who is more prone to depression? Analysis of micro-level data of patients with cancer.* Bachelor's thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies, Prague. 2024, pages 72. Advisor: PhDr. Mgr. Jana Votápková, Ph.D.

Contents

List of Tables

List of Figures

Acronyms

AME Average Marginal Effects

CDC Centers for Disease Control and Prevention

CFS chronic fatigue syndrome

GED General Educational Development

HPV human papillomavirus

IQR interquartile range

LR likelihood ratio

MLE Maximum Likelihood Estimation

MOÚ Masarykův onkologický ústav

NCHS National Center for Health Statistics

NHIS National Health Interview Survey

PMDD Premenstrual dysphoric disorder

PMS Premenstrual syndrome

Chapter 1

Introduction

According to WHO (2024), about 1 in 5 people will develop cancer in their lifetime. It results in cancer being the leading cause of death worldwide, accounting for one in six deaths. Furthermore, up to 20 % of people diagnosed with cancer suffer from depression (Breitbart, 2018). Such prevalence is significantly higher than in the general society, where approximately 280 million people in the world have depression, which represents 5 % of the whole population (WHO, 2023). Depression is also the main cause of suicide. Every year, more than 700 000 people die due to suicide, which makes it the fourth leading cause of death worldwide for people between 15 and 29 years of age.

The first year of the COVID-19 pandemic led to a so-called wake-up call, as there was an enormous 25 % increase in the incidence of depression and anxiety. WHO (2022b) states that a major explanation for such a massive increase was social isolation at the time, together with fear of suffering and death. Robinson et al. (2022) further mention estimates indicating that the development of mental disorders was more likely for people with pre-existing physical health issues, such as cancer.

The objective of this thesis is to identify determinants of depression for people aged 18+ suffering from cancer, with a particular focus on the effect of coronavirus symptoms. For the analysis, apart from the examination of the whole sample of respondents and subsample of only cancer respondents, several robustness checks will be implemented - cancer respondents will be split into two samples based on how many years it is since their cancer diagnosis ($t \leq 7$ represents patients with cancer diagnosed during the last 7 years and $7 < t \leq 15$ represents patients diagnosed with cancer 8-15 years ago), and subsample of non-cancer respondents. In each sample, logistic regression and multinomial logistic regression are employed

on two different dependent variables describing taking the medication for depression or anxiety and severity of depression states. Employment of several samples serves for the examination whether the determinants of mental disorder change for different groups of people.

The results suggest that better health status and general well-being reduce the odds and severity of depression for all people. The same effect is observed for age, but with the exception of non-cancer respondents. Moreover, living with another adult, most likely a partner, also decreases the chance of depression occurrence, however, not for cancer survivors. Contrary to that, being a woman considerably affects the prevalence of this mental disorder in the world population, especially in monthly intervals, whereas coronavirus symptoms significantly increase the likelihood of depression for people without cancer diagnosis, while there is no effect for people fighting cancer.

The thesis is structured as follows: Chapter 2 contains literature review. It comments on the types of cancer and acknowledges risk factors contributing to cancer development. Moreover, it describes the relationship between cancer and depression and classifies depression as a non-cancer cause of death. Chapter 3 describes the dataset, chosen variables, and provides adjustments of the original dataset. Chapter 4 introduces the two methodological concepts applied for our investigation. In Chapter 5, results of all our analysis are presented, which is followed by Chapter 6 where the results are properly discussed and a brief comparison of all cancer and non-cancer respondents is portrayed. And finally, Chapter 7 summarizes our entire work. Suggestion for the further extension of the research is also included.

Chapter 2

Literature review

Cancer is a worldwide health complication that can affect any part of the human body of any individual. Experts all over the world have been searching for effective treatment of cancer. Due to a close connection between cancer and depression, we will analyze the factors potentially contributing to the development of depression among cancer patients. This chapter will investigate which patients are more secured, as successful treatment often begins with a balanced psyche. First we will identify the most common types of cancer and their causes, then we will describe the relationship between depression and cancer, and finally we will explain how depression affects the suicidal mortality rate in patients suffering from cancer, as suicide may be a proxy for worsened mental well-being, and thus functions as a non-cancer cause of death.

2.1 Types of cancer

Cancer (further also "malignant tumor" or "neoplasm") is a result of a carcinogenesis process by which a normal cell is transformed into a tumor cell. When tumor cells reach their programmed size and touch other cells, they do not stop growing like normal cells, but continue to grow regardless of the surrounding tissues. When these cells get into the wall of blood or lymphatic vessels, they settle in the lymph nodes, where they form metastases. Widespread metastases are then the leading cause of cancer death.

In total there are over 200 types of cancer in the world, and Cancer Research UK (2020) classifies them in two ways. First, depending on where cancer forms in the body, we distinguish for instance lung cancer or breast cancer. Second, depending on where cancer originates, we distinguish for example brain and

spinal cord cancers, carcinoma, leukemia, lymphoma and myeloma and sarcoma (Cancer Research UK, 2020). In 2020, there were 18.1 million cancer cases in the world, of which 9.3 million cases were men and 8.8 million were women. Breast cancer takes the lead with 12.5 % of the total number of new cases, followed by lung cancer accounting for 12.2 % of the total number (WCRF, 2022).

WCRF (2022) states that among men, lung cancer was the most common type, contributing to 15.4 % of all newly diagnosed cancer cases in 2020. Together with colorectal cancer and prostate cancer, they accounted for 41.9 % of all cancer cases. Among women, breast cancer was the most common type, contributing to 25.8 % of all newly diagnosed cancer cases in 2020, and together with colorectal and lung cancers, they accounted for 44.5 % of all cancer cases.

Approximately 400,000 children and adolescents are newly diagnosed with cancer each year (WHO, 2021). The likelihood of curing childhood cancer depends on the country where the child lives. When it comes to high income countries, the likelihood that the child will be cured is more than 80 %, while if the child comes from low-income or middle-income countries, the likelihood is less than 30 % (WHO, 2021). According to WHO (2021), the most common types of childhood cancer are brain tumors, leukemia, lymphomas and solid tumors. Solid tumours, that mostly occur at childhood or adolescence, account for 30 % of all the tumours diagnosed for these age cohorts (St. Jude Children's Research Hospital, 2024). Solid tumors include, for example, neuroblastoma (a cancer usually found in the belly) and Wilms (a cancer found in one or both kidneys) tumors. These are tumors that do not contain any liquid or cysts and occur in several places, such as bones, muscles or organs. There are two main types of solid tumors - carcinomas and sarcomas, and both types are operable (St. Jude Children's Research Hospital, 2024).

The incidence of malignant neoplasms in the Czech Republic is gradually increasing, although a slowdown in this growth has been observed in recent years. In 2018, as many as 87,361 diseases were diagnosed in the Czech Republic, while the incidence of these diseases has been higher for men than for women. The ratio of this representation is 1.2:1, respectively (ÚZIS, 2021). Among the most common types were prostate, colon and rectum, breast, and lung cancers. The most common type of cancer among men in the Czech Republic was prostate cancer, which accounted for 25 % of all new cancer cases in men. The incidence of this disease continues to rise, with 7,938 new cases diagnosed in 2018. On the other hand, the most common type of cancer among women in the Czech Republic was breast cancer, which accounted for 26.5 % of all new cases among women. The

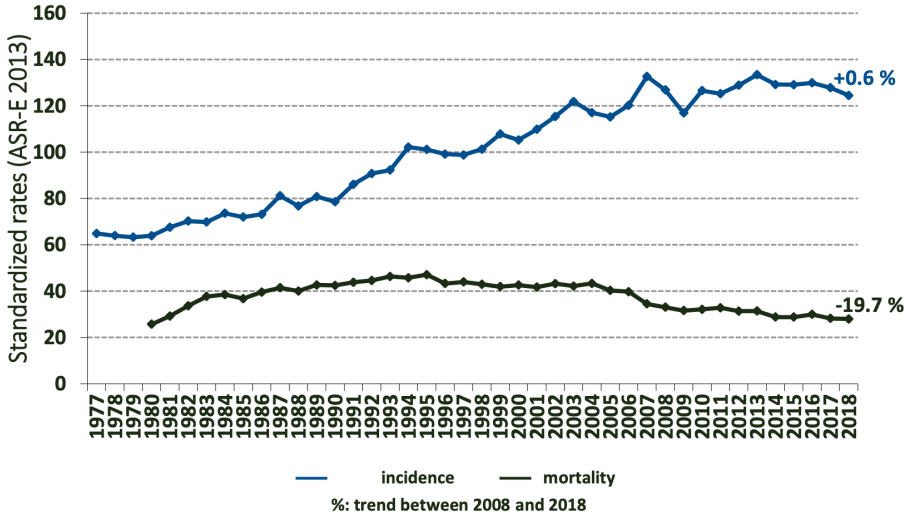
incidence of this disease is also increasing, with a sign of a slight stabilization in recent years (7,182 new cases diagnosed in 2018).

The cancer survival rate is now close to the EU average (European Commission, 2021). Actually, this rate has been gradually improving since the end of the 20th century, and for patients diagnosed between 2010 and 2014, it reached almost the same level as the EU average. There can be several reasons, but the main one is probably the establishment of specialized oncology care centers from EU financial funds. The most common cancers in the Czech Republic, prostate cancer and breast cancer, had survival rates of 85 % and 81 %, respectively, for patients diagnosed between 2010 and 2014, while the EU23 survival rates are 87 % and 82 %, respectively (European Commission, 2021).

As for mortality, the Czech Republic is among the countries with the highest cancer burden both in Europe and in the world. With more than 27,000 people dying of any type of neoplasm here every year, cancer is among the most common causes of death, following cardiovascular diseases (Dusek et al., 2014). It accounts for 28 % of all deaths in men and 23 % of all deaths in women. Dusek et al. (2014) also argue that, in addition to the cancer burden increasing with demographic aging of the population, incidence of multiple tumours within new incidence of cancer accounts for 15 %.

During the first decade of the 21st century, the overall incidence of cancer increased with a growth index of +27.6 %, while the death rate stabilized over time at a growth index of -5 %. Nevertheless, European Commission (2017) reports that over recent years, the Czech Republic has seen significant improvements in life expectancy and mortality, with both measures approaching the European average. ÚZIS (2018) strongly supports this with the fact that between the years 2008 and 2018 there has been a decrease in mortality of breast cancer and colorectal cancer by tens of percent. This is due to the introduction of screening programs that help to find tumors or, in the case of colorectal cancer, precancerous lesions, in earlier stages, and thus allowing cancer screening procedures to reduce mortality rate. Figure 2.1 shows that between 2008 and 2018 the incidence rate of breast cancer was gradually increasing, while the mortality rate decreased by 19.7 %. Čabanová (2024) argues that in the population of patients with breast cancer, thanks to screening programs three quarters of the disease is found in the first stage, which enables a subsequent decrease in mortality.

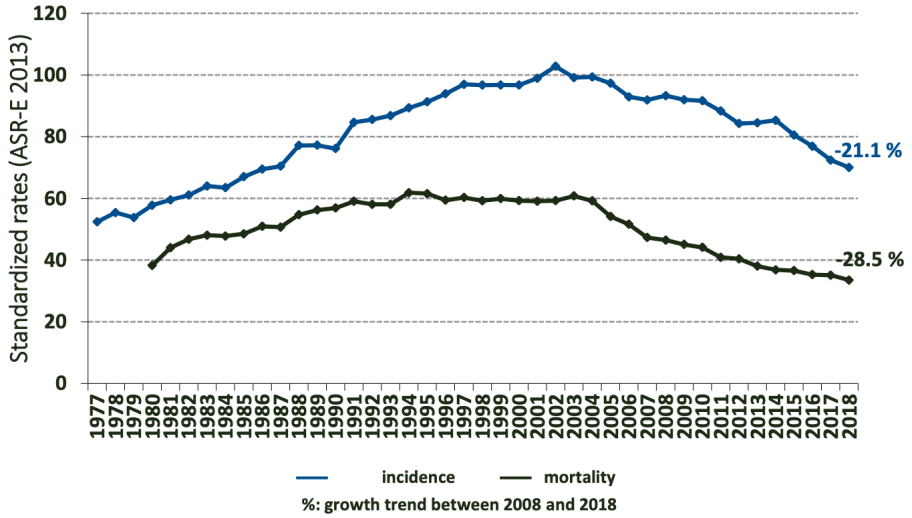
Figure 2.1: Time trends of breast cancer incidence and mortality for women



Source: ÚZIS (2018).

Further as Figure 2.2 demonstrates, between 2008 and 2018, both incidence rate and mortality rate of colorectal cancer decreased substantially by 21.1 % and 28.5 %, respectively. According to Čabanová (2024), screening is definitely the reason for lower mortality. Further, in recent years there has been a problem with the participation of patients with a positive test for occult gastrointestinal bleeding for a subsequent colonoscopy (only about 60 % of patients with a positive test will undergo a colonoscopy), which may be behind the decreased incidence.

Figure 2.2: Time trends of colorectal cancer incidence and mortality



Source: ÚZIS (2018).

2.2 Risk factors

WHO (2022a) states that the transition from normal cells to tumor cells is the result of a synergy between a person's genetic factors and categories of external agents. There are three such categories, and they include chemical carcinogens, physical carcinogens and biological carcinogens. Mesothelioma Web (2016) characterizes chemical carcinogens as molecular substances found in various compounds such as tobacco smoke, aflatoxin, and arsenic. Water-insoluble particles of soft or hard material then constitute physical carcinogens, including, for example, ultraviolet radiation or ionizing radiation, which is found in medical x-rays. As far as biological carcinogens are concerned, according to WHO (2022a) these are different types of infection, such as infections from certain viruses, bacteria, or parasites.

WHO (2022a) places great emphasis on biological carcinogens, as far as chronic infections are concerned. This is a problem mainly in low-income and middle-income countries, where in 2018 approximately 13 % of diagnosed cancer cases were associated to carcinogenic infections such as Helicobacter pylori, human papillomavirus (HPV), hepatitis B virus, hepatitis C virus, and Epstein-Barr virus. Both types of hepatitis increase the risk of liver cancer, and some types of HPV increase the risk of cervical cancer. In addition, people infected with HIV have a significantly higher risk of developing other types of cancer, such as Kaposi's sarcoma[1], and their risk of developing cervical cancer increases up to six-fold (WHO, 2022a).

Furthermore, risk factors for the development of cancer include, among others, physical inactivity, unhealthy diet, and last but not least, age. With increasing age, cellular repair mechanisms become less efficient and, at the same time, the accumulation of risks for specific types of cancer increases (WHO, 2022a).

Fitch (2022) uses three forms of prevention strategies to reduce the burden of cancer. The first of these is primary prevention, a key control strategy to prevent the incidence of cancer in the first place and the most cost-effective approach of all. It is a form of prevention where we should try to reduce our exposure to potentially modifiable cancer risk factors. This includes tobacco cessation, avoiding ultraviolet radiation exposure or maintaining a healthy body weight. Based on this, 33-50 % of all cancer cases could be prevented (Fitch, 2022). Secondary prevention strategy is then important for the early diagnosis of cancer, when

[1] Kaposi's sarcoma is a cancer that develops from cells that line the lymphatic or blood vessels and appears as tumors on the skin, in lymph nodes, inside the mouth, or in other organs.

the disease is most likely to occur. This includes cancer screening, as it helps detect premalignant lesions and thus delay the progression of cancer. Further Moleyar-Narayana and Ranganathan (2020) say that in screening, it is important to identify a person with the potential possibility of cancer occurrence, and it is a key point in reducing the incidence of cancer. The strategy of tertiary prevention serves to enable access to quality and necessary treatment for patients who are applicants for screening and they need to detect secondary malignancies early (Greene, 2016).

2.3 Depression and cancer

According to Deshields (2017), distress is a common reaction to cancer, because we suddenly feel that we have no control over our own situation. In addition, Breitbart (2018) states that about 15 % to 20 % of all cancer patients struggle with depression. Depression is a type of affective disorder, and it can be accompanied by severe anxiety states. Symptoms include exhaustion, sleep and concentration disorders, or immune disorders, and typically arise as a result of experienced trauma. It has psychological, biochemical (clinical) or biological causes. The aforementioned arise on the basis of the depressive effect of an experience or event. The second arise on the basis of an imbalance in the neurochemistry of the brain. Biological causes include genes, hormones, and brain chemicals.

Since having cancer is a depressing experience, many cancer patients are exposed to depression, even if only by the fear of getting this mental disorder (Kneier, 2012). However, the way a person reacts to a cancer diagnosis, and the emotions they feel, is not depression in itself. It develops into depression when other factors, that occurred during life before the diagnosis of cancer itself, contribute to it. Life history provides the context in which cancer is experienced. Therefore, if a person does not face depression during cancer, they most likely will not face it in normal life without cancer.

Cancer can trigger several feelings in a person that create the basis for depression. The fear of death is a typical feeling when diagnosed with cancer, which can trigger depression. Furthermore it can be uncertainty, fear of treatment and side effects or fear of changes in one's own body as a result of illness and treatment (Hong et al., 2022).

The likelihood of depression of people with cancer is up to five times higher than in the general population (Hartung et al., 2017). Among cancer patients, women have the highest burden of depression, where up to 30 % of women with

cancer suffer from depression. Of these, 22 % are lung cancer cases and 21 % are breast cancer cases. However, Pudrovska (2010) argues that cancer has a greater psychological impact on men than on women. She reasons that male masculinity suffers from cancer treatment as much as it does from losing control over one's own body.

A reduced 5-year chance of survival for people with cancer is associated with increased depressive symptoms after controlling for several things such as number of positive lymph nodes, tumor size or histopathological grade (Mausbach and Irwin, 2017). In their study, Spiegel and Giese-Davis (2003) define some reasons that depression may increase the risk of mortality in cancer patients, such as that depression may affect neuroendocrine correlates of stress that promote neoplastic growth, and stress hormones may then suppress immune resistance to tumors. Depression can affect the ability to manage daily activities. In case of cancer patients, it can significantly affect their treatment. Hartung et al. (2017) also explain that people with cancer and cancer-related depression show lower attachment to treatment, due to reduced compliance with medical prescriptions, preventive screening procedures or healthy lifestyle recommendations, and even have a lower survival rate than patients without depression. Nevertheless, compared to non-depressed patients, depressed oncology patients aged 65 and older have 38 % more outpatient health clinic visits and 61 % higher health care costs (Patrick et al., 1997). The American Society of Clinical Oncology acknowledges anxiety and depression to be an important factor, thus they have introduced screenings of cancer patients for psychological disorders (Andersen et al., 2023). Such screenings help physicians to uncover a thorough health status of their patients, such as that patients's high distress is a result of previous use of antidepressants and individually tailor treatments to each patient (Shreders et al., 2016).

There is however likely to be an endogenous relation between cancer and depression. Cancer may not be the cause of depression, but it can be a result of depression. According to Satin et al. (2009), the cancer rate is 25 % higher in patients showing depressive symptoms, and 39 % higher in patients diagnosed with depression as such. This is supported by the study conducted by Mössinger and Kostev (2023), where they found that patients 18+ with depression have an increased risk of cancer by 10 - 39 %, depending on the type of cancer. The greatest risk of occurrence is for lung cancer, followed by cancer of the gastrointestinal-tract, breast and urinary. They also directly confirm that there is an association between depression and cancer, citing smoking as one example. Substance abuse with tobacco, which is frequent for people with psychiatric illness,

creates a dysregulated inflammatory response causing physical problems, especially lung cancer (Sethi et al., 2012). Similarly, a study by Luger et al. (2014) shows that smokers, compared to never smokers and former smokers, have twofold increased risk of developing depression, which is followed by the study from Mössinger and Kostev (2023) in the beginning of this paragraph about increased risk of cancer for depression patients.

Luber et al. (2001) also argue that compared to non-depressed patients, primary care patients aged 65 and older have significantly more radiology procedures, scans, and laboratory tests. However, Noyes Jr (1999) argues that there is an association between depression and hypochondria, where the proportion of having some kind of comorbid Axis I disorder[2] (which includes depression) among hypochondriacal and non-hypochondriacal patients is 88 % vs 51 %. Further, patients with depression have high rates of somatization and increased awareness of bodily sensations. Thus, seeing a doctor many times or seeking out lots of laboratory tests can be its consequence.

According to Spiegel and Giese-Davis (2003), many of the symptoms of cancer and many of the side effects of its treatment, such as sleep and appetite disturbances or problems with concentration, are similar to the symptoms of depression. It is a statistical problem in the measurement of overlapping symptoms and can cause a number of serious complications.

2.4 Depression as a non-cancer reason for death

It is in the interest of science and the public whether psychological factors have an effect on the mortality of people, on the development of cancer and other diseases, or both, on the mortality of people with cancer or with other diseases (Coyne et al., 2007). Satin et al. (2009) concluded in their research that depression does not predict cancer progression, but it does predict mortality, in cancer patients.

Although depressive symptoms are more common in people with advanced cancer, they can still be assumed to represent a proxy for disease severity (Massie et al., 1998). Epstein and Street Jr (2007) also say that unlike patients without depression, depressed patients tend not to maintain and use the social support available to them. Smith (2015) adds that compared to depressed but otherwise healthy patients, among patients with cancer reduced appetite and poor cogni-

[2] Axis I disorders are most commonly found in public mental health illnesses that can negatively affect a person's well-being, including for example anxiety disorders or mood disorders.

tion are helpful symptoms in diagnosing depression. Also, sense of guilt is higher among depressed but otherwise healthy patients (56.5 %) and lower among depressed cancer patients (4 %).

Finally, Akechi et al. (2004) highlight suicide as the non-cancer cause of death. They say that the risk of suicide in cancer patients is higher than in the general population. In their cohort study of terminally ill cancer patients in Japan, in which they examined the patients' suicidal thoughts and their actual interest in euthanasia, they found that at the beginning of the study, 8.6 % of the patients had suicidal thoughts, and 5 % were really interested in euthanasia. However, these numbers changed during the investigation to 38.6 % and 15.8 % of patients, respectively. They concluded that suicidal tendencies can change in all patients, and that end-of-life care focusing on psychological imbalances in dying patients could be a major form of suicide prevention.

In addition to depression (which is the main cause of suicide), hopelessness is also considered to be a factor contributing to the risk of suicide in cancer patients (Anguiano et al., 2012). Up to 39.1 % of patients even worry about being burden to others. According to Anguiano et al. (2012), even the first year after the diagnosis carries a greater risk of committing suicide. From a gender perspective, men with cancer have a greater risk of suicide than women with cancer. Suicide rates also increase as the population ages, as people aged 65 and over have a greater risk of suicide than younger people, and the rate is highest among people aged 80 and over (Anguiano et al., 2012).

Chapter 3

Data Description

The aim of this chapter is to describe the data set we will be working with, along with all the important variables that will be used in the analysis.

3.1 Dataset

The source of our data is the National Health Interview Survey (NHIS) database, managed by the National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC). NHIS is a cross-sectional household interview survey of the U.S. civilian non-institutionalized population residing within the 50 states and the District of Columbia. It is conducted throughout the year and its main objective is to monitor the health of the United States population through a broad range of health characteristics based on many demographic and socioeconomic conditions. Interviews are typically organized in respondents' homes, but the completion of the interview is often done over the phone. NCHS, CDC conducts two types of surveys, the Sample Adult Interview and the Sample Child Interview. In our model, we will work with the most current sample of adults from 2022, where the number of respondents was 27,651. As a robustness check, we will further narrow down our data to only respondents suffering from cancer, and then create two subsamples based on when they were diagnosed, so that we can examine whether there is difference between in-treatment patients and cancer survivors. We will also make a sample of non-cancer respondents for comparison. The whole sample with cancer patients has 3,430 respondents. The subsample with patients diagnosed with cancer during the last 7 years has 1,519 respondents and the subsample with patients diagnosed with cancer 8-15 years ago has 819 respondents. The non-cancer sample consists of 24,221 respondents.

3.2 Variables

Values of "refused", "don't know" and "not ascertained" were jointly coded as NA for our purposes, and thus these observations were excluded. The descriptive statistics of all variables of interest are provided in Appendix A.

3.2.1 Dependent variables

In the analysis, we use two different dependent variables. Both of them were created from the original variables available in the NHIS database and both of them are related to the respondent's depressive state. Medication against depression or anxiety (*medic*) takes the value of one if the patient takes the medication against anxiety or depression, and zero otherwise. It was created from two original binary variables *DEPMED_A* and *ANXMED_A*. Severity of depression (*depression*) is a categorical variable with 6 categories describing how often and how depressed the respondent feels. The variable takes the value 1 if the respondent feels depressed a little for a few times a year, 2 if he feels depressed a lot for a few times a year, 3 if he feels depressed monthly but just a little, 4 if he feels depressed monthly a lot, 5 if he feels depressed a little but daily, and 6 if he feels depressed daily and a lot. It was created as a combination of two original categorical variables *DEPLEVEL_A* and *DEPFREQ_A*.

3.2.2 Independent variables

Independent variables describe sociodemographic characteristics, household composition, educational level, health status, life satisfaction and response to COVID-19.

Variable *age* describes the age of a respondent at the time of the interview. If the respondent is older than 85 years, he is assigned a value of 85. There are 1,002 people older than 85 which is 3.63 % of the overall sample. As APA (2022) suggests, we expect age to have a positive effect on depression.

Female is a dummy variable determining the sex of the respondent, taking the value 1 if the adult is a female and 0 if a male. Since the average is 0.544 for the whole sample and 0.577 for the cancer sample, we can say that there are more females than males. However, since Hartung et al. (2017) and Pudrovska (2010) do not agree on who has a greater risk of cancer depression, no assumptions on the effect will be raised.

One of the ways NHIS provides geographical classification of the U.S. population is based on region classification, where states are grouped into four regions used by the U.S. Census Bureau. Since according to CDC (2023) 6 out of the 10 U.S. states with the highest prevalence of depression are located in the region South, dummy independent variable *south*, created from the original variable *REGION*, is used in the analysis. It equals 1 if the household is located in Delaware, Maryland, District of Columbia, West Virginia, Virginia, Kentucky, Tennessee, North Carolina, South Carolina, Georgia, Florida, Alabama, Mississippi, Louisiana, Oklahoma, Arkansas, or Texas, and 0 otherwise. We assume that the *south* will have a positive effect on depression.

Household composition is represented by the variable *hhsize18+*. *Hhsize18+* is a categorical variable with 3 categories representing the number of persons 18 or older in the household. It takes the value 1 if the respondent is the only adult in the household, 2 if he lives in the household with another adult, and 3 if there are a total of 3 or more adults in his household. The median number of adults in the household is 2, which indicates that respondent lives with at least one other adult, probably a partner. Since partners help each other in many things, we expect this variable to have a negative effect on patients' depressive states.

Akin-Odanye et al. (2011) state that, among other socio-demographic factors, educational level has significant negative impact on depression level of breast cancer patients. Friberg et al. (2019) further add that prostate cancer men with short education fall into a risk group prone to depression. It can be seen in the Grossman model, where education plays an important role in the production function for health, as better educated people can engender more health (Nocera and Zweifel, 1998). Since this is supported by the finding of McFarland and Wagner (2015) that a college degree serves as a so-called protection against depression, the categorical variable *educ* with 10 categories, describing the highest educational attainment of the sample adult, is used in our analysis and we expect its negative effect. It takes the value 1 if the respondent attended only grades 1-11, 2 if he also attended 12th grade, but without a diploma, 3 if he is the owner of a GED certification or its equivalent, 4 if he is a high school graduate, 5 if he attended some college, but does not have a degree, 6 if he has an occupational, technical or vocational associate degree and 7 if he has an academic associate degree, 8 if he was awarded a bachelor's degree, 9 if he has achieved a master's degree, and 10 if he has a professional school or doctoral degree.

Categorical variable *hlthstat* with 5 categories refers to the respondent's general health status. It equals 1 if his health is poor, 2 if the respondent claims

his health is fair, 3 if it is good, 4 if the health status is very good, and 5 if the health is generally excellent. We assume that the better the respondent feels, the less likely he is to suffer from depression. Thus, we are expecting negative effect of *hlthstat*. Tables in Appendix B show that there is relatively high correlation between this independent variable and *depression*. We acknowledge that this finding may raise concerns, particularly those related to the endogeneity issue. One might suggest that some symptoms of depression can affect the health status as well. However, similarly to Mayer (2023), we argue that many medical conditions with symptoms affecting the general health of a person have similar symptoms to depression, while in fact, are not depression. To support the claim that endogeneity is not necessarily a problem, we point out the following cases. Firstly, fatigue and weakness, both syndromes of depression, but also symptoms that result from anemia. Moreover, fatigue is also the leading symptom of chronic fatigue syndrome (CFS). Further, bipolar disorder and cyclothymic disorder, both similar health conditions that share some symptoms with depression, mainly the swings in mood. It is therefore important to distinguish which medical complication is causing the deterioration of the general state of health in order to avoid an incorrect attribution to depression.

One of the aspect of people's general well-being is life satisfaction. It reflects factors such as health, work, and income, and its measures have been shown to be valid and reliable. *Lifesat* with 4 categories takes the value 1 if the person is very dissatisfied with his life, 2 if he is dissatisfied, 3 if he is just satisfied, and 4 if he is very satisfied with his life. We can assume that this variable will have a negative effect. One might also find concerning the ability of *lifesat* to explain big proportion of variance in the dependent variable *depression* in the whole sample. Due to the fact that such independent variables can be considered as an alternatives for dependent variable, there might occur bias in interpretation. However, the results for the cancer time-subsamples refuse those concerns as *lifesat* there is statistically insignificant. Thus, even though we admit that the correlation matrix is pointing out potential problems, we conclude that they are not major threats to our analysis, as we are interested in the explanation of differences between whole and cancer samples.

Adults who were ever tested positive for COVID-19 were asked how would they describe their coronavirus symptoms when they were at their worst. Answers are described by categorical variable *symptcvd* with 4 categories, which equals 0 if the respondent had no symptoms, 1 if he had mild symptoms, 2 if the symptoms were moderate, and 3 if he suffered from severe covid symptoms. The median in our

analysis is 2. In their research on cancer patients during the COVID-19 pandemic, Tolia et al. (2023) came to the conclusion that the number of those suffering from anxiety or depression conditions was significantly small. In addition, they emphasized the knowledge of patients being far more concerned about their cancer treatment than about COVID-19. Xiong et al. (2020) say that for the general population the COVID-19 pandemic is associated with highly significant levels of psychological hazardous effects on mental health. Thus, we assume positive effect of coronavirus symptoms on depressive states for the overall sample, but we expect its lower significant level for sample with cancer patients.

Dummy independent variable *cancer* is added to the regression in the whole sample to determine the effect of having cancer diagnosis on the depressive states of patients. It equals 1 if the respondent has ever been told he had a cancer, and 0 otherwise. According to our assumptions we expect positive effect of this variable.

Two dummy variables representing time intervals specifying when patients were diagnosed with cancer are added to the regression in the cancer sample, and then also to the regression in the whole sample to replace the *cancer*. *time_7* equals 1 if the patient was diagnosed with cancer during the last 7 years, and *time_15* equals 1 if the diagnosis was told him in the period of 8-15 years ago. Since 8-15 years is a relatively long time ago, we can expect a negative effect of the variable *time_15* on the occurrence of depression, as it is assumed that people are already cured, which led them to reevaluate their life values. Conversely, people with a diagnosis no older than 7 years are assumed to be still in treatment, so the positive effect of *time_7* is expected.

3.3 Preliminary analysis and adjustments of the original dataset

All independent variables are checked for atypical values. The interquartile range (IQR) method is used to determine upper and lower bounds for the data, especially for *age*. To avoid potential bias, as many as 64 observations were excluded as outliers from the original dataset of 27,651 observations. To ensure that our analysis is robust, based on IQR as many as 123 observations were excluded as outliers from the dataset of 3,430 observations of only cancer patients, and 57 observations were excluded as outliers from the dataset of 24,221 observations of non-cancer respondents. Further, as many as 38 observations were excluded from the subsample of 1,519 cancer patients with diagnosis no older than 7 years, and 36

observations were excluded from the subsample of 819 cancer patients diagnosed 8-15 years ago.

Independent variables are then checked for multicolinearity using a correlation matrix (see Table B.1, Table B.2, Table B.3, Table B.4 and Table B.5 in appendix). As boundaries are used correlations of 0.5 or -0.5. Any variables with a correlation greater than 0.5 in absolute values would be excluded from the analysis. But since the only two variables with such correlation are *cancer* and *time_7* in the sample of all respondents (the correlation between them is 0.641), and these two factors are never used in the same analysis, there is no need to remove any of our variables of interest.

Chapter 4

Methodology

This chapter presents the two types of models that were chosen to analyze the data set. We have decided to use a logit model for the dependent variable *medic*, and a multinomial logit model for the dependent variable *depression*. All the data analysis were completed in R Studio.

4.1 Logit model

Logit model is a statistical model used to model binary dependent variables, whose range of values is restricted only to the values 0 or 1. Applying this model will allow us to overcome the limitations of the linear probability model, such that the fitted probabilities are not restricted (meaning they can be less than zero or greater than one) or that the partial effect of any explanatory variable is constant. Thus, we are interested primarily in the response probability:

$$P(y = 1 \mid x_1, x_2, \ldots, x_k) = G(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k) \tag{4.1.1}$$

where y is the dependent variable, x_1, x_2, ..., x_k are the independent variables, G is a non-linear logistic function

$$G(z) = \frac{exp(z)}{1 + exp(z)} \tag{4.1.2}$$

ensuring that the estimated response probabilities are strictly between zero and one, β_0 is the intercept and β_1, β_2, ..., β_k are coefficients. In order to reach the equation for logistic regression, the needed odds ratio is $\frac{P(y)}{1-P(y)}$. The natural

logarithm of this expression is given as:

$$ln\left[\frac{P(y)}{1-P(y)}\right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k \tag{4.1.3}$$

where after applying exponential function on both sides, we get:

$$\frac{P(y)}{1-P(y)} = e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_k x_k} \tag{4.1.4}$$

and finally, after rearranging the equation, we reach our logistic function which is a cumulative distribution function for the standard logistic random variable such that:

$$P(y) = e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_k x_k} - P(y)e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_k x_k} \tag{4.1.5}$$

$$P(y) = \frac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_k x_k}}{1+e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_k x_k}} \tag{4.1.6}$$

Due to its non-linear nature, maximum likelihood estimation (MLE) is used for the logit model estimation. To obtain MLE of β, we need to maximize the log-likelihood function in the sample. Under very general conditions, the MLE is consistent, asymptotically normal and asymptotically efficient.

Wooldridge (2012) further suggests that the most difficult aspect of logit model is presenting and interpreting results. Usually, we are interested in the effect of independent variables on the probability that y happens, that is, that $P(y = 1 \mid \mathbf{x}) = G(\beta_0 + \mathbf{x}\boldsymbol{\beta})$. This effect can be obtained as:

$$\frac{\partial P(y = 1 \mid \mathbf{x})}{\partial x_j} = g(\beta_0 + \mathbf{x}\boldsymbol{\beta})\beta_j \tag{4.1.7}$$

where $g(z) = \frac{\partial G(z)}{\partial z}(z)$ and β_j are coefficients denoting the sign of the marginal effect of each x_j on the response probability. By computing this effect for each observation and then averaging it, we can obtain average marginal effect:

$$AME = \frac{1}{N}\sum_{1}^{N} g(\beta_0 + \mathbf{x}\boldsymbol{\beta})\beta_j \tag{4.1.8}$$

To measure goodness-of-fit, we cannot simply use R^2. One possibility is a pseudo R^2 called McFadden's R squared, which is based on the log-likelihood defined as:

$$R^2_{McFadden} = 1 - \frac{log(\mathcal{L}_{full})}{log(\mathcal{L}_{null})} \tag{4.1.9}$$

where $\mathcal{L}_{full}$ denotes the value of log-likelihood of the estimated model, and $\mathcal{L}_{null}$ denotes the value of log-likelihood of a model with an intercept only.

The easiest way how to test the significance of variables is by using the likelihood ratio (LR) test. The LR test is based on twice the difference in the log-likelihoods for the unrestricted and restricted models:

$$LR = 2(\mathcal{L}_{ur} - \mathcal{L}_r) \overset{a}{\sim} \chi^2_q \tag{4.1.10}$$

where $\mathcal{L}_{ur}$ is the log-likelihood value of the unrestricted model, $\mathcal{L}_r$ is the log-likelihood value of the restricted model and q is the number of restrictions.

4.2 Multinomial logit model

Multinomial logit model is a procedure used to generalize the logistic regression to more than two possible discrete outcomes. In such case, the dependent variable is categorical (sometimes also called nominal) with more than two categories that cannot be ordered in any meaningful way.

We denote y as a dependent variable taking on the values {0, 1, ..., J} for J a positive integer and $\mathbf{x}$ as a set of conditioning variables. As in the logit model, we are interested primarily in the response probability $P(y = j \mid \mathbf{x})$ for j = 0, 1, 2, ..., J. Once we know the probabilities for j = 1, ..., J, $P(y = 0 \mid \mathbf{x})$ is determined due to the rule that probabilities must sum to unity. After letting $\mathbf{x}$ be a $1 \times K$ vector, the multinomial logit model response probability is given by:

$$P(y = j \mid \mathbf{x}) = \frac{exp(\mathbf{x}\boldsymbol{\beta}_j)}{1 + \sum_{h=1}^{J} exp(\mathbf{x}\boldsymbol{\beta}_h)} \tag{4.2.1}$$

for j = 1, ..., J. As the probabilities must sum to unity,

$$P(y = 0 \mid \mathbf{x}) = \frac{1}{1 + \sum_{h=1}^{J} exp(\mathbf{x}\boldsymbol{\beta}_h)} \tag{4.2.2}$$

$\boldsymbol{\beta}_j$ is interpreted such that by taking the ratio between (4.2.1) and (4.2.2) for j = 1, 2, ..., J, we get:

$$\frac{P(y = j \mid \mathbf{x})}{P(y = 0 \mid \mathbf{x})} = exp(\mathbf{x}\boldsymbol{\beta}_j) \tag{4.2.3}$$

where the change in $\frac{P(y=j|\mathbf{x})}{P(y=0|\mathbf{x})}$ is $\beta_{jk} exp(\mathbf{x}\boldsymbol{\beta}_j)\Delta x_k$ for roughly continuous x_k. After

applying logarithmic function on both sides, we get:

$$log\left(\frac{P(y=j \mid \mathbf{x})}{P(y=0 \mid \mathbf{x})}\right) = \mathbf{x}\boldsymbol{\beta}_j \tag{4.2.4}$$

and this can generally extend j and h:

$$log\left(\frac{P(y=j \mid \mathbf{x})}{P(y=h \mid \mathbf{x})}\right) = \mathbf{x}(\boldsymbol{\beta}_j - \boldsymbol{\beta}_h) \tag{4.2.5}$$

Wooldridge (2010) further mentions another useful fact about multinomial logit model, such as that since $P(y = j \text{ or } y = h \mid \mathbf{x}) = P(y = j \mid \mathbf{x}) + P(y = h \mid \mathbf{x})$, it holds that:

$$P(y=j \mid y=j \text{ or } y=h, \mathbf{x}) = \frac{P(y=j \mid \mathbf{x})}{P(y=j \mid \mathbf{x}) + P(y=h \mid \mathbf{x})} = \Lambda\left[\mathbf{x}(\boldsymbol{\beta}_j - \boldsymbol{\beta}_h)\right] \tag{4.2.6}$$

where $\Lambda(\cdot)$ is the logistic function. Thus, the probability that the outcome j follows a standard logit model with parameter vector $\boldsymbol{\beta}_j - \boldsymbol{\beta}_h$ depends on the choice being either j or h.

Additionally, same as for logit model, maximum likelihood estimation (MLE) is used for the estimation of multinomial logit model. Again we need the conditional log-likelihood, which for each i can be written as:

$$\ell_i(\boldsymbol{\beta}) = \sum_{j=0}^{J} 1\left[y_i = j\right] log\left[P(y=j \mid \mathbf{x}_i)\right] \tag{4.2.7}$$

and traditionally to obtain MLE of $\boldsymbol{\beta}$, we need to maximize the $\sum_{i=1}^{N} \ell_i(\boldsymbol{\beta})$.

Because associating $\boldsymbol{\beta}_j$ with jth outcome is inaccurate, the coefficients of the multinomial logit model are difficult to interpret (Greene, 2018). To obtain marginal effect of the characteristics on the probability, we have to differentiate (4.2.1) as:

$$\boldsymbol{\delta}_j = \frac{\partial P(y=j \mid \mathbf{x})}{\partial \mathbf{x}} = P(y=j \mid \mathbf{x})\left[\boldsymbol{\beta}_j - \sum_{k=0}^{J} P(y=k \mid \mathbf{x})\boldsymbol{\beta}_k\right] = P(y=j \mid \mathbf{x})\left[\boldsymbol{\beta}_j - \bar{\boldsymbol{\beta}}\right] \tag{4.2.8}$$

There, every subvector of $\boldsymbol{\beta}$ gets in every marginal effect through the probability and through the weighted average occuring in $\boldsymbol{\delta}_j$. Unlike the coefficients, marginal effects remain the same regardless of the reference category.

Goodness-of-fit will be again measured using pseudo R^2 called McFadden's R^2 defined as:

$$R^2_{McFadden} = 1 - \frac{log(\mathcal{L}_{full})}{log(\mathcal{L}_{null})} \tag{4.2.9}$$

where $\mathcal{L}_{full}$ is the value of log-likelihood of the estimated model, and $\mathcal{L}_{null}$ is the value of log-likelihood of a model with no predictors.

4.3 Model Evaluation

To make sure validity of our models, and to see how well they fit the data, the following tests will be run to answer these questions. Results for all logit and multinomial logit models, indicating no problem in our analysis, are presented in Appendix D.

4.3.1 Goodness of Fit

Likelihood Ratio Test

In order to decide which model offers a significantly better fit, we perform likelihood ratio tests. The LR test works on the principle of comparing two regression models - the nested model (Model 1) and the overall model (Model 2). Under the null hypothesis, the full model and the nested model fit the data equally well, and thus the nested model is preferred as additional predictors do not represent a significant improvement.

Hosmer-Lemeshow Test

To determine the goodness of fit of all our models, the Hosmer-Lemeshow tests are run. The test works on the principle of comparing observed with expected frequencies of the outcome using chi-squared distributed test statistic. The evidence of good fit is a result of non-significant p-value, which means that the observed and expected frequencies do not differ.

4.3.2 Multicollinearity

Both logistic regression and multinomial logistic regression require no multicollinearity among the independent variables. For this, variance inflation factor (VIF) is used. VIF helps us understand the amount of the variance of the estimated coefficient that is inflated due to collinearity. It is measured by taking a

variable and regressing it against every other variable. A value of VIF which is greater than 5 indicates that there is a severe correlation among predictors.

Table 4.1: Model evaluation

Test	**H0**	**Short description**
Likelihood Ratio test	The full model and the nested model fit the data equally well. The nested model should be used.	To decide which model offers a significantly better fit. It compares the goodness of fit of the nested model and the full model.
Hosmer-Lemeshow test	No evidence that the observed and expected frequencies differ, thus there is evidence of good fit.	To determine the goodness of fit of the model. It compares observed with expected frequencies of the outcome and uses a chi-squared distributed test statistic.
Test	**Acceptable level**	**Short description**
Multicollinearity test	VIF < 5	To detect the amount of multicollinearity among the independent variables in a regression.

Chapter 5

Results

In this chapter, the results of our analyzes will be presented to describe the influence of various factors on the depressive states of cancer patients. The final logit model, described in section 4.1, was selected based on LR test (Appendix D).

To guarantee the robustness of the results, we will do our regression analyzes separately for different subsamples. The whole sample of respondents, the cancer subsamples divided based on time of diagnosis, and the sample of non-cancer respondents.

5.1 Logistic Regression

Due to non-linear nature of logit models, the results cannot be presented directly. Thus, we use average marginal effects (AME), all summarised in one table in Appendix C, to interpret the obtained coefficients from our analyzes.

5.1.1 Results for cancer patients

Table 5.1 presents results of our analysis for cancer patients. The estimated coefficients and standard errors, together with AME, are included.

Both age and gender are highly significant at 1 % level, but their signs are in opposite direction. The results suggest that with increasing age by one additional year, among cancer patients the probability of taking the medication for depression decreases by 0.5 %. Not only is the effect small despite its significance, but also the result is not in line with our assumption that age increases depression. A potential explanation for this result is that if people live to a later age despite a cancer diagnosis, they are likely to change their priorities and the reflection on

their life. Furthermore, being female has large impact on the depression as it increases the probability of taking the medication by 13.2 %.

According to our results, the variable *hhsize18+* is significant at 10 % level. Thus, for people with cancer living in the household with at least one other adult the probability of depression medication decreases by 3.9 %, which is in line with our assumption.

Table 5.1: Logistic Regression - the cancer sample

	Dependent variable: medic	
	Logit	AME
age	-0.029^{***}	-0.005
	(0.007)	
female	0.814^{***}	0.132
	(0.183)	
hhsize18+	-0.241^{*}	-0.039
	(0.134)	
educ	0.115^{***}	0.019
	(0.037)	
hlthstat	-0.448^{***}	-0.072
	(0.083)	
lifesat	-0.250^{*}	-0.040
	(0.130)	
symptcvd	0.037	0.006
	(0.090)	
time_7	-0.182	-0.029
	(0.185)	
time_15	-0.502^{**}	-0.081
	(0.229)	
Constant	2.241^{***}	-
	(0.751)	
Observations	940	
Pseudo R^2	0.107	
Log Likelihood	-465.215	
Akaike Inf. Crit.	950.430	
Note:	*p<0.1; **p<0.05; ***p<0.01	

The effect of *educ*, an additional educational attainment, is highly statistically significant at 1 % level, but positive. Therefore, our hypothesis that the more educated an individual is, the less likely he is to suffer from depression, cannot be in the cancer sample supported. A potential explanation is that more educated

people are subject to higher level of stress, possibly at work, which leads to serious psychological problems like burnout symptom which is depressive in nature (Bianchi et al., 2018).

The variable *hlthstat* describing patient's general health status is highly significant at 1 % level and negative, which supports our assumption that the better the patient feels, the less likely he is to suffer from depression. The same can be inferred about the effect of life satisfaction. The effect of *lifesat* is statistically significant at 10 % significance level and negative, confirming that people's general well-being decreases the probability of depression.

From the regression results we see can that the effect of *symptcvd* (coronavirus symptoms) is statistically insignificant at any reasonable level of significance. Thus, our assumption about the influence of COVID-19 on depression among cancer patients can be supported.

Compared to insignificant *time_7*, dummy variable *time_15* is statistically significant at 5 % level. Its effect tells us that if the patient was diagnosed with cancer 8-15 years ago, the probability that he will take the depression medication decreases by 8.1 %. Scott (2016) explains that after surviving a cancer diagnosis, people start to think of their day-to-day life differently and they appreciate the simple things much more. This, and our result from the analysis, support our hypothesis about reevaluating life values among cancer survivors.

5.1.2 The whole sample regression results

This subsection describes how the results change after running the regression for the whole sample of respondents. Table 5.2 presents the results when once a dummy variable *cancer*, indicating if the respondent has ever been told he had cancer, is included, and then when *cancer* is replaced by two dummies *time_7* and *time_15* describing time intervals when patients were diagnosed.

For the whole sample with dummy *cancer* it can be seen that, apart from minor exceptions in significance and sign of the effects, there have been almost no changes compared to the sample of cancer patients.

The effect of *age*, now significant at 5 % level, is a bit smaller, resulting in decrease in probability of taking the medication by 0.1 % with one additional year of age.

Household composition is now highly statistically significant, but its impact has become smaller in that living with one additional adult decreases the probability of taking the medication for depression by only 1.6 % in the whole sample.

Table 5.2: Logistic Regression - the whole sample

	Dependent variable: medic			
	with *cancer*		with time intervals	
	Logit	AME	Logit	AME
age	-0.004^{**}	-0.001	-0.003^{*}	-0.000
	(0.002)		(0.002)	
female	0.842^{***}	0.116	0.845^{***}	0.116
	(0.058)		(0.058)	
hhsize18+	-0.114^{***}	-0.016	-0.116^{***}	-0.016
	(0.041)		(0.041)	
educ	0.089^{***}	0.012	0.090^{***}	0.012
	(0.012)		(0.012)	
hlthstat	-0.462^{***}	-0.063	-0.465^{***}	-0.064
	(0.030)		(0.030)	
lifesat	-0.415^{***}	-0.057	-0.416^{***}	-0.057
	(0.048)		(0.048)	
symptcvd	0.137^{***}	0.019	0.138^{***}	0.019
	(0.031)		(0.031)	
cancer	0.196^{**}	0.027	-	-
	(0.089)			
time_7	-	-	0.214^{*}	0.029
			(0.122)	
time_15	-	-	-0.142	-0.019
			(0.178)	
Constant	0.576^{***}	-	0.556^{***}	-
	(0.213)		(0.213)	
Observations	9,931		9,931	
Pseudo R^2	0.085		0.085	
Log Likelihood	$-4{,}337.121$		$-4{,}337.603$	
Akaike Inf. Crit.	8,692.243		8,695.207	
Note:			*p<0.1; **p<0.05; ***p<0.01	

Both *lifesat* and *symptcvd* have gained significance and are now highly significant at 1 % level. Moreover, the effects of both variables are as we expected them to be. Negative effect of *lifesat* tells us that the happier a person is, the less he suffers from depression. Further, as the severity of the coronavirus increases, the probability of medication increases by 1.9 %. Here, from the highly significant effect at the 1 % level for all respondents and from the statistically insignificant effect for respondents suffering from cancer, we can see that our assumption that COVID-19 increases the probability of depression was correct. The global pan-

demic has had a significant impact on mental health for the general population, but not for the cancer population.

The fact that the patient was diagnosed with cancer during his lifetime is statically significant at the 5 % level. Having cancer diagnosis increases the depression medication by 2.7 %, and since the effect is positive, it supports our hypothesis about cancer patients being prone to depression.

All other variables regarding gender, educational attainment or general health status remain statistically significant determining the probability of taking the medication against depression in the same direction as in the cancer sample.

From Table 5.2 we can further see that the results for the whole population, when cancer is replaced by dummies representing the time when cancer was first diagnosed, are essentially identical to the results for the first model with *cancer*.

Probably the biggest change occurs with *age*, which lost significance in the overall sample, and is now significant at the 10 % significance level only. In addition, its impact on the probability of taking the medication for depression approaches zero.

New added dummy variable *time_7* is significant at 10 % significance level, and its effect is positive. This supports our hypothesis that people with a diagnosis no older than 7 years are assumed to be still in treatment. Thus, if a patient has a relatively recent diagnosis of cancer, his probability of taking medication for depression increases by 2.9 %. On the other hand, the dummy variable *time_15* is insignificant at any reasonable level. Therefore, the data do not provide evidence of the assumed negative effect due to reevaluation of life values.

All other variables kept exactly the same significance as in the previous model, and their effects almost did not change either.

5.1.3 Cancer subsamples results

This subsection presents results of two regressions for cancer patients based on how long it is since their first cancer diagnosis. The regression results for both groups of patients diagnosed (1) during the last 7 years and (2) 8-15 years ago are summarised in Table 5.3.

As we can see, the number of significant variables is smaller for both subsamples than in the whole sample and whole cancer sample.

In both cases, age together with gender remain significant at 1 % level, and their signs do not change either. One additional year of age is associated with 0.5 % decrease in probability of medication for the first subsample, and with 0.4 %

decrease for the second subsample. Similarly, being female increases the probability of depression medication by 15 % for people diagnosed during the last 7 years, and by 8.4 % for people diagnosed 8-15 years ago.

Hhsize18+ is still, same as for all cancer patients, significant at 5 % level and decreases the probability for depression by 8.1 % for people with a diagnosis no more than 7 years old. However, for patients suffering with cancer for more than 7 years it has become insignificant. This is plausible, considering that if we assume that these people are already cured, their dependence on others is no longer so significant and they are now able to take care of themselves. Thanks to the gathered self-management, such survivors are able to keep their psychological comfort in the first place without the help of others.

Table 5.3: Logistic Regression - time of cancer diagnosis

	Dependent variable: medic			
	(1) t $\leq$ 7		(2) 7 $<$ t $\leq$ 15	
	Logit	AME	Logit	AME
age	-0.032^{***}	-0.005	-0.027^{***}	-0.004
	(0.010)		(0.009)	
female	0.913^{***}	0.150	0.624^{***}	0.084
	(0.268)		(0.209)	
hhsize18+	-0.493^{**}	-0.081	0.121	0.016
	(0.197)		(0.162)	
educ	0.078	0.013	0.084^{*}	0.011
	(0.054)		(0.045)	
hlthstat	-0.519^{***}	-0.085	-0.500^{***}	-0.067
	(0.128)		(0.100)	
lifesat	-0.081	-0.013	-0.180	-0.024
	(0.198)		(0.160)	
symptcvd	-0.021	-0.003	-	-
	(0.134)			
Constant	2.601^{**}	-	1.381	-
	(1.062)		(0.903)	
Observations	416		764	
Pseudo R^2	0.110		0.079	
Log Likelihood	-208.107		-328.155	
Akaike Inf. Crit.	432.213		670.309	
Note:			*p<0.1; **p<0.05; ***p<0.01	

Educational level remains significant only for the subsample of patients diagnosed with cancer 8-15 years ago, and still its effect does not support our hypothesis that it decreases the occurrence of depression. Although it is inconsistent

with our assumption, the fact that higher education increases the depression is in line with the study by Maneeton et al. (2012). They found that cancer patients educated for more than 13 years are at higher risk of the prevalence of depressive disorder.

Hlthstat remains highly significant for both cases, leading to decrease in the probability of taking the medication by 8.5 % and 6.7 %, respectively. On the contrary, *lifesat* has lost its significance completely. This removes the doubts about endogeneity of this variable. It could be attributed to the fact that cancer overweights the life satisfaction as people are mainly focusing on treating cancer and stuff related to diagnosis, and they do not really care how satisfied they are in everyday life.

The variable determining the severity of the coronavirus symptoms is excluded from the sample for patients diagnosed 8-15 years ago, as covid occurred only 5 years ago. As for the whole cancer sample, this variable is insignificant for people who we assume are still being treated for cancer.

5.1.4 Results for non-cancer patients

For comparison, this section presents results for a sample of respondents, where people who have ever been diagnosed with cancer are excluded. Estimated coefficients together with AME are displayed in Table 5.4.

The biggest change that emerges is the significance of the effect of age on the probability of taking the medication for depression. For people without cancer diagnosis every additional year of life is statistically insignificant at any reasonable level, compared to all cancer samples, where this effect was significant at the 1 % significance level. For people without cancer, we do not expect a reevaluation of life values because they did not have to overcome a dangerous disease, so the insignificant effect does not surprise us.

Further, the effect of another adult in the respondent's household gained significance compared to cancer patients. *Hhsize18+* is now significant at the 5 % level and is associated with a 1.4 % decrease in the probability of depression medication. Life satisfaction also gained significance. With the effect significant at the 1 % level, it is still true that the happier the respondent is, the less he suffers from depression.

For non-cancer patients, coronavirus symptoms are significant at the 1 % level. The effect tells us that as the progress of the coronavirus disease worsens, the probability of depression for people without cancer increases by 2.1 %. The results

suggest that people without cancer are not preoccupied with other worries like cancer, and are more affected by the fear of the pandemic.

As in all other samples, women are more prone to depression also in the non-cancer sample. Likewise, the variables regarding education and health status retained their significance and effect as for cancer patients and the overall sample.

Table 5.4: Logistic Regression - the non-cancer sample

	Dependent variable: medic	
	Logit	AME
age	−0.002	−0.000
	(0.002)	
female	0.835***	0.112
	(0.062)	
hhsize18+	−0.106**	−0.014
	(0.043)	
educ	0.085***	0.011
	(0.013)	
hlthstat	−0.460***	−0.061
	(0.033)	
lifesat	−0.447***	−0.060
	(0.052)	
symptcvd	0.159***	0.021
	(0.034)	
Constant	0.568**	-
	(0.228)	
Observations	8,938	
Pseudo R^2	0.082	
Log Likelihood	−3,825.883	
Akaike Inf. Crit.	7,667.765	
Note:	*p<0.1; **p<0.05; ***p<0.01	

5.2 Multinomial Logistic Regression

The results of multinomial logistic regressions are shown in Table 5.5, Table 5.6, Table 5.7, Table 5.8, Table 5.9 and Table 5.10. Due to inaccurate association of $\boldsymbol{\beta}_j$ with jth outcome, AME will be used for interpretation of the obtained effects. Note that the dependent variable *medic* in the logistic regression is much more objective as it denotes if the person takes the medication for depression or anxiety,

or not. On the other hand, our response variable *depression* in the multinomial regression has rather self-assessment nature.

5.2.1 Results for cancer patients

Table 5.5 gives us outcomes from the regression in sample with cancer patients.

The results suggest that among cancer patients, age decreases the probability of having severe daily or monthly depression. Moreover, the effects are highly significant, suggesting that one additional year of age reduces the probability of severe daily depression by 1.8 %, and the probability of severe monthly depression by 0.7 %. There is also highly significant increase in the probability of mild daily depression, where another year of age increases the chance of occurrence by 0.7 %. Nevertheless, we can conclude that with increasing age, people with cancer are more resistant to external negative exposures, and thus are less likely to develop depression.

Table 5.5: Multinomial Logistic Regression - the cancer sample

	Dependent variable: depression					
	(1) A little/year	(2) A lot/year	(3) A little/month	(4) A lot/month	(5) A little/day	(6) A lot/day
age	0.012**	0.006*	0.000	−0.007***	0.007***	−0.018***
	(0.006)	(0.003)	(0.004)	(0.002)	(0.003)	(0.004)
agesq	0.000	0.000*	0.000	0.000**	0.000**	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
female	−0.020**	−0.026***	0.015***	0.039***	−0.007***	0.000
	(0.009)	(0.005)	(0.003)	(0.011)	(0.002)	(0.003)
south	0.036***	−0.025***	0.018***	−0.003**	0.002*	−0.028***
	(0.007)	(0.005)	(0.007)	(0.001)	(0.001)	(0.005)
hhsize18+	0.012	−0.020***	−0.029	0.029***	0.015***	−0.007
	(0.013)	(0.006)	(0.018)	(0.008)	(0.004)	(0.004)
hlthstat	0.027	−0.006	0.030*	−0.001	−0.021**	−0.029***
	(0.022)	(0.013)	(0.016)	(0.008)	(0.008)	(0.011)
lifesat	0.091***	0.035***	−0.033***	−0.005*	−0.002	−0.085***
	(0.013)	(0.008)	(0.012)	(0.003)	(0.004)	(0.013)
symptcvd	−0.061***	0.020	0.020	−0.001	0.050***	−0.029*
	(0.023)	(0.017)	(0.020)	(0.005)	(0.011)	(0.016)
time_7	−0.064***	0.005	0.106***	−0.040***	−0.006***	−0.001
	(0.017)	(0.003)	(0.020)	(0.011)	(0.002)	(0.004)
time_15	−0.001	−0.036***	0.081***	0.006***	−0.054***	0.003
	(0.016)	(0.008)	(0.014)	(0.002)	(0.015)	(0.004)
Observations	260	260	260	260	260	260
Pseudo R^2	0.185	0.185	0.185	0.185	0.185	0.185
Akaike Inf. Crit.	602.264	602.264	602.264	602.264	602.264	602.264
Note:						*p<0.1; **p<0.05; ***p<0.01

Being a female also has a large impact on the depressive states of people with cancer. It is statistically significant at 1 %, and the effects tell us that being a

woman is associated with a huge increase in the probability of strong monthly depression by 3.9 %, and mild monthly depression by 1.5 %. On the other hand, it decreases the probability of the incidence of mild daily depression, strong yearly depression and mild yearly depression by 0.7 %, 2.6 % and 2 %, respectively. These results could indicate that women are generally less prone to depression, except for the monthly occurrence, which might be connected to premenstrual syndrome (PMS) causing substantial mood instability.

Further, from Table 5.5 it can be seen that living in the Southern States is for patients suffering from cancer also highly significant, but the effect varies. It decreases the probability of severe daily depression by 2.8 %, and of severe depression a few times a year by 2.5 %, but it increases the probability of mild monthly depression occurrence by 1.8 % and the occurrence of mild annual depression by 3.6 %.

Patient's general health status is significant for the probability of having any form of daily depression. Both the large impact and negative sign were expected. With improving patient's health status, the probability that the patient will have strong daily depression decreases by 2.9 %, and the probability that he will have mild daily depression decreases by 2.1 %.

The fact that *lifesat* is almost fully significant and mostly negative supports our hypothesis that this variable reduces the risk of depression.

As expected, coronavirus symptoms seem to be insignificant when reporting depression status among cancer patients, which corresponds to the study by Tolia et al. (2023) who found that patients do not worry about coronavirus as much as their own cancer treatment and is in line with our previous results.

5.2.2 The whole sample regression results

This subsection describes how the results change after running the regression for the whole sample of respondents. Once with dummy variable *cancer* (Table 5.6) and once with dummies *time_7* and *time_15* (Table 5.7).

From Table 5.6 it can be seen that age has lost its significance in the whole sample, even though its signs remained practically unchanged.

Female is now highly significant for all alternatives and mostly positive, which in particular leads to large increase in the probability of severe daily depression by 1.8 %, in the probability of severe monthly depression by 2.2 %, in the probability of mild monthly depression by 3.5 % and of severe depression a few times a year by 0.8 %. It is interesting to mention that the probability that the patient will

suffer from mild depression only a few times a year decreases with female gender by 8.1 %. Compared to the cancer sample, we can see that the effect is highly statistically significant for all alternatives. However, it is important to mention that in both cases the female gender has a significant positive effect on the occurrence of depression in monthly intervals. As we already pointed out, it is likely to be explained by PMS, which affects up to 75 % of menstruating women. Although the symptoms vary and are mostly mild, in some cases they can develop into severe symptoms such as depression, anxiety or anger. According to Leonard (2020), in such a case it is premenstrual dysphoric disorder (PMDD), which affects 3-8 % of people with a menstrual cycle. Such people then attempt suicide in 15 % of cases during their lifetime.

Variables *hhsize18+*, *hlthlstat* and *lifesat* have gained their significance and all of them still have the expected negative effect according to our assumptions. The only exception for all three variables is mild annual depression. In all cases, the effect is statistically significant at the 5 % significance level and positive. Living in a household with another adult increases the probability of this depression by 4.1 %, better health status increases the probability by 7.5 %, and life satisfaction increases it by an enormous 14.4 %.

Table 5.6: Multinomial Logistic Regression - the whole sample with *cancer*

	Dependent variable: depression					
	(1) A little/year	(2) A lot/year	(3) A little/month	(4) A lot/month	(5) A little/day	(6) A lot/day
age	0.004^{**}	0.002^{*}	-0.005^{***}	-0.001	0.001	-0.001
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
agesq	0.000	0.000^{**}	0.000^{***}	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
female	-0.081^{***}	0.008^{**}	0.035^{***}	0.022^{***}	-0.002^{***}	0.018^{***}
	0.008	(0.003)	(0.012)	(0.002)	(0.001)	(0.002)
south	-0.003	-0.007^{***}	0.014	0.009^{***}	-0.003^{***}	-0.009^{***}
	(0.008)	(0.002)	(0.013)	(0.001)	(0.001)	(0.002)
hhsize18+	0.041^{***}	-0.013^{*}	0.005	-0.015^{***}	-0.007^{***}	-0.011^{*}
	(0.010)	(0.007)	(0.008)	(0.005)	(0.002)	(0.006)
hlthstat	0.075^{***}	-0.009^{*}	0.000	-0.014^{***}	-0.014^{***}	-0.037^{***}
	(0.007)	(0.005)	(0.006)	(0.003)	(0.003)	(0.004)
lifesat	0.144^{***}	0.003	-0.028^{***}	-0.022^{***}	-0.011^{***}	-0.086^{***}
	(0.009)	(0.007)	(0.010)	(0.004)	(0.002)	(0.007)
symptcvd	-0.021^{**}	0.017^{***}	-0.013^{*}	0.003	0.000	0.014^{***}
	(0.010)	(0.006)	(0.007)	(0.004)	(0.003)	(0.005)
cancer	-0.044^{***}	0.006^{***}	0.021^{***}	0.011^{***}	0.000	0.006^{***}
	0.001	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)
Observations	2,761	2,761	2,761	2,761	2,761	2,761
Pseudo R^2	0.108	0.108	0.108	0.108	0.108	0.108
Akaike Inf. Crit.	5,632.906	5,632.906	5,632.906	5,632.906	5,632.906	5,632.906

Note: *p<0.1; **p<0.05; ***p<0.01

As we assumed, coronavirus symptoms are more significant for the whole sample compared to the cancer sample. Further, its effect is mostly positive which suggests that *symptcvd* is associated with the increase in the probability of occurrence of strong daily depression by 1.4 %, and in the probability of strong annual depression by 1.7 %.

Having cancer diagnosis is significant for almost all categories of depression, and the effects are large and positive, as we expected.

Table 5.7 shows us that when the variable *cancer* is replaced by *time_7* and *time_15*, there is almost no change in the effects.

Both dummy variables *time_7* and *time_15*, indicating time of respondent's cancer diagnosis, are highly statistically significant at 1 % level. Moreover, their effect is as we expected. *time_7* is mostly positive, which corresponds to our prediction about people who are still in treatment. Conversely, the effect of *time_15* is once positive and once negative, which could indicate that people who are cured have reevaluated their life priorities.

All other variables remained significant as before, and their effects are almost identical compared to the first model in the whole sample.

Table 5.7: Multinomial Logistic Regression - the whole sample with time intervals

	Dependent variable: depression					
	(1) A little/year	(2) A lot/year	(3) A little/month	(4) A lot/month	(5) A little/day	(6) A lot/day
age	0.004**	0.002*	−0.005***	−0.001	0.001	−0.001
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
agesq	0.000	0.000**	0.000***	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
female	−0.082***	0.008**	0.036***	0.022***	−0.002***	0.018***
	(0.008)	(0.003)	(0.012)	(0.002)	(0.001)	(0.002)
south	−0.004	−0.008***	0.015	0.009***	−0.003***	−0.009***
	(0.008)	(0.002)	(0.013)	(0.001)	(0.000)	(0.002)
hhsize18+	0.042***	−0.013*	0.005	−0.015***	−0.008***	−0.011*
	(0.010)	(0.007)	(0.008)	(0.005)	(0.002)	(0.006)
hlthstat	0.075***	−0.009*	0.000	−0.015***	−0.014***	−0.037***
	(0.007)	(0.005)	(0.006)	(0.003)	(0.003)	(0.004)
lifesat	0.145***	0.003	−0.028***	−0.022***	−0.011***	−0.086***
	(0.009)	(0.007)	(0.010)	(0.004)	(0.002)	(0.007)
symptcvd	−0.021**	0.017***	−0.014*	0.003	0.000	0.014***
	(0.010)	(0.006)	(0.007)	(0.004)	(0.003)	(0.005)
time_7	−0.079***	0.016***	0.069***	−0.022***	0.004***	0.012***
	(0.002)	(0.001)	(0.003)	(0.002)	(0.000)	(0.001)
time_15	−0.034***	−0.029***	0.052***	0.028***	−0.027***	0.010***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)	(0.001)
Observations	2,761	2,761	2,761	2,761	2,761	2,761
Pseudo R^2	0.110	0.110	0.110	0.110	0.110	0.110
Akaike Inf. Crit.	5,630.699	5,630.699	5,630.699	5,630.699	5,630.699	5,630.699

Note: *p<0.1; **p<0.05; ***p<0.01

5.2.3 Cancer subsamples results

The regression results for group of patients diagnosed during the last 7 years are summarised in Table 5.8. The regression results for group of patients diagnosed 8-15 years ago are summarised in Table 5.9.

We can see that the number of significant variables is for both cancer subsamples very similar to the whole sample and whole cancer sample.

For people diagnosed with cancer during the last 7 years, age is highly statistically significant at the 1 % level and its effect is mostly negative. Further, the impact is large as each additional year of life decreases the probability of strong daily depression by 2.7 %, and probability of mild monthly depression by 3.8 %. In contrast, for people with cancer diagnosis 8-15 years old age is not that significant. As we have already explained, it can be due to the fact that older people who have survived cancer have reevaluated their life attitudes and that is why age does not have such a big effect on depression for them.

Female and *lifesat* in both subsamples retained practically the same significance and signs as in the whole sample and whole cancer sample.

Table 5.8: Multinomial Logistic Regression - (1) cancer t $\leq$ 7

	Dependent variable: depression					
	(1) A little/year	(2) A lot/year	(3) A little/month	(4) A lot/month	(5) A little/day	(6) A lot/day
age	0.075***	−0.005	−0.038***	0.000	−0.005**	−0.027***
	(0.009)	(0.006)	(0.009)	(0.002)	(0.002)	(0.005)
agesq	−0.001***	0.000	0.000***	0.000	0.000	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
female	−0.037***	−0.024***	−0.003	0.000	0.031***	0.033***
	(0.009)	(0.008)	(0.004)	(0.000)	(0.007)	(0.006)
south	−0.129***	−0.012	0.137***	0.042**	−0.009**	−0.029***
	(0.028)	(0.008)	(0.033)	(0.019)	(0.004)	(0.005)
hhsize18+	0.006	−0.004	−0.020	0.002	0.051***	−0.034***
	(0.030)	(0.010)	(0.040)	(0.011)	(0.010)	(0.007)
hlthstat	0.093***	−0.010	0.025	0.000	−0.016**	−0.093***
	0.027	(0.018)	(0.024)	(0.002)	(0.007)	(0.017)
lifesat	−0.022	0.046***	0.059***	0.001	−0.021***	−0.063***
	(0.016)	(0.014)	(0.014)	(0.002)	(0.006)	(0.011)
symptcvd	−0.065*	0.008	0.019	0.001	0.073***	−0.036***
	(0.034)	(0.025)	(0.034)	(0.006)	(0.014)	(0.008)
Observations	117	117	117	117	117	117
Pseudo R^2	0.354	0.354	0.354	0.354	0.354	0.354
Akaike Inf. Crit.	271.173	271.173	271.173	271.173	271.173	271.173

Note: *p<0.1; **p<0.05; ***p<0.01

For people who we assume are still being treated for cancer we can observe an unusually large and positive effect of living in the Southern States. As we can see,

living in the south increases the probability of prevalence of mild monthly depression by 13.7 %. Further, *south* is for these people associated with the increase in the probability of having severe monthly depression by 4.2 %. However, the rest of the effects are negative, so our hypothesis that Southern States increases the risk of depression cannot be fully supported.

Another unusually large and unexpectedly positive effect for patients in cancer treatment can be observed for a variable *hhsize18+*. The results suggest that the probability of having mild daily depression will increase by 5.1 % if the respondent lives in a household with another adult.

Respondent's health status has lost its significance for people diagnosed with cancer 8-15 years ago. Since after the entire cancer treatment cancer survivors are expected to be able to find little things to be grateful for, their real health status no longer play such a psychological role for them.

Table 5.9: Multinomial Logistic Regression - (2) cancer $7 < t \leq 15$

	Dependent variable: depression					
	(1) A little/year	(2) A lot/year	(3) A little/month	(4) A lot/month	(5) A little/day	(6) A lot/day
age	−0.025	0.076***	−0.023***	0.001	−0.001	−0.029***
	(0.018)	(0.020)	(0.006)	(0.002)	(0.003)	(0.007)
agesq	0.000	−0.001***	0.000***	0.000	0.000	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
female	−0.130***	0.040***	0.044***	0.017**	0.029***	−0.001
	(0.014)	(0.011)	(0.012)	(0.007)	(0.009)	(0.005)
south	0.043**	0.062***	−0.041***	−0.006*	−0.034***	−0.025***
	(0.018)	(0.016)	(0.009)	(0.004)	(0.011)	(0.008)
hhsize18+	−0.054***	0.020***	0.042	−0.033**	0.001	0.024
	(0.020)	(0.006)	(0.026)	(0.013)	(0.004)	(0.017)
hlthstat	0.070***	−0.017	−0.015	−0.008	0.004	−0.033**
	(0.023)	(0.012)	(0.015)	(0.008)	(0.011)	(0.016)
lifesat	0.135***	0.061***	−0.024***	−0.013**	−0.049***	−0.109***
	(0.023)	(0.016)	(0.008)	(0.006)	(0.016)	(0.023)
Observations	197	197	197	197	197	197
Pseudo R^2	0.213	0.213	0.213	0.213	0.213	0.213
Akaike Inf. Crit.	401.574	401.574	401.574	401.574	401.574	401.574

Note: *p<0.1; **p<0.05; ***p<0.01

5.2.4 Results for non-cancer patients

For comparison, the results of regression for non-cancer patients are presented in Table 5.10.

We can see that the results for *age* are almost identical to the results for the whole sample. Compared to cancer patients, an additional year of life lost significance, and the effects on depression are almost negligible for non-cancer

respondents. As we have already mentioned, people without a history of cancer are not expected to reassess life values.

Female gender is still highly significant and positive for monthly occurrence of depression, as it increases the probability of having mild monthly depression by 3.7 %, and the probability of having severe monthly depression by 1.9 %. This shows that the PMS effect occurs both in women with cancer and in women without it.

It can be seen that respondents who have not been diagnosed with cancer are not affected by whether they reside in the Southern States. Compared to the cancer sample, this variable is almost insignificant.

Hlthstat is now more significant and the negative effect matches our assumptions. As the respondent's health condition improves, the probability of daily strong depression decreases by 3.7 %, of daily mild depression by 1.4 %, the probability of strong monthly depression decreases by 1.6 %, and of strong annual depression by 1 %.

Respondent's general well-being remains statistically significant and negative even for non-cancer patients.

Table 5.10: Multinomial Logistic Regression - the non-cancer sample

	Dependent variable: depression					
	(1) A little/year	(2) A lot/year	(3) A little/month	(4) A lot/month	(5) A little/day	(6) A lot/day
age	0.005***	0.002	−0.006***	−0.001	0.001	−0.001
	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
agesq	0.000	0.000*	0.000***	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
female	−0.086***	0.011	0.037***	0.019***	−0.001	0.020***
	(0.010)	(0.009)	(0.013)	(0.002)	(0.001)	(0.005)
south	−0.006	−0.006	0.011	0.010***	−0.003***	−0.006
	(0.010)	(0.009)	(0.013)	(0.001)	(0.001)	(0.005)
hhsize18+	0.047***	−0.013*	0.006	−0.019***	−0.011***	−0.010*
	(0.010)	(0.007)	(0.009)	(0.006)	(0.002)	(0.006)
hlthstat	0.081***	−0.010*	−0.005	−0.016***	−0.014***	−0.037***
	0.008	(0.005)	(0.006)	(0.004)	(0.003)	(0.005)
lifesat	0.152***	−0.001	−0.027***	−0.025***	−0.011***	−0.087***
	(0.010)	(0.008)	(0.011)	(0.004)	(0.002)	(0.007)
symptcvd	−0.018*	0.017***	−0.016**	0.004	−0.005	0.018***
	(0.010)	(0.006)	(0.008)	(0.005)	(0.003)	(0.005)
Observations	2,485	2,485	2,485	2,485	2,485	2,485
Pseudo R^2	0.110	0.110	0.110	0.110	0.110	0.110
Akaike Inf. Crit.	5,017.0	5,017.0	5,017.0	5,017.0	5,017.0	5,017.0
Note:						*p<0.1; **p<0.05; ***p<0.01

The coronavirus symptoms are more significant compared to cancer patients, which is consistent with our assumption. However, their effect on the respondents'

self-reported level of depression varies. But it is certainly important that it has a significant positive effect on severe depression, when with worsening COVID-19 symptoms the probability of daily and annual depression increases by 1.8 % and 1.7 %, respectively.

Chapter 6

Discussion

The results of our analyzes discover connection between determinants contributing to the development of depression and its seriousness among cancer patients. Robustness checks were carried out on the overall and non-cancer samples. The essential element for this relationship is gender of the individual. Being female increases the probability of taking the medication for depression by 10-15 % for all people, regardless of whether they have cancer or not. At the same time, it significantly increases the chance of depression in any form at monthly intervals. The greater vulnerability to mental imbalances is likely the effect of intense hormonal fluctuation caused by a menstrual cycle in women.

Since surviving cancer comes with many new opportunities and people often believe that their life will be different and better after finishing the treatment, we observe a decreasing probability of depression for cancer survivors as they get older. Contrary to that, for people who have not had to face such a diagnosis in their lives, additional year of age is not reflected in their depressive disorder, as it is statistically insignificant.

We can say that the household composition has a significant effect on the probability of depression and is associated with decreasing depression prevalence. However, this cannot be said for a sample of people who were diagnosed with cancer 8-15 years ago and it is therefore assumed that they are already cured. For these people, the number of adults in the household is insignificant when referring to depression. It might be an effect of gained strength of character through cancer survivorship. As people have been through a lot, they have acquired coping skills to be able to handle stressful situations by themselves without the help of other people around them. On the contrary, among people still in the process of cancer treatment, a positive effect of another adult in the household was observed, which

increases the probability that the respondent will suffer from daily depression by 5.1 %. It often happens that a person with cancer worries much more about their loved ones than about themselves - how their partner will be able to take care of the household alone during a possible hospitalization or how the partner will be able to deal with the consequences in case of unsuccessful treatment. Not having to worry about your loved ones would solve worsening depression in such cases.

Effects of general health status and life satisfaction are present with significant decreases in the incidence of depression for all people, with one exception. In the two groups divided based on how long respondents have the cancer diagnosis, the life satisfaction is insignificant. In the first case, people may be so busy treating cancer that they do not really care how happy they are in real life. In the second case, when people are already cured of cancer, they tend to have a different outlook on life and attach less importance on less important things. Therefore, their life satisfaction is on a slightly different scale. Conversely, educational attainment was found to be the most surprising, as it is positively correlated with depression symptoms in all cases. However, it is a bit intuitive that a higher education goes hand in hand with better job positions, which are often associated with higher stress levels due to excessive workloads and control over the work of their subordinates. For people who are now being treated for cancer, this effect is insignificant, since they are most likely not working at the moment and thus education does not affect their current mental condition.

The results of the effect of the severity of coronavirus symptoms are important to us. In the main cancer sample, likewise in the group of patients with cancer no more than 7 years, the coronavirus symptoms were found to be insignificant when reporting the probability of depression medication as well as the severity of depression. The situation is however different among respondents not suffering from cancer. Those people report statistically significant and positive effect between coronavirus symptoms and depression states, because as the progress of covid worsens, the odds of depression occurrence increases by 2.1 %. Furthermore, for people without cancer, covid meant an increased chance of severe daily depression by almost 2 %. When the coronavirus pandemic hit and the quarantine was imposed, all people were locked at home and glued to their television screens watching the events of the world. At that moment, no one knew how it all will turn out and how to defend against the disease. For many, it was an unexpected shock, and when they were diagnosed with covid, they panicked about what to do with their deteriorating condition. This all leading to the creation of pressure on mental health and subsequent psychological distress associated with coronavirus

symptoms and isolation. Meanwhile, people with cancer have probably already been through a similar trauma. A cancer diagnosis is quite a shocking piece of information, so they have already been through the whole process of stressing about the future. Thus, this new disease, that was not known much about at the time, just did not stress them out.

Among cancer patients, the fact that people are likely to have successfully completed cancer treatment significantly reduces the likelihood of depression. People have become mentally immune due to cancer and are not that sensitive to negative circumstances. On the other hand, in the entire sample, the fact that a patient has ever been diagnosed with cancer intuitively suggests that his chances of depression are greater.

Hypothesis of living in the Southern States increasing the severity of depression is not supported in neither sample, as the significance is completely different for each sample.

We have to also acknowledge that all mild depression results in our multinomial regression seem odd. We assume these results may be caused by human error. People probably wrongly evaluated their past, or mild depression was so negligible to them that they do not even remember if they were a little depressed last year. However, we did not gain evidence that this is in fact the case.

Chapter 7

Conclusion

The purpose of this thesis was to analyse the impact of sociodemographic characteristics and health-related factors on the incidence of depression among the American cancer population. The special focus was placed on how people with cancer reacted to the COVID-19 pandemic. In total, we used five different samples. The entire original sample, which was further divided into a subsample of non-cancer respondents and a subsample of cancer respondents. The subsample of cancer respondents was then further divided into two small samples based on how long patients have been living with the diagnosis. For the analysis, questionnaire of NHIS from 2022 was used.

For the investigation of the binary dependent variable representing taking medication for depression or anxiety, we applied logistic regression. Among important independent variables used in our analysis were *age*, *female*, *hhsize18+* describing number of persons 18 or older in the respondent's household, *educ* indicating the highest educational level, *hlthstat* explaining the subjective assessment of the state of health at the time of interview, *lifesat* which portrays how satisfied the respondent is with his life, and seriousness of coronavirus symptoms (*symptcvd*). Region category variable *south* does not play a role in taking the medication, and thus is not included. By employing this dependent variable in all our five samples we examined whether determinants of mental health change between all, cancer, and non-cancer respondents.

In the second part of the analysis, we estimated the effect of the similar set of independent variables on the severity of depression symptoms using multinomial logistic regression. Here, the highest educational attainment seems to play no role on the level of depression, and thus is excluded. Again, we employed this dependent variable in all five samples so that we could see the changes between

those groups.

Most of the variables included in all analysis were found to be statistically significant. Overall, the variable with the strongest effect and significance was *female*. Being female had the highest statistical significance and, on average, was associated with an 12 % increase in the probability of depression.

Main point of interest was the difference in reaction to the COVID-19 pandemic between cancer and non-cancer patients. Consistent with Tolia et al. (2023), our analysis suggests that for people diagnosed with cancer and for people currently in cancer treatment the coronavirus symptoms play no role in the depression. On the contrary, for people who have never been diagnosed with cancer the coronavirus symptoms were highly statistically significant, increasing the threat of depression by 2.1 %.

In all our samples, age was negatively correlated with depression occurrence and despite its significance, the effect of aging was noticeably small. Moreover, for respondents without cancer, age did not prove to be significant on the depression. Further, general health status and life satisfaction were also negatively correlated, but in the cancer subsamples the life satisfaction was insignificant. Depression also decreases with additional adult in the household, but there is no effect of household composition for cancer survivors.

The results do not correspond to the majority of studies we found about relationship between depression and educational achievement. Thus, our hypothesis on decreasing depression with higher level of education is not supported, and in addition, for people currently being treated for cancer, education does not play a role. The same can be said about southern region. Living in the Southern States does not support our theory of increasing severity of depression symptoms.

In the cancer sample, the variable representing people for whom we assume their cancer treatment has ended was significantly negatively correlated with the incidence of depression, while variable representing ongoing cancer treatment was found to be insignificant. In the whole sample, both ever being told about cancer diagnosis and having ongoing cancer treatment were significantly positively correlated with depression.

In summary, we came to the conclusion that, among both cancer patients and non-cancer patients, women are much more prone to depression compared to men, and also that the risk of depression for cancer patients decreases with age. Furthermore, among the most significant variables are those related to general health status and household composition.

The results also confirmed the hypothesis about our primary point of interest

that the coronavirus pandemic did not affect the psychological health of people with cancer, as they already had enough to worry about.

As a result of the coronavirus pandemic, when people were unable to access face-to-face care and looked for help on the internet, the concept of mobile health (mHealth) apps has taken off. Chow et al. (2020) cite mobile phone apps as a potential way to expand effective care in a cost-effective manner, as 81 % of American adults own a cell phone. Apps for patients with cancer help to reduce the symptoms of depression and anxiety and gain more autonomy and self-acceptance. The first mobile application for mental health support for oncology patients in the Czech Republic, developed in the middle of 2022, is called MOÚ MindCare. Its goal is to improve the mental state of patients with the help of three interventions - mindfulness, positive psychology and autogenic training (Světlák et al., 2021). However, since the concept of digital health is relatively new, more research needs to be done so that the technologies can be adapted to all minorities battling cancer who are at risk for depression.

Our study thus implies that adjustments should be made to clinical practice guidelines focusing on the female gender as a risk group. Furthermore, great emphasis should be placed on improving family policy, as living with another adult reduces depression among people fighting cancer. Developers of health applications should therefore focus on special sections dedicated to women and people who live alone, as these two factors represent vulnerable groups prone to depression.

This analysis was performed on a cross-sectional dataset collected continuously throughout the year. Panel data could serve as a further extension of this thesis, as such data allow to see changes in the dependent variable and predict trends. Additionally, including respondent's life history or cancer-related changes could help in a deeper understanding of depression among cancer patients.

NHIS acknowledgement

National Center for Health Statistics. National Health Interview Survey, 2022. Public-use data file and documentation. https://www.cdc.gov/nchs/nhis/data-questionnaires-documentation.htm. 2023.

Bibliography

Akechi, T., Okuyama, T., Sugawara, Y., Nakano, T., Shima, Y., and Uchitomi, Y. (2004). Suicidality in terminally ill Japanese patients with cancer: Prevalence, patient perceptions, contributing factors, and longitudinal changes. *Cancer: Interdisciplinary International Journal of the American Cancer Society*, 100(1):183–191.

Akin-Odanye, E., Asuzu, C. C., and Popoola, O. A. (2011). Measured effect of some socio-demographic factors on depression among breast cancer patients receiving chemotherapy in Lagos State University Teaching Hospital (LASUTH). *African Health Sciences*, 11(3).

Andersen, B. L., Lacchetti, C., Ashing, K., Berek, J. S., Berman, B. S., Bolte, S., Dizon, D. S., Given, B., Nekhlyudov, L., Pirl, W., et al. (2023). Management of Anxiety and Depression in Adult Survivors of Cancer: ASCO Guideline Update. *Journal of Clinical Oncology*, 41(18):3426–3453.

Anguiano, L., Mayer, D. K., Piven, M. L., and Rosenstein, D. (2012). A literature review of suicide in cancer patients. *Cancer nursing*, 35(4):E14–E26.

APA (2022). How to prevent depression as you age. `https://www.apa.org/topics/aging-older-adults/depression`.

Bianchi, R., Schonfeld, I. S., and Laurent, E. (2018). Burnout syndrome and depression. *Understanding depression: Volume 2. Clinical manifestations, diagnosis and treatment*, pages 187–202.

Breitbart, W. (2018). Cancer, Depression, and Suicide Risk: Signs to Watch For. `https://www.cancer.net/blog/2018-08/cancer-depression-and-suicide-risk-signs-watch`.

Cancer Research UK (2020). Types of cancer. `https://www.cancerresearchuk.org/what-is-cancer/how-cancer-starts/types-of-cancer`.

CDC (2023). National, State-Level, and County-level Prevalence Estimates of Adults Aged ≥ 18 Years Self-Reporting a Lifetime Diagnosis of Depression – United States, 2020. https://www.cdc.gov/mmwr/volumes/72/wr/mm7224a1.htm.

Chow, P. I., Showalter, S. L., Gerber, M., Kennedy, E. M., Brenin, D., Mohr, D. C., Lattie, E. G., Gupta, A., Ocker, G., and Cohn, W. F. (2020). Use of mental health apps by patients with breast cancer in the united states: pilot pre-post study. *JMIR cancer*, 6(1):e16476.

Coyne, J. C., Stefanek, M., and Palmer, S. C. (2007). Psychotherapy and survival in cancer: the conflict between hope and evidence. *Psychological bulletin*, 133(3):367.

Deshields, T. L. (2017). How to Recognize Cancer Distress – and Cope with It. https://www.cancer.net/blog/2017-11/how-recognize-cancer-distress-%E2%80%94-and-cope-with-it.

Dusek, L., Muzik, J., Maluskova, D., Májek, O., Pavlik, T., Koptíková, J., Melichar, B., Büchler, T., Finek, J., Cibula, D., et al. (2014). Cancer incidence and mortality in the Czech Republic. *Klin Onkol*, 27(6):406–423.

Epstein, R. M. and Street Jr, R. L. (2007). Patient-centered communication in cancer care: promoting healing and reducing suffering.

European Commission (2017). Czech Republic Country Health Profile 2017. https://health.ec.europa.eu/system/files/2017-12/chp_cs_english_0.pdf.

European Commission (2021). Czechia Country Health Profile 2021. https://health.ec.europa.eu/system/files/2021-12/2021_chp_cs_english.pdf.

Fitch, M. I. (2022). Reducing the global burden of cancer. *Asia-Pacific Journal of Oncology Nursing*, 9(9).

Friberg, A. S., Rask Moustsen, I., Benzon Larsen, S., Hartung, T., Wreford Andersen, E., Halgren Olsen, M., Tjønneland, A., Kjaer, S. K., Johansen, C., Brasso, K., et al. (2019). Educational level and the risk of depression after prostate cancer. *Acta oncologica*, 58(5):722–729.

Greene, H. (2016). Cancer prevention, screening, and early detection. *Advanced Oncology Nursing Certification Review and Resource Manual. 3rd ed. Oncology Nursing Society*, pages 1–34.

Greene, W. H. (2018). Econometric Analysis.

Hartung, T., Brähler, E., Faller, H., Härter, M., Hinz, A., Johansen, C., Keller, M., Koch, U., Schulz, H., Weis, J., et al. (2017). The risk of being depressed is significantly higher in cancer patients than in the general population: prevalence and severity of depressive symptoms across major cancer types. *European Journal of Cancer*, 72:46–53.

Hong, Y., Yuhan, L., Youhui, G., Zhanying, W., Shili, Z., Xiaoting, H., and Wenhua, Y. (2022). Death anxiety among advanced cancer patients: a cross-sectional survey. *Supportive Care in Cancer*, pages 1–9.

Kneier, A. W. (2012). Coping with Depression. *Cureus*.

Leonard, J. (2020). Depression during period: Everything you need to know. `https://www.medicalnewstoday.com/articles/327490?c=186101158224`.

Luber, M. P., Meyers, B. S., Williams-Russo, P. G., Hollenberg, J. P., DiDomenico, T. N., Charlson, M. E., and Alexopoulos, G. S. (2001). Depression and service utilization in elderly primary care patients. *The American Journal of Geriatric Psychiatry*, 9(2):169–176.

Luger, T. M., Suls, J., and Vander Weg, M. W. (2014). How robust is the association between smoking and depression in adults? A meta-analysis using linear mixed-effects models. *Addictive behaviors*, 39(10):1418–1429.

Maneeton, B., Maneeton, N., and Mahathep, P. (2012). Prevalence of depression and its correlations: a cross-sectional study in Thai cancer patients. *Asian Pacific Journal of Cancer Prevention*, 13(5):2039–2043.

Massie, M., Popkin, M., and Holland, J. (1998). Psycho-oncology.

Mausbach, B. T. and Irwin, S. A. (2017). Depression and healthcare service utilization in patients with cancer. *Psycho-oncology*, 26(8):1133–1139.

Mayer, K. (2023). Things That May Look Like Depression but Aren't. `https://www.webmd.com/depression/looks-like-depression-but-not`.

McFarland, M. J. and Wagner, B. G. (2015). Does a college education reduce depressive symptoms in American young adults? *Social Science & Medicine*, 146:75–84.

Mesothelioma Web (2016). Physical Carcinogens – What Are They and How Do They Act? `https://www.mesotheliomaweb.org/carcinogens.htm`.

Moleyar-Narayana, P. and Ranganathan, S. (2020). Cancer screening.

Mössinger, H. and Kostev, K. (2023). Depression is associated with an increased risk of subsequent cancer diagnosis: a retrospective cohort study with 235,404 patients. *Brain Sciences*, 13(2):302.

Nocera, S. and Zweifel, P. (1998). The demand for health: an empirical test of the Grossman model using panel data. In *Health, the medical profession, and regulation*, pages 35–49. Springer.

Noyes Jr, R. (1999). The relationship of hypochondriasis to anxiety disorders. *General Hospital Psychiatry*, 21(1):8–17.

Patrick, D., Unutzer, J., and Simon, G. (1997). Depressive symptoms and the cost of health services in HMO patients aged 65 years and older: A four year perspective study. *JAMA*, 277(20):1618–1623.

Pudrovska, T. (2010). Why is cancer more depressing for men than women among older white adults? *Social Forces*, 89(2):535–558.

Robinson, E., Sutin, A. R., Daly, M., and Jones, A. (2022). A systematic review and meta-analysis of longitudinal cohort studies comparing mental health before versus during the COVID-19 pandemic in 2020. *Journal of affective disorders*, 296:567–576.

Satin, J. R., Linden, W., and Phillips, M. J. (2009). Depression as a predictor of disease progression and mortality in cancer patients: a meta-analysis. *Cancer*, 115(22):5349–5361.

Scott, H. (2016). Life Isn't the Same: How Cancer Changes You. `https://www.roswellpark.org/cancertalk/201612/life-isnt-same-how-cancer-changes-you`.

Sethi, G., Shanmugam, M. K., Ramachandran, L., Kumar, A. P., and Tergaonkar, V. (2012). Multifaceted link between cancer and inflammation. *Bioscience reports*, 32(1):1–15.

Shreders, A., Niazi, S., Hodge, D., Chimato, N., Agarwal, A., Gustetic, E., Hammond, W. A., Dholaria, B. R., and Ailawadhi, S. (2016). Universal screening for depression in cancer patients and its impact on management patterns.

Smith, H. R. (2015). Depression in cancer patients: Pathogenesis, implications and treatment. *Oncology letters*, 9(4):1509–1514.

Spiegel, D. and Giese-Davis, J. (2003). Depression and cancer: mechanisms and disease progression. *Biological psychiatry*, 54(3):269–282.

St. Jude Children's Research Hospital (2024). What is a Solid Tumor? `https://www.stjude.org/treatment/disease/solid-tumors/what-is-solid-tumor.html`.

Světlák, M., Šumec, R., Slezáčková, A., Humpolíček, P., Lekárová, M., Barešová, Z., Vigašová, D., Malatincová, T., Šedo, J., Halámková, J., et al. (2021). eHealth v podpoře duševního zdraví onkologicky nemocných: mobilní aplikace MindCare a její možnosti v rámci "cancer survivorship".

Tolia, M., Symvoulakis, E. K., Matalliotakis, E., Kamekis, A., Adamou, M., Kountourakis, P., Mauri, D., Dakanalis, A., Alexidis, P., Varveris, A., et al. (2023). COVID-19 emotional and mental impact on cancer patients receiving radiotherapy: an interpretation of potential explaining descriptors. *Current Oncology*, 30(1):586–597.

ÚZIS (2018). Appendix 2: Example summary report. `https://www.ipaac.eu/res/file/outputs/wp7/comprehensive-ict-model-multiple-data-sources-a2.pdf`.

ÚZIS (2021). Celková zátěž zhoubnými novotvary v ČR. `https://www.uzis.cz/index.php?pg=aktuality&aid=8466`.

Čabanová, A. (2024). Co lze ještě zlepšit na onkologických screeninzích. `https://www.tribune.cz/zdravotnictvi/co-lze-jeste-zlepsit-na-onkologickych-screeninzich/`.

WCRF (2022). Worldwide cancer data. `https://www.wcrf.org/cancer-trends/worldwide-cancer-data/`.

WHO (2021). Childhood cancer. `https://www.who.int/news-room/fact-sheets/detail/cancer-in-children`.

WHO (2022a). Cancer. https://www.who.int/news-room/fact-sheets/detail/cancer.

WHO (2022b). COVID-19 pandemic triggers 25 % increase in prevalence of anxiety and depression worldwide. https://www.who.int/news/item/02-03-2022-covid-19-pandemic-triggers-25-increase-in-prevalence-of-anxiety-and-depression-worldwide.

WHO (2023). Depressive disorder (depression). https://www.who.int/news-room/fact-sheets/detail/depression.

WHO (2024). Global Cancer Burden Growing, Amidst Mounting Need for Services. https://www.who.int/news/item/01-02-2024-global-cancer-burden-growing--amidst-mounting-need-for-services.

Wooldridge, J. M. (2010). Econometric Analysis of Cross Section and Panel Data.

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

Xiong, J., Lipsitz, O., Nasri, F., Lui, L. M., Gill, H., Phan, L., Chen-Li, D., Iacobucci, M., Ho, R., Majeed, A., et al. (2020). Impact of COVID-19 pandemic on mental health in the general population: A systematic review. *Journal of affective disorders*, 277:55–64.

Appendix A

Descriptive Statistics

Table A.1: Descriptive Statistics for the cancer sample

Statistic	Mean	St. Dev.	Min	Max	Median
dependent					
medic	0.216	0.412	0	1	0
depression	2.019	1.686	1	6	1
independent					
age	69.754	10.994	39	85	71
hhsize18+	1.655	0.620	1	3	2
educ	6.132	2.426	1	10	6
hlthstat	3.182	1.120	1	5	3
lifesat	3.381	0.650	1	4	3
symptcvd	1.560	0.928	0	3	2
dummy					
female	0.575	0.494	0	1	1
south	0.369	0.483	0	1	0
time_7	0.438	0.496	0	1	0
time_15	0.239	0.427	0	1	0

Table A.2: Descriptive Statistics for the non-cancer sample

Statistic	Mean	St. Dev.	Min	Max	Median
dependent					
medic	0.167	0.373	0	1	0
depression	1.893	1.524	1	6	1
independent					
age	50.748	18.054	18	85	51
hhsize18+	1.805	0.685	1	3	2
educ	5.915	2.429	1	10	6
hlthstat	3.613	1.034	1	5	4
lifesat	3.393	0.599	1	4	3
symptcvd	1.548	0.868	0	3	2
dummy					
female	0.539	0.498	0	1	1
south	0.368	0.482	0	1	0

Table A.3: Descriptive Statistics for the whole sample

Statistic	Mean	St. Dev.	Min	Max	Median
dependent					
medic	0.173	0.379	0	1	0
depression	1.914	1.549	1	6	1
independent					
age	52.949	18.439	18	85	54
hhsize18+	1.787	0.679	1	3	2
educ	5.944	2.429	1	10	6
hlthstat	3.560	1.055	1	5	4
lifesat	3.391	0.605	1	4	3
symptcvd	1.550	0.874	0	3	2
dummy					
female	0.544	0.498	0	1	1
south	0.368	0.482	0	1	0
cancer	0.124	0.330	0	1	0
time_7	0.055	0.228	0	1	0
time_15	0.030	0.170	0	1	0

Table A.4: Descriptive Statistics for cancer t $\leq$ 7

Statistic	Mean	St. Dev.	Min	Max	Median
dependent					
medic	0.220	0.414	0	1	0
depression	2.014	1.638	1	6	1
independent					
age	67.439	12.029	34	85	68
hhsize18+	1.691	0.621	1	3	2
educ	6.148	2.419	1	10	6
hlthstat	3.148	1.126	1	5	3
lifesat	3.383	0.642	1	4	3
symptcvd	1.565	0.926	0	3	1
dummy					
female	0.535	0.499	0	1	1
south	0.359	0.480	0	1	0

Table A.5: Descriptive Statistics for cancer 7 $<$ t $\leq$ 15

Statistic	Mean	St. Dev.	Min	Max	Median
dependent					
medic	0.177	0.382	0	1	0
depression	1.929	1.677	1	6	1
independent					
age	70.167	10.592	41	85	71
hhsize18+	1.672	0.612	1	3	2
educ	6.215	2.362	1	10	7
hlthstat	3.258	1.121	1	5	3
lifesat	3.405	0.645	1	4	3
dummy					
female	0.535	0.499	0	1	1
south	0.363	0.481	0	1	0

Appendix B

Correlation Matrices

Table B.1: Correlation Matrix for the cancer sample

	medic	depression	age	female	south	hhsize18+	educ	hlthstat	lifesat	symptcvd	time_7	time_15
medic	1	0.389	−0.137	0.114	0.044	−0.018	−0.032	−0.181	−0.169	0.070	0.005	−0.054
depression	0.389	1	−0.114	0.055	0.014	−0.062	−0.151	−0.336	−0.431	0.136	−0.015	−0.026
age	−0.137	−0.114	1	−0.085	0.013	−0.184	−0.069	−0.024	0.041	−0.092	−0.127	0.005
female	0.114	0.055	−0.085	1	−0.015	−0.067	−0.098	0.042	−0.036	0.062	−0.081	−0.045
south	0.044	0.014	0.013	−0.015	1	−0.045	−0.053	−0.070	−0.010	0.046	−0.014	−0.009
hhsize18+	−0.018	−0.062	−0.184	−0.067	−0.045	1	0.024	−0.011	0.102	0.014	0.044	0.017
educ	−0.032	−0.151	−0.069	−0.098	−0.053	0.024	1	0.273	0.149	−0.025	0.0002	0.020
hlthstat	−0.181	−0.336	−0.024	0.042	−0.070	−0.011	0.273	1	0.419	−0.153	−0.031	0.035
lifesat	−0.169	−0.431	0.041	−0.036	−0.010	0.102	0.149	0.419	1	−0.073	0.004	0.021
symptcvd	0.070	0.136	−0.092	0.062	0.046	0.014	−0.025	−0.153	−0.073	1	−0.005	0.007
time_7	0.005	−0.015	−0.127	−0.081	−0.014	0.044	0.0002	−0.031	0.004	−0.005	1	−0.495
time_15	−0.054	−0.026	0.005	−0.045	−0.009	0.017	0.020	0.035	0.021	0.007	−0.495	1

Table B.2: Correlation Matrix for the non-cancer sample

	medic	depression	age	female	south	hhsize18+	educ	hlthstat	lifesat	symptcvd
medic	1	0.370	0.021	0.135	0.001	−0.041	0.005	−0.198	−0.153	0.083
depression	0.370	1	−0.036	0.057	0.016	−0.082	−0.147	−0.308	−0.385	0.079
age	0.021	−0.036	1	0.048	−0.002	−0.204	−0.069	−0.230	−0.0005	−0.059
female	0.135	0.057	0.048	1	0.022	−0.042	0.024	−0.006	0.021	0.062
south	0.001	0.016	−0.002	0.022	1	−0.021	−0.071	−0.040	0.020	−0.013
hhsize18+	−0.041	−0.082	−0.204	−0.042	−0.021	1	−0.045	0.054	0.090	−0.006
educ	0.005	−0.147	−0.069	0.024	−0.071	−0.045	1	0.259	0.144	−0.016
hlthstat	−0.198	−0.308	−0.230	−0.006	−0.040	0.054	0.259	1	0.377	−0.101
lifesat	−0.153	−0.385	−0.0005	0.021	0.020	0.090	0.144	0.377	1	−0.060
symptcvd	0.083	0.079	−0.059	0.062	−0.013	−0.006	−0.016	−0.101	−0.060	1

Table B.3: Correlation Matrix for the whole sample

	medic	depression	age	female	south	hhsize18+	educ	hlthstat	lifesat	symptcvd	cancer	time_7	time_15
medic	1	0.374	0.020	0.133	0.007	−0.041	0.001	−0.200	−0.155	0.080	0.045	0.031	0.003
depression	0.374	1	−0.034	0.059	0.017	−0.079	−0.146	−0.313	−0.391	0.087	0.036	0.021	0.009
age	0.020	−0.034	1	0.043	0.001	−0.213	−0.056	−0.238	0.001	−0.057	0.317	0.177	0.149
female	0.133	0.059	0.043	1	0.018	−0.046	0.010	−0.004	0.013	0.061	0.025	−0.005	−0.002
south	0.007	0.017	0.001	0.018	1	−0.023	−0.068	−0.043	0.017	−0.008	−0.001	−0.006	−0.002
hhsize18+	−0.041	−0.079	−0.213	−0.046	−0.023	1	−0.039	0.055	0.091	−0.004	−0.070	−0.032	−0.028
educ	0.001	−0.146	−0.056	0.010	−0.068	−0.039	1	0.254	0.144	−0.019	0.031	0.022	0.019
hlthstat	−0.200	−0.313	−0.238	−0.004	−0.043	0.055	0.254	1	0.380	−0.108	−0.133	−0.092	−0.052
lifesat	−0.155	−0.391	0.001	0.013	0.017	0.091	0.144	0.380	1	−0.063	−0.008	−0.004	0.002
symptcvd	0.080	0.087	−0.057	0.061	−0.008	−0.004	−0.019	−0.108	−0.063	1	0.007	0.005	0.007
cancer	0.045	0.036	0.317	0.025	−0.001	−0.070	0.031	−0.133	−0.008	0.007	1	0.641	0.465
time_7	0.031	0.021	0.177	−0.005	−0.006	−0.032	0.022	−0.092	−0.004	0.005	0.641	1	−0.042
time_15	0.003	0.009	0.149	−0.002	−0.002	−0.028	0.019	−0.052	0.002	0.007	0.465	−0.042	1

Table B.4: Correlation Matrix for cancer t $\leq$ 7

	medic	depression	age	female	south	hhsize18+	educ	hlthstat	lifesat	symptcvd
medic	1	0.391	−0.122	0.118	0.035	−0.044	−0.046	−0.171	−0.175	0.056
depression	0.391	1	−0.112	0.004	0.078	−0.044	−0.088	−0.337	−0.372	0.196
age	−0.122	−0.112	1	−0.125	−0.005	−0.215	−0.100	−0.050	0.049	−0.067
female	0.118	0.004	−0.125	1	−0.001	−0.011	−0.081	0.054	−0.005	0.056
south	0.035	0.078	−0.005	−0.001	1	−0.051	−0.055	−0.061	−0.023	0.081
hhsize18+	−0.044	−0.044	−0.215	−0.011	−0.051	1	0.032	0.023	0.091	0.037
educ	−0.046	−0.088	−0.100	−0.081	−0.055	0.032	1	0.268	0.131	0.001
hlthstat	−0.171	−0.337	−0.050	0.054	−0.061	0.023	0.268	1	0.423	−0.191
lifesat	−0.175	−0.372	0.049	−0.005	−0.023	0.091	0.131	0.423	1	−0.072
symptcvd	0.056	0.196	−0.067	0.056	0.081	0.037	0.001	−0.191	−0.072	1

Table B.5: Correlation Matrix for cancer 7 < t $\leq$ 15

	medic	depression	age	female	south	hhsize18+	educ	hlthstat	lifesat
medic	1	0.335	−0.117	0.109	0.019	0.021	−0.001	−0.200	−0.130
depression	0.335	1	−0.174	0.117	−0.064	−0.018	−0.252	−0.294	−0.467
age	−0.117	−0.174	1	−0.078	0.085	−0.125	−0.070	−0.050	0.026
female	0.109	0.117	−0.078	1	−0.064	−0.111	−0.088	0.005	−0.070
south	0.019	−0.064	0.085	−0.064	1	−0.051	−0.031	−0.072	−0.014
hhsize18+	0.021	−0.018	−0.125	−0.111	−0.051	1	0.042	0.021	0.122
educ	−0.001	−0.252	−0.070	−0.088	−0.031	0.042	1	0.279	0.171
hlthstat	−0.200	−0.294	−0.050	0.005	−0.072	0.021	0.279	1	0.411
lifesat	−0.130	−0.467	0.026	−0.070	−0.014	0.122	0.171	0.411	1

Appendix C

Average Marginal Effects

Table C.1: AME

	Dependent variable: medic					
	Cancer sample	Whole sample with *cancer*	Whole sample with time intervals	Cancer t $\leq$ 7	Cancer 7 $<$ t $\leq$ 15	Non-cancer sample
age	-0.005^{***}	-0.001^{**}	-0.000^{*}	-0.005^{***}	-0.004^{***}	-0.000
female	0.132^{***}	0.116^{***}	0.116^{***}	0.150^{***}	0.084^{***}	0.112^{***}
hhsize18+	-0.039^{*}	-0.016^{***}	-0.016^{***}	-0.081^{**}	0.016	-0.014^{**}
educ	0.019^{***}	0.012^{***}	0.012^{***}	0.013	0.011^{*}	0.011^{***}
hlthstat	-0.072^{***}	-0.063^{***}	-0.064^{***}	-0.085^{***}	-0.067^{***}	-0.061^{***}
lifesat	-0.040^{*}	-0.057^{***}	-0.057^{***}	-0.013	-0.024	-0.060^{***}
symptcvd	0.006	0.019^{***}	0.019^{***}	-0.003	-	0.021^{***}
cancer	-	0.027^{**}	-	-	-	-
time_7	-0.029	-	0.029^{*}	-	-	-
time_15	-0.081^{**}	-	-0.019	-	-	-
Observations	940	9,931	9,931	416	764	8,938

Appendix D

Model Evaluation tests results

LR test

Likelihood ratio test (binary model) results

Likelihood ratio test	H0: The full model and the nested model fit the data equally well
Chisq = 7.9248	Pr(>Chisq) = 0.2437

Logistic Regression - Model 1 & Model 2

	Dependent variable: medic	
	Model 1	Model 2
age	-0.040^{***}	-0.030^{**}
	(0.011)	(0.012)
female	0.909^{***}	0.884^{***}
	(0.259)	(0.264)
south		0.687^{*}
		(0.370)
west		0.235
		(0.421)
midwest		0.126
		(0.414)
hhsize18+	−0.218	−0.246
	(0.197)	(0.205)
hhsize<18		0.205
		(0.242)
hispanic		0.526
		(0.520)
educ	0.151^{***}	0.154^{***}
	(0.056)	(0.057)
hlthstat	-0.373^{***}	-0.337^{***}
	(0.118)	(0.122)
lifesat	-0.536^{***}	-0.579^{***}
	(0.198)	(0.206)
smknow		0.210
		(0.173)
symptcvd	−0.157	−0.191
	(0.132)	(0.137)
time_7	0.048	0.060
	(0.268)	(0.276)
time_15	-0.757^{**}	-0.731^{**}
	(0.353)	(0.354)
Constant	3.858^{***}	2.565^{**}
	(1.070)	(1.249)
Observations	427	427
Pseudo R^2	0.155	0.171
Log Likelihood	−212.677	−208.714
Akaike Inf. Crit.	445.354	449.429

We can see that since in both cases the corresponding p-value is greater than .05, we fail to reject the null hypothesis. Therefore, full model and nested model fit the data equally well, and the nested model is preferred. Insignificant variables excluded from the models are: *west* (dummy variable representing whether the household is located in the Western States), *midwest* (dummy representing whether the household is in the Midwestern United States), *hhsize<18* (categorical variable determining the number of persons under 18 in the household), *hispanic* (dummy variable indicating whether the respondent is hispanic) and *smknow* (variable indicating the severity of the smoking habit). Further, *south* (dummy variable if the household is in the Southern States) is insignificant for the probability of depression, and conversely, *educ* (categorical variable describing the highest educational attainment) is insignificant for the severity of depressive symptoms.

Likelihood ratio test (multinomial model) results

Likelihood ratio test	H0: The full model and the nested model fit the data equally well
Chisq = 25.162	Pr(>Chisq) = 0.7171

Multinomial Logistic Regression - Model 1 & Model 2

	Dependent variable: depression									
	Model 1					Model 2				
	(2)	(3)	(4)	(5)	(6)	(2)	(3)	(4)	(5)	(6)
age	−0.078	0.583***	−0.120	0.084	−43.201***	−0.960***	0.678***	−0.028	−0.454***	0.825***
	(0.069)	(0.069)	(0.077)	(0.073)	(0.216)	(0.078)	(0.083)	(0.098)	(0.090)	(0.283)
agesq	0.0002	−0.004***	0.0004	−0.0004	0.294***	0.006***	−0.005***	−0.0004	0.003***	−0.028***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.004)	(0.001)	(0.001)	(0.001)	(0.001)	(0.005)
female	−1.205***	−0.210**	0.959***	−0.732***	116.921***	−1.104***	0.215	0.872***	−1.501***	42.294***
	(0.122)	(0.106)	(0.045)	(0.057)	(0.004)	(0.049)	(0.147)	(0.095)	(0.074)	(0.009)
south	−1.140***	−1.094***	−1.120***	1.342***	27.871***	−0.255	−0.607**	51.198***	29.453***	70.312***
	(0.127)	(0.054)	(0.099)	(0.031)	(0.010)	(0.187)	(0.239)	(0.191)	(0.048)	(0.007)
west						1.348***	0.931***	52.859***	29.004***	108.746***
						(0.159)	(0.194)	(0.148)	(0.049)	(0.004)
midwest						2.230***	−0.254***	52.198***	−48.417	50.271***
						(0.122)	(0.058)	(0.204)		(0.001)
hhsize18+	−0.703	−0.709***	0.681	0.166	49.494***	−0.444	−0.720*	0.451	−0.195	7.257***
	(0.444)	(0.202)	(0.415)	(0.127)	(0.014)	(0.304)	(0.427)	(0.449)	(0.119)	(0.016)
hhsize<18						−4.122***	−32.753	−0.354	−2.459***	−10.416***
						(0.072)		(0.217)	(0.094)	(0.002)
hispanic						2.239***	−58.424	−68.930***	−74.663***	22.808***
						(0.041)		(0.000)	(0.000)	(0.001)
educ						0.461**	−0.188	−0.046	−0.083	−0.801***
						(0.211)	(0.231)	(0.207)	(0.265)	(0.071)
hlthstat	−0.230	0.002	0.052	−0.021	−43.359***	−0.635*	0.172	0.142	0.111	−2.389***
	(0.345)	(0.338)	(0.370)	(0.487)	(0.001)	(0.368)	(0.384)	(0.375)	(0.459)	(0.010)
lifesat	0.854***	−0.391**	0.290	−3.161***	−268.083***	1.141***	−0.174	0.857***	−2.559***	−121.174***
	(0.165)	(0.177)	(0.251)	(0.076)	(0.022)	(0.096)	(0.271)	(0.144)	(0.106)	(0.031)
smknow						−41.092***	0.985**	0.902**	−0.313	17.088***
						(0.009)	(0.482)	(0.435)	(0.206)	(0.026)
symptcvd	1.018**	0.068	0.780	2.376***	−123.391***	1.077**	0.400	0.517	2.998***	−40.442***
	(0.422)	(0.427)	(0.511)	(0.101)	(0.026)	(0.503)	(0.384)	(0.443)	(0.153)	(0.035)
time_7	0.811***	0.453***	−1.027***	0.894***	143.979***	1.578***	1.052***	−1.299***	1.232***	145.855***
	(0.163)	(0.139)	(0.149)	(0.036)	(0.006)	(0.063)	(0.245)	(0.123)	(0.068)	(0.007)
time_15	−379.581	0.942***	0.005	−544.384***	168.523***	−51.690***	1.587***	0.076	−73.698	113.428***
		(0.086)	(0.079)	(0.000)	(0.002)	(0.000)	(0.105)	(0.079)		(0.003)
Constant	−0.155***	−18.979***	−0.157***	−2.500***	2,129.842***	68.083***	−25.020***	−56.761***	−11.661***	191.921***
	(0.011)	(0.003)	(0.016)	(0.005)	(0.008)	(0.009)	(0.004)	(0.015)	(0.007)	(0.010)
Observations	122	122	122	122	122	122	122	122	122	122
Pseudo R^2	0.467	0.467	0.467	0.467	0.467	0.552	0.552	0.552	0.552	0.552
Akaike Inf. Crit.	267.916	267.916	267.916	267.916	267.916	302.754	302.754	302.754	302.754	302.754

Note: *p<0.1; **p<0.05; ***p<0.01

Goodness of fit test

From the results of Hosmer-Lemeshow tests we can see that for all models there are small values of test statistics with large p-values. This means that we fail to reject the null hypotheses, which indicates good fits.

Hosmer and Lemeshow test (binary model) results

Logistic Regression - the cancer sample		**Logistic Regression - the non-cancer sample**	
Goodness of fit test	H0: Evidence of good fit	Goodness of fit test	H0: Evidence of good fit
X-squared	8.1924	X-squared	3.7362
df	8	df	8
p-value	0.4149	p-value	0.8801
Logistic Regression - the whole sample with *cancer*		**Logistic Regression - the whole sample with time intervals**	
Goodness of fit test	H0: Evidence of good fit	Goodness of fit test	H0: Evidence of good fit
X-squared	3.786	X-squared	2.9433
df	8	df	8
p-value	0.8759	p-value	0.9379
Logistic Regression - (1) cancer t $\leq$ 7		**Logistic Regression - (2) cancer 7 $<$ t $\leq$ 15**	
Goodness of fit test	H0: Evidence of good fit	Goodness of fit test	H0: Evidence of good fit
X-squared	6.7981	X-squared	4.2937
df	8	df	8
p-value	0.5586	p-value	0.8297

Hosmer and Lemeshow test (multinomial model) results

Multinomial Logistic Regression - the cancer sample		**Multinomial Logistic Regression - the non-cancer sample**	
Goodness of fit test	H0: Evidence of good fit	Goodness of fit test	H0: Evidence of good fit
X-squared	36.546	X-squared	27.791
df	40	df	40
p-value	0.6266	p-value	0.9277
Multinomial Logistic Regression - the whole sample with *cancer*		**Multinomial Logistic Regression - the whole sample with time intervals**	
Goodness of fit test	H0: Evidence of good fit	Goodness of fit test	H0: Evidence of good fit
X-squared	32.621	X-squared	28.957
df	40	df	40
p-value	0.7899	p-value	0.9023
Multinomial Logistic Regression - (1) cancer $\leq$ 7		**Multinomial Logistic Regression - (2) cancer 7 $<$ t $\leq$ 15**	
Goodness of fit test	H0: Evidence of good fit	Goodness of fit test	H0: Evidence of good fit
X-squared	18.85	X-squared	31.433
df	40	df	40
p-value	0.9982	p-value	0.8315

Multicollinearity test

Based on our research, there is an empirical justification for including the square of the variable *age*. Depression increases with the age as people become more experienced and went through several difficult obstacles, but at some point, the severity of depression does not increase and starts to decrease (older people have a different perspective on life). So, the relationship between depression and age is inverted U-shaped. Since we can see that the only worrying VIF values are between *age* and *agesq*, where one variable is a deterministic non-linear function of the other, we do not need to worry about multicollinearity.

Variation Inflation Factor (VIF) results

	Multicollinearity test (binary model)					VIF > 5: problematic amount of collinearity				
	age	female	hhsize18+	educ	hlthstat	lifesat	symptcvd	time_7	time_15	cancer
Cancer sample	1.103011	1.060336	1.085699	1.095249	1.315158	1.226657	1.037032	1.276200	1.246127	
Non-cancer sample	1.106013	1.005159	1.032303	1.125431	1.320582	1.162643	1.019511			
Whole sample (*cancer*)	1.228477	1.006140	1.037877	1.117426	1.339904	1.168533	1.021481			1.131833
Whole sample (time int.)	1.167035	1.005556	1.037882	1.117761	1.340292	1.169155	1.021533	1.044586	1.026511	
(1) cancer ≤ 7	1.104218	1.065264	1.105995	1.092346	1.368557	1.227123	1.051098			
(2) cancer $7 < t \leq 15$	1.039436	1.046436	1.045733	1.125408	1.284159	1.219098				

Variation Inflation Factor (VIF) results

	Multicollinearity test (multinomial model)						VIF > 5: problematic amount of collinearity				
	age	agesq	female	south	hhsize18+	hlthstat	lifesat	symptcvd	time_7	time_15	cancer
Cancer sample	94.4477	94.4114	1.0382	1.0428	1.0928	1.3678	1.3715	1.0523	1.3233	1.2986	
Non-cancer sample	35.1778	35.0294	1.0137	1.0078	1.0262	1.2573	1.1879	1.0255			
Whole sample (*cancer*)	36.8386	37.0759	1.0117	1.0085	1.0299	1.2854	1.2004	1.0264			1.1427
Whole sample (time int.)	36.7912	36.8629	1.0111	1.0084	1.0303	1.2872	1.2025	1.0255	1.0603	1.0286	
(1) cancer ≤ 7	79.4665	80.3105	1.0312	1.1686	1.1728	1.2703	1.2882	1.0987			
(2) cancer $7 < t \leq 15$	115.2468	114.6936	1.0443	1.0640	1.0573	1.2053	1.2440				