The effect of socioeconomic status on overall mortality after diagnosed with colorectal cancer

A survival analysis

Hege Wermundsen Skappel

Supervisors:
Tron Anders Moger
Eline Aas

Master thesis submitted as a part of the Master of Philosophy Degree in Health Economics, Policy and Management

UNIVERSITY OF OSLO
FACULTY OF MEDICINE
DEPARTMENT OF HEALTH MANAGEMENT AND HEALTH ECONOMICS

May 2018
The effect of socioeconomic status on overall mortality after diagnosed with colorectal cancer

* A survival analysis
The association between socioeconomic status and colorectal cancer mortality

Hege Wermundsen Skappel

http://www.duo.uio.no/

Trykk: Reprosentralen, Universitetet i Oslo
Abstract

**Title:** The effect of socioeconomic status on overall mortality after diagnosed with colorectal cancer.

**Background:** Social inequality in health persists even in egalitarian and highly developed countries. There is a social gradient found in incidence and mortality from several non-communicable diseases, cancer being one of them. Colorectal cancer is the third most common form of cancer and one of the most common causes of cancer death worldwide. Survival analysis is an appropriate way of analysing time and cause to such an event.

**Objective:** To estimate the effects of socioeconomic status on overall mortality after diagnosed with colorectal cancer in a Norwegian cohort, and to compare results from the Cox proportional hazard and the Weibull models to see how they differ.

**Method:** An individual level dataset conducted from a randomized controlled trial for colorectal cancer, Norwegian colorectal cancer prevention (NORCCAP), was used. 1298 individuals diagnosed with colorectal cancer were included. The Cox Proportional Hazard model and the Weibull model were applied to calculate hazard ratios for mortality risk by level of education, income, family status and place of residence, adjusted for age, gender and period. The multivariable models were compared regarding their predictive power using Harrell’s C-statistic and by survival curves.

**Results:** An education level of 10-12 years and an income of 200 000 – 300 000 NOK and 400 000 – 600 000 had a positive impact on mortality compared to a low level of education and lower income. Family status and place of residence did not have a significant effect on mortality. The Cox Proportional Hazard model and the Weibull Proportional Hazard model gave almost identical hazard ratios and p-values, but the Cox model gave a slightly higher predictive power than the Weibull model. The survival curves gave quite different results even though the multivariable analyses gave virtually identical hazard ratios and p-values.

**Conclusion:** Socioeconomic status seems to have an impact on mortality after diagnosed with colorectal cancer given that education and income had a significant effect on mortality. Similar hazard ratios from the Cox model and the Weibull model nevertheless gave some deviations in estimated survival over time in these data.
Acknowledgment

I would like to thank my supervisors Tron Anders Moger and Eline Aas. I am beyond grateful for all the help you have provided in this thesis. Your commitment has both inspired and motivated, and without your guidance, this thesis would not have been possible to finish.

I would also like to thank my family and friends for all support and for always believing in me.

Thanks to my fellow students, you have prevented numerous meltdowns.
# Table of Contents

1 **Introduction** .................................................................................................................. 1

2 **Background** ...................................................................................................................... 3
   2.1 Social inequality in health ............................................................................................... 3
   2.2 Possible explanations for social inequality in health ....................................................... 4
   2.3 Social inequality and cancer mortality ............................................................................. 5
   2.4 Colorectal cancer ........................................................................................................... 6
   2.5 Colorectal cancer risk factors and prevention ................................................................. 8
   2.6 Socioeconomic status and colorectal cancer ................................................................. 10
   2.7 The measurement of socioeconomic status ................................................................. 12
   2.8 Weibull and Cox proportional hazard models ............................................................... 14

3 **Methods** .......................................................................................................................... 16
   3.1 Statistical methodology ................................................................................................. 16
      3.1.1 Survival analysis ..................................................................................................... 16
      3.1.2 Censoring .............................................................................................................. 16
      3.1.3 Survival function and hazard rate .......................................................................... 17
      3.1.4 Parametric survival analysis .................................................................................. 19
      3.1.5 Semiparametric survival analysis ......................................................................... 21
      3.1.6 Predicting survival and applications to time-dependent transitions probabilities .... 22
      3.1.7 Harrell’s C-statistic ............................................................................................... 22
   3.2 Data .................................................................................................................................. 24
      3.2.1 The Colorectal Cancer dataset .............................................................................. 24
      3.2.2 Outcome variable .................................................................................................. 24
      3.2.3 Main covariates: SES variables .............................................................................. 24
      3.2.4 Background variables as potential confounders .................................................... 25
   3.3 Statistical analysis ........................................................................................................... 26

4 **Results** .............................................................................................................................. 28
   4.1 Descriptive statistic of data ............................................................................................ 28
   4.2 Weibull and Cox PH regression – univariable analysis ................................................. 30
   4.3 Weibull and Cox PH regression – multivariable analysis ............................................ 32
   4.4 Model validation – Schoenfeld residual and Harrell’s C-statistic ................................ 34
   4.5 Survival curves .............................................................................................................. 35

5 **Discussion** ......................................................................................................................... 39
   5.1 Main findings .................................................................................................................. 39
   5.2 Previous research .......................................................................................................... 41
5.3 Interpretation of results................................................................. 42
5.4 Limitations.................................................................................... 45

6 Conclusion....................................................................................... 47

References ......................................................................................... 48

Appendices ......................................................................................... 53

A.1 Schoenfeld Residuals................................................................. 53
A.2 Committee Decision and Right to Appeal from REK .................... 54
LIST OF FIGURES

Figure 1. The different time-dependent hazard rates from the Weibull model ................. 20
Figure 2. Survival curves from Weibull and Cox model for age .................................. 35
Figure 3. Survival curves from Weibull and Cox model for gender ............................... 36
Figure 4. Survival curves from Weibull and Cox model for education ........................... 37
Figure 5. Survival curves from Weibull and Cox model for income ............................. 38

LIST OF TABLES

Table 1. Descriptive characteristics of the study sample .............................................. 29
Table 2. Comparison of Weibull and Cox univariable regression .................................. 31
Table 3. Comparison of Weibull and Cox multivariable regression ............................... 33
Table A1. Test of proportionality of hazards on Schoenfeld residuals ......................... 53
Social inequality in health persists even in egalitarian and highly developed countries. There is a social gradient found in incidence and mortality from several non-communicable diseases, cancer being one of them (World Health Organisation, 2017). Social inequality in health can be perceived as a problem for several reasons. It is a problem of justice, because people in lower social classes are deprived of life chances and freedom, and equality in the right to health is linked to the inherent dignity of human. Social inequality is also a living condition problem, as failing health is an important, if not the most important factor behind social exclusion in Norway. Poor health hamper an active, creative, productive and socially participating life (Dahl, Bergsli, & van der Wel, 2014).

Cancer is one of the leading causes of morbidity and mortality globally and nationally (World Health Organisation, 2017). Colorectal cancer (CRC) is the third most common form of cancer and one of the most common causes of cancer death worldwide (World Health Organisation, 2018), and an appropriate way of analysing time until and cause of such an event is survival analysis (Briggs, Claxton, & Sculpher, 2006; Cleves, Gutierrez, Gould, & Marchenko, 2008). Parametric methods, for example, using the Weibull distribution, have been widely used in fitting survival data. The Cox semi-parametric method is also widely used for modelling such data. These methods account for the relationship between survival and other variables such as age, gender, and family history of cancer. In medical sciences, researchers tend to use the Cox model instead of parametric methods to analyse survival data. This might be because there are fewer assumptions in the use of Cox semi-parametric method (Carroll, 2003). However, some studies argue that parametric methods can provide more accurate estimates in some circumstances (Pourhoseingholi, et al., 2007; Bradley, 1977). Both are used in economic evaluation, for instance, to estimate time-dependent transition probabilities in Markov models, with the Weibull being the most common.

The primary focus of this master thesis is to estimate the effects of socioeconomic status, such as education, income, family status, and place of residence on overall mortality after diagnosed with CRC in a Norwegian cohort. The secondary focus is to compare results from the Cox proportional hazard and the Weibull models to see how they differ, and briefly discuss the implications of choosing one over the other if used in economic evaluation.
To the best of my knowledge, there have not been any studies on socioeconomic status’ impact on mortality after diagnosed with colorectal cancer with data from the 21st century in Norway. In addition, to the best of my knowledge, no one has compared the Cox model and the Weibull model regarding the implications of choosing one over the other if used in economic evaluation or for mortality of patients with colorectal cancer.

The next chapter provides background information on social inequalities in health and cancer mortality. Furthermore, the epidemiology of colorectal cancer and previous research of socioeconomic status and colorectal cancer is presented. Towards the end of the chapter, different measurements of socioeconomic status and a short introduction on survival analysis is given. In chapter 3 the theoretical framework of survival analysis is described, as well as the data used in the analysis. The descriptive data and the results from the analyses will be presented in chapter 4. In Chapter 6, findings of the study will be interpreted, discussed, and compared to previous research. Limitations of the study will also be discussed. Chapter 7 will provide a brief summary of the main findings in a conclusion.
2 BACKGROUND

2.1 SOCIAL INEQUALITY IN HEALTH

Social inequalities in health can be defined as inequalities that are unnecessary, preventable, and unfair (Dahlgren & Whitehead, 2009, s. 14). The theory of social inequality and its effects on health is extensive. The links between health, wealth, and education have been studied in several populations, with the general finding being that better health and longer life is associated with higher socioeconomic status, compared to those with a low level of education and lower income (Adams, Hurd, McFadden, Merrill, & Ribeiro, 2003). This is called the social gradient in health, and means that pathogenic factors apply to all individuals, but affect them in different degrees according to their place in the social hierarchy. The higher the socioeconomic status, the better the health, and vice versa (Dahl, Bergsli, & van der Wel, 2014).

The social gradient in health indicates that inequality is found throughout all the layers of socioeconomic status in society, and not only in those at the bottom of the social ladder. However, the difference in inequality is greater between those with low and middle income compared to those with a high or very high income, meaning that the gradient is steepest at the bottom (Dahl, Bergsli, & van der Wel, 2014).

There are increasing health inequalities in all countries in Europe, and removing these inequalities is now one of the most important challenges for the public health (Dahlgren & Whitehead, 2009). Social inequalities in health are also found in Norway, even though Norway is a wealthy country with consistently good health, high life expectancy and a public health system that is almost free of charge at the point access (Dahlgren & Whitehead, 2009). With regard to income, nor Denmark, Sweden or Finland have detected significant associations between income inequality and health, while there has been found in Norway where higher levels of regional income inequality were associated with higher mortality (Dahl, Elstad, Hofoss, & Martin-Mollard, 2006).
2.2 Possible explanations for social inequality in health

Several theories try to explain the mechanisms behind social inequality in health. The Norwegian Directorate for Health and Social Affairs invited one of Norway's foremost researchers in the field, Jon Ivar Elstad, at the Norwegian Institute for Growth, Welfare and Aging, to write an overview of research on the causes of social inequalities in health (Elstad, 2005). According to this overview, there are several possible explanations for the association between health and social inequalities. However, which theory that explain most of the social inequality found in health, has no consensus.

The mobility explanation is based on health status determining social class, and not the other way around. The explanation says that those with good health tend to rise in the socioeconomic hierarchy because they have the strength and power to do so. Conversely, health problems could weaken the opportunities for education, good work or promotion. In that case, people with good health will be able to finish higher levels of education and succeed in demanding well-paid jobs, whilst people with poor health might lose employment and income and end up in a lower social class than they originally were (Elstad, 2005).

The materialist approach explains health inequalities through differences in socioeconomic positions. In this approach, social class is a determinant for health because social class is linked to differential exposures to the material world, which can be either advantageous or not to health. If good material conditions such as clean water and healthy food, sanitation, and, most important, equal access to health care, were available for all, social inequality in health could be reduced (Elstad, 2005).

Another approach which stresses that culture determines, or frames behavioural choices, is the cultural-behavioural approach. The approach suggests that health-damaging behaviour by individuals creates social inequalities in health. This includes decisions affecting health, such as, engaging in higher risk lifestyles with, for example, drinking, smoking, or an unhealthy diet. Almost all these health-risk behaviours are more prevalent among lower SES groups. (Elstad, 2005).

Both the materialistic and the cultural-behavioural approach essentially say that human beings literally do not act with consciousness, willingness and emotions to external and internal influences. An alternative model is the psycho-social perspective that focuses on individuals
in their social environment (Elstad, 2005). This perspective argues that individuals with lower socioeconomic status suffer from poorer health because of experiencing more negative life events, less autonomy at work, less social support, more stress and less job security (Sundmacher, Scheller-Kreinsen, & Busse, 2011).

The life-course perspective adds a dimension of time and explains inequalities in health as a result of biological and social circumstances early in life. Health is thus also determined by past living conditions and events and is no longer just a result of current conditions and lifestyle choices (Elstad, 2005).

2.3 SOCIAL INEQUALITY AND CANCER MORTALITY

Cancer and cardiovascular disease are the most dominant causes of death in industrialized countries, with the latter being most common. In Norway, death rates for cardiovascular diseases have declined over the last few decades, while there have been rather weak improvements in survival from many of the most common cancer types and the incidence have increased considerably. There are much smaller social differences in total cancer mortality than in cardiovascular mortality (Kravdal, 2000), however, a social gradient is found in both incidence and mortality for several cancer sites (Dahl, Bergsli, & van der Wel, 2014).

A consistent social gradient for men across populations, with a higher risk for the more disadvantaged, are found for respiratory cancers and cancers in the mouth, stomach, and pharynx. A negative social gradient for women is found for cervical cancer and cancers of the oesophagus, stomach and, less consistently, liver, but not for respiratory sites. A positive social gradient in men are found for colon cancer, while for women, a positive social gradient is well established for breast cancer (Faggiano, Partanen, Kogevinas, & Boffetta, 1997). However, the inequality varies between countries in both incidence and mortality from cancer (World Health Organisation, 2018).

Kravdal (2000) has addressed the social inequalities in cancer survival. The study wanted to investigate if the increase in mortality among cancer patients, compared to similar non-cancerous people, is different for diverse social classes. The pattern Kravdal (2000) finds in his study shows that education improves survival, and within groups with the same education, income and occupation status have a favourable impact. People with primary school have a 15% higher increase in mortality compared to people with education beyond the upper
secondary school. This yields for both women and men, controlled for age, period, histology and stage at the time of diagnosis. For those with even higher levels of education, the survival are even better.

Kravdal (2000) presents three factors that are crucial for cancer survival as suggested explanations for survival inequalities: biologic characteristics of the tumour, characteristics of the host, and treatment. He argues that education and income may influence survival through these factors as there is, for example, much empirical evidence for a social effect operating through stage at diagnosis. One author has found that half of the relationship between socioeconomic status and survival for colorectal cancer disappears when it is controlled for stage (Auvinen, 1992). The treatment itself is another possible explaining mechanism of the social pattern Kravdal (2000) sees, and he argues that the results from the study can provide a basis for testing whether the treatment given to those in lower social classes in society is worse than the treatment to people of higher social classes, and that this has consequences for survival or the standard of living as a patient.

2.4 COLORECTAL CANCER

Colorectal cancer (CRC) is tumours of the colon and rectum. Most cases of CRC are sporadic, meaning there is no known hereditary (genetic) components, and develop slowly over several years through adenomatous polyps (Brenner, Kloor, & Pox, 2014). Changes in bowel habits, blood in the stool, and anaemia are cardinal symptoms and sings of CRC. In later stages, fatigue, anorexia, weight loss, pain, jaundice, and other signs and symptoms of locally advanced and metastatic disease occur. CRC is traditionally diagnosed by sigmoidoscopy and colonoscopy with biopsy. There are several ways to treat colorectal cancer depending on the cancer stage and where the tumour is located. The main treatment is surgery, however, laparoscopic technique (keyhole surgery) seems to give equally good results as open surgery but provides shorter injections and less pain. In addition, chemotherapy and radiation therapy are used (Potter & Hunter, 2008).

Approximately 1.4 million new cases of colorectal cancer and almost 700,000 deaths occurred worldwide in 2012 (Arnold, Sierra, Laversanne, & Soerjomataram, 2017). In Norway, colorectal cancer is the second most common cancer diagnosis when looking at both genders, after prostate cancer for men and breast cancer for women. More than 4000 new cases were
registered in 2014, and there has been a significant increase in the incidence rates of CRC for both sexes in the past 60 years. The increase has been stronger in Norway than the other Nordic countries, and the rate in Norway is also higher compared with other European countries (The Norwegian Directorate of Health, 2017a). The global burden of CRC is expected to increase to more than 2.2 million new cases and 1.1 million deaths by 2030, given the temporal profiles and demographic projections (Arnold, Sierra, Laversanne, & Soerjomataram, 2017). The number of cases is also expected to continue to increase in Norway. Projections estimate that there will be 5500 new cases annually of CRC in 2024-28. The aging population is the main reason for this assumed increase, (Engholm, et al., 2017) since CRC is uncommon for individuals younger than 50 years with a mean age of 70 at diagnosis. However, globally the incidence is increasing in young people and declining in elderly (Holt, Kozuch, & Mewar, 2009).

It has been a steady increase in survival the last 40 years. The 5-year relative survival rate (the proportion surviving at least 5 years, adjusted for survival in the normal population) for CRC is now about 60% for colon cancer and 66% for rectal cancer in Norway, compared to under 30% in the early 1970s. The relative survival rate falls rapidly in the first few years and levels off after about 5 years (The Norwegian Directorate of Health, 2017a). However, CRC is still one of the leading causes of death by cancer in Norway, only second to lung cancer (Cancer Registry of Norway, 2018).

Comparing global patterns and trends in colorectal cancer incidence and mortality show large differences. In rapidly transitioning countries with medium and high human development index (HDI) including those in the Baltics, China, Russia, Brazil it is observed increases in both incidence and mortality in the most recent decade. In very high HDI countries, including Canada, UK, Denmark, Norway and Singapore, increases in incidence with concomitant decreases in mortality is reported. In a number of the highest HDI-indexed countries, including the United States, Japan and France, decreases in both incidence and mortality is observed (Arnold, Sierra, Laversanne, & Soerjomataram, 2017).
2.5 COLORECTAL CANCER RISK FACTORS AND PREVENTION

Colorectal cancer has at least two features that contribute to its preventability. First, tumours of the colon and rectum have several established or suspected risk factors that are quantitatively important as well as potentially modifiable. Second, the existence of adenomatous polyps and the accessibility of the colonic mucosa to endoscopy (Potter & Hunter, 2008).

Inflammatory bowel disease (IBD), ulcerative colitis and Crohn’s disease are established risk factors, and up to 5% of all CRC in patients under the age of 50 occur in those with IBD (Potter & Hunter, 2008; Ullman & Itzkowitz, 2011). Cigarette smoking have originally not been associated with an elevated risk of CRC, because most of the early studies examining the relationship between smoking and CRC did not find an association. In contrast, a systematic review and meta-analysis found that the majority of studies with a follow-up after 1970 found a correlation (Liang, Chen, & Giovannucci, 2009). They argue that the reason for this might be a 35 to 40-year lag time between exposure and disease, which would not be captured by earlier studies and studies with shorter follow-up time. In other words, the association between smoking and CRC is controversial, but worth mentioning. For alcohol consumption, most studies have shown an increased risk for CRC. An overall and dose–response meta-analysis of published studies supports the evidence for a causal relation between high intakes of alcohol and increased risk for colorectal cancer and provide additional evidence of an association for moderate intakes of alcohol (Fedirko, et al., 2011). There is also increasing evidence that indicates that obesity may be associated with the risk of CRC. A Systematic Review of Prospective Studies for obesity and risk of CRC concludes that both of general and central obesity were positively associated with the risk of CRC in this meta-analysis (Ma, et al., 2013). Other suspected risk factors are high consumption of red and processed meat and diabetes, while physical activity (Boyle, Kegele, Bull, Heyworth, & Fritschi, 2012), aspirin/nonsteroidal anti-inflammatory drug use (Friis, Riis, Erichsen, Baron, & Sorensen, 2015; Drew, Cao, & Chan, 2016), and endoscopic screening, with removal of precancerous lesions (Elmunzer, et al., 2012), have a protective effect.

Endoscopic screening can be a protective factor in reducing incidence and mortality of CRC, as it can detect adenomatous polyps (Manser & Bauerfeind, 2014). Adenomatous polyps are known to be precursors to CRC. The risk of getting CRC can increase up to 30% if one has a family history/genetic predisposition (The Norwegian Directorate of Health, 2017a). The goal
of screening for colorectal cancer is to reduce the incidence and mortality of the disease in the population by finding and eliminating precursors for cancer and finding cancer in an early symptom-free phase, giving fewer cancer cases, better prognosis and reduced mortality. Many countries have introduced screening for colorectal cancer to their residents, but Norway does not currently have a national screening program. The Norwegian Directorate of Health has investigated how a national screening program for colorectal cancer should be organized in Norway, so hopefully, it will be introduced soon. In the period 2012-2018, it is an ongoing pilot program for colorectal cancer screening in the Southern and Eastern Norway Regional Health Authority. In the pilot, 140,000 women and men aged 50-74 from Østfold, Akershus and Buskerud will be offered screening for colorectal cancer. The program is organised as a randomised study with two offers for intestinal screening. Participants are randomised to either receive an immunochemical test for invisible blood in the stool (FIT) or a sigmoidoscopy examination (lower intestine intestinal examination) (The Norwegian Directorate of Health, 2017b).

In addition, a study on the effect of flexible sigmoidoscopy screening on colorectal cancer incidence and mortality in a Norwegian cohort (Holme, et al., 2014) found that flexible sigmoidoscopy screening reduced colorectal cancer incidence by 27% in men and 13% in women and reduced colorectal cancer mortality by 42% in men and 9% in women with a follow-up time of 10 years. Another study, using the same Norwegian cohort but with longer follow-up time (Holme, et al., 2018) found that the 15-year risks for CRC in women were 1.86% in the screening group and 2.05% in the control group, and that the 15-year risks for CRC death were 0.60% in the screening group and 0.59% in the control group. For men, the 15-year risks for CRC were 1.72% in the screening group and 2.50% in the control group. The 15-year risk for CRC death were 0.49% in the screening group and 0.81% in the control group. These two studies found an effect of sigmoidoscopy screening, but the long-term effect is stronger among men, but has little or no effect among women. In comparison, a UK Flexible Sigmoidoscopy Screening trial (Atkin, et al., 2017) found that once-only sigmoidoscopy screening was effective in both women and men after 17 years of follow-up, although less so in women, which is consistent with the findings from the Norwegian studies.
2.6 Socioeconomic Status and Colorectal Cancer

Colorectal cancer is increasing in number and has become a significant burden of disease in both developed and developing countries (International Agency for Research on Cancer, 2012). Evidence of increasing inequality in incidence and mortality from CRC by socioeconomic status has emerged (Egeberg, Halkjaer, Rottmann, Hansen, & Holten, 2008). However, existing evidence is inconclusive (Leufkens, et al., 2012). A systematic review of the impact of socioeconomic status on incidence, mortality, and survival of colorectal cancer reports that most studies found a social gradient but between Europe and North America, the findings differ. In the United States, lower SES was associated with an increased incidence of CRC, while an inverse association was observed in Europe. However, the findings do not differ when considering mortality. The same studies indicate a negative social gradient for CRC mortality in both Europe and the United States (Manser & Bauerfeind, 2014). In addition, a study on global patterns and trends in colorectal cancer incidence and mortality (Arnold, Sierra, Laversanne, & Soerjomataram, 2017) reports that CRC has replaced infection-related cancers as the second most common cancer type in several middle-income countries. The study further argue that this highlights the major challenge of CRC control in countries undergoing significant socioeconomic transition, and the importance of continued efforts to monitor trends in CRC incidence and mortality worldwide. The number of patients with CRC will continue to increase in future decades beyond those already projected because of an aging population, if targeted resource-dependent actions are not carried out.

Several of the lifestyle and behavioural risk factors for CRC can be related to SES. A study on socioeconomic factors, health behaviours, and mortality from 1998 found that lower levels of education and income were associated with a significantly higher prevalence of health-risk behaviours, including smoking, being overweight, and physical inactivity (Lantz, et al., 1998). The results also show that lower income lead to a significant increase in mortality risk, yet the influence of major health-risk behaviours explained only a modest proportion of this relationship. Other studies also indicate an inverse relationship between adverse health behaviour and SES (Institute of Medicine, 2001), potentially explaining the negative social gradient for CRC mortality. Another possible explanation for the negative social gradient for CRC mortality might be the fact that a low-SES population less frequently participates in cancer screening programs, resulting that CRC is not diagnosed at an early stage. Furthermore, access to health services is, in general, worse for people with a low SES (Manser & Bauerfeind, 2014). These explanations may be appropriate for the US, Canada and
some European countries, but for Norway that has not yet implemented a national screening program for colorectal cancer and at the same time have public health care with equal access to all, these explanations should not hold. One study from Norway, which looked at trends in educational inequalities in mortality from 1971 to 2002 (Elstad, Torstensrud, Lyngstad, & Kravdal, 2011), found a weak educational gradient for colorectal cancer. The educational differences in CRC deaths were quite small across the entire period, with a tendency to lower death rates among higher educated. The study argues that the full explanation of the persistence of educational patterns in specific cancer mortality is certainly complex since cancer mortality is a function of incidence influenced by exposures to risk factors and perhaps some inherited factors, and survival that depends, among other things, on health service factors.

In a paper published in 2017 socioeconomic and racial/ethnic disparities in cancer mortality, incidence, and survival in the United States between 1950 and 2014 were analysed (Singh & Jemal, 2017). They found that socioeconomic trends in US colorectal cancer mortality changed dramatically between 1950 and 2013, where colorectal cancer mortality increased a little per year in the least-affluent group, whereas it fell consistently in the higher SES groups. Before the 1990s the least-affluent group had much lower mortality rates than the most-affluent counterparts. The most-affluent group had a 50% higher colorectal mortality rate than the least-affluent in the 1950s, and in 2009–2013, there was an inverse SES gradient, with those in the least-affluent group having a 30% higher colorectal cancer mortality rate than their most-affluent counterparts. At the turn of the 21st century, all groups had a decreasing mortality rate that continued to 2009-2013.

To the best of my knowledge, there have not been any studies on socioeconomic status and mortality after diagnosed with colorectal cancer with data from the 21st century in Norway.
2.7 THE MEASUREMENT OF SOCIOECONOMIC STATUS

Socioeconomic status can be used as a convenient summary term for various socioeconomic factors. The concept designates the position of an individual within a given social structure (Manser & Bauerfeind, 2014). This classification is based on socioeconomic factors such as education, income, assets, housing and occupation. However, there is little evidence as to which socioeconomic factors are of most importance (Pearce, 1997). In addition, it has been argued that socioeconomic status cannot be directly measured as it is a latent variable like, for example, mood or well-being (Oakes & Rossi, 2003). In other words, SES does not have a universal measurement. However, different factors as a measurement of SES can be included as individual variables or as an index. An index combines numerical information from several indicators for the purpose of giving a better or an alternative target to a phenomenon than what the indicators can provide individually (Alver, Hesselberg, & Lyshol, 2009).

Education is a frequently used measure of socioeconomic status, because it is relatively easy to measure due to capturing a large proportion of the population, and because education has a strong influence on later occupation and income. Completed years at school usually measure education, and in analyses, it can be included as a number of years or made into categories based on the school system (Dahl, Bergsli, & van der Wel, 2014). Education is usually fixed early in life, and therefore it is appropriate to consider education when investigating SES throughout life because SES will most likely not change later in life (Berkman & Macintyre, 1997). Furthermore, education has a positive impact on a person's ability to perceive and benefit from relevant health information, and communication with health workers (Dahl, Bergsli, & van der Wel, 2014).

Income represents financial and material resources. It is easy to rank, and it can measure a unique aspect of the social class. Income varies within occupation and is sensitive to changes in life circumstances (Berkman & Macintyre, 1997). However, it has some weaknesses. Among other things, it is not easy to measure and is often mistaken. For example, some revenue types are not registered. In addition, income will vary in time depending on, for example, labour or it may be that a man or a woman has low income because the partner has a high income. Then the household income will be high, but the personal income low. Income is usually included in analyses by categories appropriate for the study population (Dahl, Bergsli, & van der Wel, 2014).
Education and income can be interrelated. These two measurements of SES can affect health both combined and separately as a socioeconomic position (Dahl, Bergsli, & van der Wel, 2014). Two other variables that might not be a direct SES measurement, but still can have an impact on a person’s SES, are family status and how centrally you live (Næss, Rognerud, & Strand, 2007).

Family status represents whether a person has a partner/being married and whether a person has a child in the family. Most studies point to the fact that those who are married have better mental and physical health than single people. Particularly divorced and widows have more depression and anxiety and higher mortality, and this may be since marriage gives social support, and that married have less financial problems than singles (Næss, Rognerud, & Strand, 2007). In addition, the presence of children may contribute to the relationship as the poor prognosis among single people has been suggested to be partly due to lack of support from or involvement with children (Kravdal, 2003). We may, therefore, assume that one-person families have lower SES than other families and that the family status can have an impact on mortality.

Considering how centrally one live, one can assume that there are more people with lower SES in less central areas than in more central areas, because of, for example, housing prices. Assuming further that the residence of individuals may have an impact on the use of health services because of access, since the availability of specialist healthcare in cities and more central areas is higher, there are more doctors and better health offers in the cities and access to general practitioners are more stable in central areas. In addition, this leads to less travel time to the general practitioners and specialists (The Norwegian Directorate of Health, 2014). Residence may thus affect the treatment and lead to a social difference on mortality after CRC because of the distance to a hospital. In addition, Norway has a very diverse physical environment. The country has densely populated urban areas as well as, for example, scattered inland settlements. The variation might create differences in economic activity and lifestyles (Kravdal, 2006).
2.8 Weibull and Cox proportional hazard models

Survival analysis concerns analysing the time to the occurrence of an event for different subgroups, and Weibull and Cox proportional hazard models are two different regression techniques in survival analysis. Survival analysis and these two approaches are fully described in the Method section. However, as a brief introduction, The Cox proportional hazards regression model is firmly established as the accepted, statistical norm in the analysis of survival data arising in clinical trials. The vast popularity of this model mostly comes from a wide experience in its application and the fact that it is distribution free. About the underlying distribution of survival time, no assumption must be made to make inferences about relative death rate. However, if the distribution of survival times can be well approximated, parametric analyses can be useful allowing a broader set of inferences to be made (Carroll, 2003). A number of studies have been conducted to compare parametric and semi-parametric survival regression methods (Orbe, Ferreira, & Núñez-Antón, 2002; Giolo, Krieger, Mansur, & Pereira, 2012; Pourhoseingholi, et al., 2007; Ghadimi, et al., 2011; Georgousopoulou, Pitsavos, Yannakoula, & Panagiotakos, 2015). However, to the best of my knowledge, no one has compared them regarding the implications of choosing one over the other if used in economic evaluation or for mortality of patients with colorectal cancer.

In the field of health economic evaluation, the Weibull model is the most commonly used. Economic evaluation in health care can be defined as the comparison of alternative options regarding their cost and consequences (Briggs, Claxton, & Sculpher, 2006).

Decision analytic models is widely used in cost-effectiveness analysis in the field of economic evaluation (Williams, Lewsey, Mackay, & Andrew Briggs, 2017). These are commonly Markov models for transitions between critical clinical states. Expected long-term health effects and cost of each intervention are estimated from the effect and cost of occupying each state, and the probability that an individual occupies each state over time. Modelling enables clinical evidence to be extrapolated over the long term, all available evidence on alternative interventions to be synthesized, and the need for further research to be assessed (Claxton, 1999). The modelling of survival is an important component, since health effects are usually measured by quality-adjusted life years (QALYs). The principal source of evidence on the effect of an intervention is usually a randomized controlled trial, where the individual patient-level data is used to build survival regression models for each of the transitions. Because economic evaluations attempt to reflect all differences in costs and outcomes, a lifetime
horizon is usually advocated. Since the study time is usually limited, the end point of interest is not observed for all individuals. Therefore, extrapolation is often required to predict the complete survival impact. Because of extrapolation, parametric survival models, most commonly the Weibull distribution, are used to extrapolate survival and construct transition probabilities for economic models (Briggs, Claxton, & Sculpher, 2006). Briggs also argues that the Cox model use to inform time-dependency in Markov models is limited, because the model does not specify how the risk of an event changes over time, for example, the functional form of the hazard function.
3 METHODS

3.1 STATISTICAL METHODOLOGY

The theory material in chapter 3.1 is adapted from the book “An Introduction to Survival Analysis Using Stata” by Mario Cleves 2008. Further references will be made subsequently.

3.1.1 Survival analysis

Survival analysis is used to examine survival data, i.e. to model the time it takes for an event to occur, called survival time. Usually this event is death, and an individual is observed until death occurs. This is the origin of the term "survival analysis". The event does not have to be death, and other examples of such events are: divorce, falling asleep, cancer diagnosis, time to progression, and time to relapse. Where one tries to understand a cause or establish risk factors, all of these can be subject to scientific interest. In classical survival analysis, one focuses on a single event for each individual and describing the occurrence of the event by means of survival functions and hazard rates and analysing the dependence on covariates by means of regression models. The survival function is defined as the probability that a subject survives beyond any given specified time without experiencing any event of interest, and the hazard rate is defined as the event rate at a specified time conditional on survival until or beyond that given time. In this master thesis, classical survival analysis is used.

3.1.2 Censoring

The study time is usually limited in survival analysis, which leads to some incomplete observations referred to as censoring, or censored survival times. There are different types of censoring, and the most common one is right-censoring, which occurs when the event of interest is not observed during the study time, or when the individual is removed from the study and cannot be monitored. To understand censoring better, let $T_i^*$ be the time until the event of interest occurs for individual $i$ and let $C_i$ be the right-censoring time, i.e. time at the end of the study. Observed lifetime for individual $i$ will then be $T_i = \min(T_i^*, C_i)$, which is time until either the event occurs, or the study ends, depending on what happens first. If one let $D_i$ be an indicator variable for censoring, one gets
The indicator variable will be 1 in cases where the event actually occurs (for example when the individual dies before the end of the study) and 0 when the event is censored.

### 3.1.3 Survival function and hazard rate

It is quite simple to analyse survival data, even though ordinary statistical methods cannot handle right-censored survival data. There are two basics that permeate the entire theory of survival analysis, namely survival function and hazard rate. You start with a set of individuals at time zero and wait for an event that might happen. The survival function, $S(t)$, often interesting to plot as a survival curve, gives the probability of surviving beyond time $t$. If the nonnegative random variable $T$ denotes the survival time, one may write more formally:

$$S(t) = P(T > t) \quad (2)$$

The survival function specifies the unconditional probability that the event of interest has not happened by time $t$. It always starts at 1, and it will often tend to zero as $t$ increases because over time more and more individuals will experience the event of interest.

The cumulative distribution function $F(t)$ gives the cumulative probability of failure up to time $t$:

$$F(t) = P(T \leq t) \quad (3)$$

The probability density function, $f(t)$, can easily be obtained from equation 2 as it can from equation 3 because $S(t) = P(T > t) = 1 - F(t)$, which gives:

$$f(t) = \frac{dF(t)}{dt} = \frac{d}{dt} \{1 - S(t)\} = -S'(t) \quad (4)$$

The hazard function, $h(t)$, is the instantaneous rate of failure, also known as the conditional failure rate. It is the limited probability that the failure event occurs in a given interval,
The hazard rate or function can vary from zero to infinity, where zero means no risk at all and infinity means the certainty of failure is that instant. The hazard rate can increase, decrease, remain constant, or take on more serpentine shapes over time. There is a close relationship between the amount of risk that has been accumulated upon to a certain time and the probability of survival past that time, and the hazard rate measures the rate at which risk is accumulated.

Given one of the four functions that describe the probability distribution of failure times, the other three are completely determined. In particular, one may derive all the other three functions from the hazard function (equation 5). To show this, it is convenient to define another function, the cumulative hazard function,

\[ H(t) = \int_0^t h(u) \, du \]  

and thus

\[ H(t) = \int_0^t f(u) \frac{1}{S(u)} \, du = -\int_0^t \frac{1}{S(u)} \left\{ \frac{d}{du} S(u) \right\} \, du = -\ln[S(t)] \]

The cumulative hazard function measures the total amount of risk that has been accumulated up to time \( t \). We can see from the equation the inverse relationship between the probability of survival and the accumulated risk, and can now conveniently write:

\[ S(t) = \exp\{-H(t)\} \]
\[ F(t) = 1 - \exp\{-H(t)\} \]
\[ f(t) = h(t)\exp\{-H(t)\} \]
The first of these formulae describes that the survival function (equation 2) equals the exponential of the negative integral of the hazard function between integration limits of 0 and t.

3.1.4 Parametric survival analysis

A parametric survival model is one in which the survival time is assumed to follow a known distribution. A distribution commonly used for survival time is the Weibull distribution. If the researcher is comfortable with the underlying distributional assumption, then parameters can be estimated that completely specify the survival and hazard functions. This simplicity and completeness are the main appeals of using a parametric approach.

The probability density function $f(t)$ can be expressed in terms of unknown parameters. Once a probability density function is specified for survival time, the corresponding survival and hazard functions can be determined as explained above.

The Weibull distributions have two parameters $\lambda$ and $p$, which gives the Weibull probability density function

$$f(t) = \lambda pt^{p-1}exp(-\lambda t^p)$$

which gives $h(t) = \lambda pt^{p-1}$ where $p > 0$, $S(t) = exp(-\lambda t^p)$ and $H(t) = \lambda t^p$

The parameter $\lambda$ is parameterized with regression parameters. The parameter $p$ is called a shape parameter and determines the shape of the hazard function. For example, if $p > 1$ then the hazard increases as time increases, and if $p < 1$ the hazard decreases over time. If $p = 1$ then the hazard is constant. The Weibull is suitable for modelling data that exhibit monotone hazard rates.
The Weibull proportional hazard (PH) model assumes a baseline hazard $h_0(t) = \lambda t^{p-1}$ where $\lambda$ is parameterized as $\exp(\beta_0)$ which gives $h_0(t) = pt^{p-1}\exp(\beta_0)$. $\beta_x$ is a vector of regression coefficients and $x_i$ is the predictor vector. Given a set of covariates, $x_i$, under the PH model,

$$h(t \mid x_i) = h_0(t)\exp(x_i\beta_x)$$

$$= pt^{p-1}\exp(\beta_0 + x_i\beta_x)$$

and this yields

$$H(t \mid x_i) = \exp(\beta_0 + x_i\beta_x) t^p$$

$$S(t \mid x_i) = \exp\{-\exp(\beta_0 + x_i\beta_x) t^p\}$$

By exponentiating the estimated intercept coefficient, the estimated scale parameter is obtained.

The hazard ratio (HR) is obtained by substituting a variable, for example, gender where “male” = 0 and “female” = 1 into the hazard functions (see equation 7). After cancelling we
obtain the familiar result $\exp(\beta_1)$. Note that this result depends on $p$ having the same value for $x_i = 0$ and $x_i = 1$, otherwise time ($t$) would not cancel in the expression for the HR (i.e., PH assumption not satisfied).

$$
HR = \frac{\exp(\beta_0 + \beta_1 \times 1)pt^{p-1}}{\exp(\beta_0 \times 0) pt^{p-1}} = \exp(\beta_1(1 - 0)) = \exp(\beta_1)
$$

(8)

The PH assumption requires that the HR is constant over time, or equivalently, that the hazard for one individual is proportional to the hazard for any other individual, where the proportionality constant is independent of time. It holds if $p$ does not vary over different levels of covariates. The PH assumption allows for the estimation of a hazard ratio enabling a comparison of rates among different populations.

### 3.1.5 Semiparametric survival analysis

The Cox proportional hazards regression model, by contrast, is not a fully parametric model. Rather it is a semi-parametric model because even if the regression parameters $\exp(\beta_0)$ are known, the distribution of the outcome remains unknown. The baseline survival (or hazard) function is not specified in a Cox model.

The Cox model is by far the most popular because of its elegance and computational feasibility, and it assumes that the covariates multiplicatively shift the baseline hazard function. The model asserts that the hazard rate for the $i$th subject in the data is

$$
h(t \mid x_i) = h_0(t)\exp(x_i \beta_x)
$$

(9)

where the regression coefficients, $\beta_x$, are to be estimated from the data.

The baseline hazard, $h_0(t)$, is given no particular parameterization and can in fact be left unestimated. In order to plot the survival function, one would combine the estimated regression coefficients with a non-parametric estimator for the cumulative hazard, typically the Nelson-Aalen estimator. The model makes no assumptions about the shape of the hazard over time, but it assumes that whatever the general shape, it is the same for everyone. One
subject's hazard is a multiplicative replica of another's hazard. To see how a change in one covariate affects the hazard, one can use the hazard ratio.

The hazard ratio (HR) is easily calculated by dividing the hazard for one individual with the hazard for a different individual. The two individuals being compared can be distinguished by their values for the set of predictors, that is, the $x_i$, for example "male" = 0 and "female" = 1. The change in hazard affected by a change in one covariate is then

$$HR = \frac{\exp(\beta_i \times 1)h_0(t)}{\exp(\beta_i \times 0)h_0(t)} = \exp(\beta_i(1 - 0)) = \exp(\beta_i)$$

(10)

This means that when all other covariates are constant, $\exp(\beta_i)$ is the time-independent hazard ratio for the change of one factor in the $i$th covariate.

3.1.6 Predicting survival and applications to time-dependent transitions probabilities

After estimating a model from either Weibull or Cox (and using Nelson-Aalen based estimates for the non-parametric baseline in Cox), one can create time-dependent transition probabilities using equation

$$1 - \frac{S(t)}{S(t-u)}$$

(11)

where $u$ define the length of the cycle. The different subgroups like gender etc. can be distinguished by the fact that the survival function will depend on the covariates.

3.1.7 Harrell’s C-statistic

Model validation is an important step in the model building process because it provides opportunities to assess the reliability of models before their deployment. Predictive accuracy measures the model's ability to predict future risk, and one important aspect is discrimination. Discrimination characterizes the model’s ability to correctly classify subjects for their actual outcomes (Guo, So, & Jang, 2017).

Akaike information criterion (AIC) and Bayesian information criterion (BIC) are criteria commonly used for model selection. The two measures estimate the quality of a model
relative to other models (Bradburn, Clark, Love, & Altman, 2003). AIC can also be used to compare models run with different parametric forms, with the lowest AIC indicative of the best fit. However, AIC cannot be used to compare parametric and semi-parametric models, since parametric models are based on observed event times and semi-parametric models are based on the order of event times (Columbia University Mailman School of Public Health, 2015).

There are a variety of methods to assess the discriminatory performance by a prediction model. The C-statistic (or concordance statistic) is the most commonly used discrimination measure in the context of regression with binary outcomes. The concept of concordance is that a subject who experiences a particular outcome has a higher predicted probability of that outcome than a subject who does not experience the outcome. The C-statistic can be calculated as the proportion of pairs of subjects whose observed and predicted outcomes agree (are concordant) among all possible pairs in which one subject experiences the outcome of interest and the other subject does not. The model can discriminate better the higher the C-statistic is between subjects who experience the outcome of interest and subjects who do not the higher c-statistics. The C-statistic ranges from 0 to 1, where a value of 0.5 means that the model is no better than predicting an outcome than random chance. A value below 0.5 indicates a poor model, while a value over 0.7 indicate a good model. The model predicts perfectly those group members who will experience a certain outcome and those who will not if it gets a value of 1 (Guo, So, & Jang, 2017). In the context of survival analysis, various C-statistics have been formulated to deal with right-censored data where Harrell's C is one of them.

The traditionally C-statistic is defined for binary outcomes and cannot be directly applied to survival data. By contrast, Harrell’s versions of the C-statistic are defined specifically for right-censored data. Harrell's method provides a simple approach by discarding the pairs that have become incomparable due to censoring. Although it is simple, the approach has the shortcoming that the estimates depend on the censoring variable (Guo, So, & Jang, 2017). The higher C-statistic, the better model also applies for Harrell’s C.
3.2 DATA

The data used in this study is from a randomized controlled trial for colorectal cancer, Norwegian colorectal cancer prevention (NORCCAP), which was carried out from 1999 to 2001. Two counties were represented: Telemark and Oslo. Telemark, with 165,855 inhabitants in 2003, has both urban and rural areas, while Oslo represent a typical urban area with 517,401 inhabitants in 2003. Each of the three years, 7000 women and men aged 50-64 were invited to participate, 3500 from each county. The participation rate was 65%. The rest of the population in the two counties in the same age group (n = 79,000) constituted the control group. In NORCCAP individuals were invited to a screening with once-only flexible sigmoidoscopy or flexible sigmoidoscopy in a combination with faecal occult blood tests.

3.2.1 The Colorectal Cancer dataset

On the basis of data from NORCCAP an individual level dataset was constructed from 1997 to 2011. Statistics Norway was commissioned to merge the various files. The personal identification number was used to merge the information from the screening with information from register data. Information on time of colorectal cancer diagnosis was found from the Cancer Registry of Norway. The dataset also included information on socio-economics and demographics factors from Statistics Norway: Age, gender, county of residence, native country, level of education, income and wealth, labour market participation, recipient of social insurance and pension. Time and cause of death were obtained from the Cause of Death register from Institute of Public Health. Only individuals diagnosed with colorectal cancer were included in the dataset used in this master thesis, which resulted in 1525 individuals. It was uncertain whether the descriptive variables were retrieved at the time of diagnosis for subjects diagnosed in 1998 or earlier. Therefore, only subjects diagnosed 1.1.1999 or later were included. This resulted in 1446 subject remaining in the dataset. I was interested in four socioeconomic variables and three background variables.

3.2.2 Outcome variable

The outcome of interest was time to death after being diagnosed with colorectal cancer, with a censoring indicator $D_i = 1$ if death was observed and $D_i = 0$ otherwise.

3.2.3 Main covariates: SES variables

Education

Education was chosen as one of the measures for SES, and was a categorical variable with low, middle and high education. “Low” comprises primary and lower secondary education,
which is mandatory in Norway and is called “grunnskole” (0-9 years). “Middle” comprises upper secondary education (10-12 years) and “High” comprises higher education (≥ 13 years). Low education was set as reference group.

**Income**

Income was chosen as the second measurement for SES, and was divided into six groups; ≤100 000, 100 000 – 200 000, 200 000 – 300 000, 300 000 – 400 000, 400 000 – 600 000, >600 000. The income is derived from the year they were diagnosed and is given in NOK. The variable was categorical, and the lowest level of income was set as reference group.

**Classification of centrality, residence**

Classification of centrality describes centrality according to the Standard for Local Government Classification (Statistics Norway, 2017). The classification represents different grouping of centrality index, where group 1 is the category with the most central municipalities (highest centrality index) and group 6 contains the least central (lowest centrality index). This is the standard classification of centrality from Statistic Norway, and the calculation of the index is based on travel time to jobs and service functions from all inhabited basic circuits. The variable is categorical, and group 1 was set as reference group.

**Family status**

Family status was a variable created by merging marital status and family type and was included as a categorical variable divided into married with children, married without children, one-person family and other families. Married also included cohabitating couples with joint children or no children. One-person families include unmarried, widower/widow, and divorced/separated and other families include single parents and persons in registered partnership. The family status represents the subject’s status at the time of diagnosis. Married with children was set as the reference group.

All the SES variables were included individually in the model, and not included as an index.

**3.2.4 Background variables as potential confounders**

**Age at diagnosis**

Age is a likely confounding factor since it is reasonable to assume that age affects survival per se and that the need for healthcare will increase with age. It may also be that certain age groups have particularly high healthcare needs. Age was included as a categorical variable, divided into four groups: 50-59, 60-69, 70-79. The youngest group was set as the reference group.
Gender

Gender was included as a control variable since there is reason to believe that the SES is different between the genders, especially in terms of income as it is well known that women earn less than men. Gender was included as a binomial variable, female/male, with male as the reference category.

Year of diagnosis

Year of diagnosis were included as a categorically variable with the groups; 1999-2001, 2002-2003, 2004-2005, 2006-2007, 2008-2009, 2010-2011, to see if better treatment may have existed and in that case had an impact on mortality. The latest year of diagnosis was set as the reference group.

3.3 Statistical analysis

A descriptive table of the variables selected was made, including total amount of subjects and how many deaths that occurred, to see how the sample was distributed over the variables included. There were 227 records with unknown information, and these records were excluded from the analysis so that there are no missing values in all the used variables for the remaining 1298 cases. The most recent years of diagnosis have a shorter follow-up time. Since it was only 50-64 year olds who were invited from 1999 to 2001, the oldest age group was not diagnosed in the two latest groups for year at diagnosis, and hence will also have shorter follow-up time. This will influence interpretation of the crude descriptive statistics shown in the table.

The variables that could have been included as continuous variables (age, education, classification of centrality, and year at diagnosis) were tried as both continuous and categorical in the models. None of the variables were significant when included as continuous variables, and thus I included them as categorical.

Survival analysis with the Weibull PH model and the Cox PH model was chosen as two appropriate models for analysis of the data and for comparing a parametric and semi-parametric model. Univariate analysis with both models was performed for all relevant variables, chosen based on previous findings in the literature. Possible confounders were investigated, and since marital status was a confounder for family type and vice versa, these
were merged. For assessing effect of each variable, Hazard Ratio (HR) with 95% confidence intervals was estimated for each model.

Weibull and Cox multiple regression analysis were performed with all the relevant variables. Variables that did not seem important after the univariable analyses were included one by one to see if it had an effect on the other variables. Harrell’s C was used as a discrimination measure for the Weibull and Cox model and was used to choose the best model and to compare the models. For assessing the PH assumption, a test based on Schoenfeld residuals was performed. The Schoenfeld Residuals Test tests the independence between residuals and time and hence is used to test the proportional Hazard assumption. The test is analogous to testing whether the slope of scaled residuals on time is zero or not. The PH assumption is violated if the slope is not zero. There is separate residual for each individual for each covariate in this test, and the covariate value for individuals that failed minus its expected value is defined as Schoenfeld residuals (Cleves, Gutierrez, Gould, & Marchenko, 2008). The Schoenfeld residuals are conducted from the Cox model, but the results are relevant for the validity of both models since the PH assumption applies for both Cox and Weibull.

Survival curves for some of the variables from the Weibull and the Cox model was conducted to graphically show the probability of survival versus time, and to see if choosing one over the other will have an impact if the results are to be used to create time-dependent transitions probabilities in economic evaluation. It will not be discussed whether the models are significantly different from each other as this goes beyond this thesis. The focus will be on the point estimates. Education and income are the two most central SES variables and age and gender are two typical variables survival probabilities are differentiated by in economic evaluations. Because of this, survival curves for education, income, age and gender was chosen.

All statistical analyses were performed using Stata 15. P-value less than 0.05 was considered as statistically significant.
4 RESULTS

4.1 DESCRIPTIVE STATISTIC OF DATA

Out of 1298 patients, 478 (37%) died and 820 (63%) were censored. Characteristics of the participants are presented in Table 1. The table shows the frequency within each category who died and survived during the follow-up period.

Education seemed to affect mortality as there were 10% more people dying within low education than within middle and high education. The income category followed the same distribution as more people died within the lower income groups. However, there were fewer deaths within the income category 400 000-600 000 than the 600 000+ group.

Family status seemed to only have a small impact on mortality, except for living alone where there was a 5-7% higher share of deaths compared to the other groups. Regarding centrality, over 70% of all subject lived in the most central municipalities. Nevertheless, the mortality distribution was fairly even, except for a smaller share of deaths in the second most and second least central municipalities.

When it comes to year of diagnosis, there were major differences from how early and how late subjects were diagnosed with regard to mortality. Only 18% of those diagnosed in 2010-2011 died, compared to 55% of those diagnosed in 2002-2003. However, this is due to follow-up time and is exactly what the survival analysis takes into account. How the proportion was distributed for the different age groups is also due to follow-up time as it was women and men aged 50-64 who were invited to the study. The gender mortality was approximately equal, with only 3% difference in proportion of deaths between males and females.
Table 1. Descriptive characteristics of the study sample

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Deaths (% dead in each category)</th>
<th>Alive (% alive in each category)</th>
<th>Total (% of total in each category)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age at diagnosis</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50 - 59</td>
<td>105 (43%)</td>
<td>140 (57%)</td>
<td>245 (19%)</td>
</tr>
<tr>
<td>60 - 69</td>
<td>291 (35%)</td>
<td>534 (65%)</td>
<td>825 (63%)</td>
</tr>
<tr>
<td>70 - 79</td>
<td>66 (29%)</td>
<td>162 (71%)</td>
<td>228 (18%)</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>253 (37%)</td>
<td>438 (63%)</td>
<td>691 (53%)</td>
</tr>
<tr>
<td>Female</td>
<td>209 (34%)</td>
<td>398 (66%)</td>
<td>607 (47%)</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>126 (43%)</td>
<td>169 (57%)</td>
<td>295 (23%)</td>
</tr>
<tr>
<td>Middle</td>
<td>226 (34%)</td>
<td>442 (66%)</td>
<td>668 (51%)</td>
</tr>
<tr>
<td>High</td>
<td>110 (33%)</td>
<td>225 (67%)</td>
<td>335 (26%)</td>
</tr>
<tr>
<td><strong>Income (NOK)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-100 000</td>
<td>63 (41%)</td>
<td>90 (59%)</td>
<td>153 (12%)</td>
</tr>
<tr>
<td>100 000 - 200 000</td>
<td>131 (41%)</td>
<td>186 (59%)</td>
<td>317 (24%)</td>
</tr>
<tr>
<td>200 000 - 300 000</td>
<td>151 (34%)</td>
<td>298 (66%)</td>
<td>449 (34%)</td>
</tr>
<tr>
<td>300 000 - 400 000</td>
<td>60 (34%)</td>
<td>119 (66%)</td>
<td>179 (14%)</td>
</tr>
<tr>
<td>400 000 - 600 000</td>
<td>34 (27%)</td>
<td>91 (73%)</td>
<td>125 (10%)</td>
</tr>
<tr>
<td>600 000 →</td>
<td>23 (31%)</td>
<td>52 (69%)</td>
<td>75 (6%)</td>
</tr>
<tr>
<td><strong>Family status</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married with children</td>
<td>129 (33%)</td>
<td>258 (67%)</td>
<td>387 (30%)</td>
</tr>
<tr>
<td>Married without children</td>
<td>129 (33%)</td>
<td>264 (67%)</td>
<td>393 (30%)</td>
</tr>
<tr>
<td>One-person family</td>
<td>164 (40%)</td>
<td>241 (60%)</td>
<td>405 (31%)</td>
</tr>
<tr>
<td>Other families</td>
<td>40 (35%)</td>
<td>73 (65%)</td>
<td>113 (9%)</td>
</tr>
<tr>
<td><strong>Classification of centrality</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Most central municipalities</td>
<td>339 (36%)</td>
<td>605 (64%)</td>
<td>944 (73%)</td>
</tr>
<tr>
<td>2. most central municipalities</td>
<td>1 (17%)</td>
<td>5 (83%)</td>
<td>6 (1%)</td>
</tr>
<tr>
<td>3. most central municipalities</td>
<td>78 (36%)</td>
<td>140 (64%)</td>
<td>218 (17%)</td>
</tr>
<tr>
<td>3. least central municipalities</td>
<td>30 (36%)</td>
<td>54 (64%)</td>
<td>84 (6%)</td>
</tr>
<tr>
<td>2. least central municipalities</td>
<td>8 (26%)</td>
<td>23 (74%)</td>
<td>31 (2%)</td>
</tr>
<tr>
<td>Least central municipalities</td>
<td>6 (40%)</td>
<td>9 (60%)</td>
<td>15 (1%)</td>
</tr>
<tr>
<td><strong>Year of diagnosis</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010-2011</td>
<td>57 (18%)</td>
<td>262 (82%)</td>
<td>319 (25%)</td>
</tr>
<tr>
<td>2008-2009</td>
<td>72 (28%)</td>
<td>188 (72%)</td>
<td>260 (20%)</td>
</tr>
<tr>
<td>2006-2007</td>
<td>81 (40%)</td>
<td>123 (60%)</td>
<td>204 (16%)</td>
</tr>
<tr>
<td>2004-2005</td>
<td>81 (45%)</td>
<td>99 (55%)</td>
<td>180 (14%)</td>
</tr>
<tr>
<td>2002-2003</td>
<td>94 (55%)</td>
<td>78 (45%)</td>
<td>172 (13%)</td>
</tr>
<tr>
<td>1999-2001</td>
<td>77 (47%)</td>
<td>86 (53%)</td>
<td>163 (12%)</td>
</tr>
</tbody>
</table>
4.2 **Weibull and Cox PH regression – univariable analysis**

Table 2 shows the univariable analysis for both Weibull and Cox.

In the analyses, education seemed to be an important predictor for mortality as it was significant in both Weibull and Cox. In both analyses, middle and high education showed an almost identical lower mortality risk compared to low education with about 29-30% lower risk (HR 1-0.71=0.29 in the Weibull model and 1-0.70=0.30 in the Cox model).

Family status also seemed to be an important predictor as the one-person family group was significant in both models. The analyses showed that one-person families had higher mortality risk compared married with children with 35-36%.

Income and classification of centrality seemed to not be important predictors for mortality, as none of them had a significant p-value. Still, from the point estimates income had a higher risk for 100 000-200 000 compared to the reference group, while the other income groups showed a lower risk with the categories 200 000-300 000 and 400 000-600 000 having a 20-21% and 29-31% lower risk compared to the reference group.

Age at diagnosis seemed to be an important predictor as the oldest group was significant in the Weibull model. Gender and year at diagnosis was not significant, and thus seemed not to be important predictors for mortality.
### Table 2. Comparison of Weibull and Cox univariable regression

<table>
<thead>
<tr>
<th>Variables/Category</th>
<th>Weibull HR (95% CI)</th>
<th>p-value</th>
<th>Cox HR (95% CI)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age at diagnosis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50 – 59 (ref.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60 – 69</td>
<td>1.15 (0.91, 1.42)</td>
<td>0.224</td>
<td>1.08 (0.87, 1.35)</td>
<td>0.492</td>
</tr>
<tr>
<td>70 – 76</td>
<td>1.41 (1.04, 1.92)</td>
<td><strong>0.028</strong></td>
<td>1.26 (0.92, 1.71)</td>
<td>0.145</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female (ref.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.87 (0.73, 1.04)</td>
<td>0.124</td>
<td>0.87 (0.72, 1.03)</td>
<td>0.123</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low (ref.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Middle</td>
<td>0.70 (0.56, 0.87)</td>
<td><strong>0.001</strong></td>
<td>0.71 (0.56, 0.87)</td>
<td><strong>0.002</strong></td>
</tr>
<tr>
<td>High</td>
<td>0.71 (0.55, 0.92)</td>
<td><strong>0.009</strong></td>
<td>0.71 (0.55, 0.91)</td>
<td><strong>0.009</strong></td>
</tr>
<tr>
<td><strong>Income (NOK)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-100 000 (ref.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100 000 - 200 000</td>
<td>1.08 (0.80, 1.44)</td>
<td>0.628</td>
<td>1.06 (0.79, 1.41)</td>
<td>0.697</td>
</tr>
<tr>
<td>200 000 - 300 000</td>
<td>0.80 (0.60, 1.06)</td>
<td>0.125</td>
<td>0.79 (0.59, 1.05)</td>
<td>0.109</td>
</tr>
<tr>
<td>300 000 - 400 000</td>
<td>0.88 (0.62, 1.24)</td>
<td>0.461</td>
<td>0.86 (0.60, 1.21)</td>
<td>0.390</td>
</tr>
<tr>
<td>400 000 - 600 000</td>
<td>0.71 (0.46, 1.06)</td>
<td>0.095</td>
<td>0.69 (0.46, 1.04)</td>
<td>0.078</td>
</tr>
<tr>
<td>600 000 →</td>
<td>0.85 (0.53, 1.36)</td>
<td>0.506</td>
<td>0.82 (0.50, 1.31)</td>
<td>0.401</td>
</tr>
<tr>
<td><strong>Family status</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married with children (ref.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married without children</td>
<td>0.97 (0.76, 1.23)</td>
<td>0.799</td>
<td>0.99 (0.77, 1.26)</td>
<td>0.952</td>
</tr>
<tr>
<td>One-person family</td>
<td>1.36 (1.08, 1.70)</td>
<td><strong>0.008</strong></td>
<td>1.35 (1.07, 1.69)</td>
<td><strong>0.009</strong></td>
</tr>
<tr>
<td>Other families</td>
<td>1.09 (0.77, 1.54)</td>
<td>0.610</td>
<td>1.09 (0.77, 1.54)</td>
<td>0.626</td>
</tr>
<tr>
<td><strong>Classification of centrality, residence</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Most central municipalities (ref.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. most central municipalities</td>
<td>0.45 (0.06, 3.23)</td>
<td>0.430</td>
<td>0.47 (0.06, 3.34)</td>
<td>0.451</td>
</tr>
<tr>
<td>3. most central municipalities</td>
<td>0.90 (0.70, 1.14)</td>
<td>0.390</td>
<td>0.92 (0.72, 1.17)</td>
<td>0.503</td>
</tr>
<tr>
<td>3. least central municipalities</td>
<td>0.88 (0.06, 1.26)</td>
<td>0.491</td>
<td>0.90 (0.62, 1.29)</td>
<td>0.568</td>
</tr>
<tr>
<td>2. least central municipalities</td>
<td>0.52 (0.26, 1.05)</td>
<td>0.071</td>
<td>0.56 (0.27, 1.11)</td>
<td>0.099</td>
</tr>
<tr>
<td>Least central municipalities</td>
<td>1.08 (0.48, 2.42)</td>
<td>0.846</td>
<td>1.09 (0.48, 2.44)</td>
<td>0.835</td>
</tr>
<tr>
<td><strong>Year of diagnosis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010-2011 (ref.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008-2009</td>
<td>0.81 (0.57, 1.14)</td>
<td>0.238</td>
<td>0.74 (0.51, 1.04)</td>
<td>0.089</td>
</tr>
<tr>
<td>2006-2007</td>
<td>0.97 (0.68, 1.35)</td>
<td>0.838</td>
<td>0.91 (0.64, 1.27)</td>
<td>0.571</td>
</tr>
<tr>
<td>2004-2005</td>
<td>0.90 (0.63, 1.27)</td>
<td>0.564</td>
<td>0.93 (0.65, 1.31)</td>
<td>0.669</td>
</tr>
<tr>
<td>2002-2003</td>
<td>1.01 (0.71, 1.41)</td>
<td>0.976</td>
<td>1.11 (0.78, 1.56)</td>
<td>0.546</td>
</tr>
<tr>
<td>1999-2001</td>
<td>0.76 (0.53, 1.08)</td>
<td>0.127</td>
<td>0.91 (0.63, 1.29)</td>
<td>0.593</td>
</tr>
</tbody>
</table>

Ref. = Reference category.
HR = Hazard Ratio, CI = Confidence interval
Bold p-value is significant
4.3 **WEIBULL AND COX PH REGRESSION – MULTIVARIABLE ANALYSIS**

Table 3 shows results from the final models of Weibull and Cox multivariable regression. The results from both models was approximately the same, and there were more significant findings in the multivariable analyses than in the univariable analyses.

The education variable had a small decrease in effect compared to the univariable analyses. Middle education was significant in both models with p-values <0.05. Compared to people with low education, people with middle education had a 26%-27% lower risk of dying. High education was not significant as it was in the univariable analyses. Regardless, higher education gave a 21-22% lower risk of dying compared to low education.

None of the income groups were significant in the univariable analyses. However, after adjusting for the other variables, the groups with 200 000 to 300 000 and 400 000 to 600 000 NOK income was significant in both models with p-values <0.05. The results show that having an income of 200 000 to 300 000 decrease the risk of dying with 28% compared to the lowest income groups if all the other variables are held constant. The risk of dying decreases even more if one has an income of 400 000 to 600 000, with a 38% lower risk. The other income groups were not significant in either model, but the HR shows a decrease in risk of dying for all groups compared to the lowest income group.

After adjusting for the other variables, family status was no longer significant. The reason was mainly confounding by the other SES variables on family status. The effect from living in a one-person family was still significant after adjusting for the background variables (age at diagnosis, gender and year of diagnosis).

Classification of centrality was still not significant in both models after adjusting for the other variables. Thus, classification of centrality does not seem to be an important predictor for mortality.
<table>
<thead>
<tr>
<th>Variables/Category</th>
<th>Weibull</th>
<th></th>
<th>p-value</th>
<th>Cox</th>
<th></th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR (95% CI)</td>
<td>p-value</td>
<td>HR (95% CI)</td>
<td>p-value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cons</td>
<td>0.37 (0.21, 0.66)</td>
<td><strong>0.001</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Age at diagnosis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50 – 59 (ref.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60 – 69</td>
<td>1.14 (0.92, 1.55)</td>
<td>0.283</td>
<td>1.13 (0.92, 1.54)</td>
<td>0.319</td>
<td></td>
<td></td>
</tr>
<tr>
<td>70 – 79</td>
<td>1.49 (1.00, 2.13)</td>
<td><strong>0.037</strong></td>
<td>1.49 (1.00, 2.14)</td>
<td><strong>0.036</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (ref.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>0.77 (0.63, 0.94)</td>
<td><strong>0.009</strong></td>
<td>0.77 (0.63, 0.94)</td>
<td><strong>0.009</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low (ref.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Middle</td>
<td>0.73 (0.58, 0.91)</td>
<td><strong>0.006</strong></td>
<td>0.74 (0.59, 0.92)</td>
<td><strong>0.008</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>0.78 (0.59, 1.04)</td>
<td>0.089</td>
<td>0.79 (0.60, 1.06)</td>
<td>0.119</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Income (NOK)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-100 000 (ref.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100 000 – 200 000</td>
<td>0.97 (0.72, 1.33)</td>
<td>0.872</td>
<td>0.97 (0.72, 1.32)</td>
<td>0.870</td>
<td></td>
<td></td>
</tr>
<tr>
<td>200 000 – 300 000</td>
<td>0.72 (0.53, 0.98)</td>
<td><strong>0.037</strong></td>
<td>0.72 (0.52, 0.98)</td>
<td><strong>0.039</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>300 000 – 400 000</td>
<td>0.79 (0.53, 1.15)</td>
<td>0.212</td>
<td>0.80 (0.53, 1.16)</td>
<td>0.224</td>
<td></td>
<td></td>
</tr>
<tr>
<td>400 000 – 600 000</td>
<td>0.62 (0.39, 0.97)</td>
<td><strong>0.037</strong></td>
<td>0.62 (0.39, 0.97)</td>
<td><strong>0.037</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>600 000 –</td>
<td>0.74 (0.43, 1.24)</td>
<td>0.252</td>
<td>0.71 (0.43, 1.22)</td>
<td>0.226</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Family status</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married with children (ref.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married without children</td>
<td>0.87 (0.67, 1.13)</td>
<td>0.310</td>
<td>0.88 (0.68, 1.15)</td>
<td>0.374</td>
<td></td>
<td></td>
</tr>
<tr>
<td>One-person family</td>
<td>1.18 (0.92, 1.51)</td>
<td>0.186</td>
<td>1.19 (0.93, 1.52)</td>
<td>0.172</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other families</td>
<td>1.03 (0.71, 1.46)</td>
<td>0.920</td>
<td>1.03 (0.71, 1.46)</td>
<td>0.916</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Classification of centrality, residence</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Most central municipalities (ref.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. most central municipalities</td>
<td>0.54 (0.07, 3.70)</td>
<td>0.511</td>
<td>0.57 (0.08, 3.87)</td>
<td>0.539</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. most central municipalities</td>
<td>0.92 (0.71, 1.18)</td>
<td>0.489</td>
<td>0.93 (0.71, 1.19)</td>
<td>0.534</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. least central municipalities</td>
<td>0.89 (0.60, 1.31)</td>
<td>0.555</td>
<td>0.89 (0.60, 1.31)</td>
<td>0.566</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. least central municipalities</td>
<td>0.60 (0.29, 1.21)</td>
<td>0.150</td>
<td>0.61 (0.30, 1.23)</td>
<td>0.164</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Least central municipalities</td>
<td>1.10 (0.48, 2.45)</td>
<td>0.845</td>
<td>1.10 (0.48, 2.44)</td>
<td>0.847</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Year of diagnosis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010-2011 (ref.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008-2009</td>
<td>0.83 (0.58, 1.18)</td>
<td>0.304</td>
<td>0.75 (0.53, 1.08)</td>
<td>0.125</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2006-2007</td>
<td>0.96 (0.67, 1.37)</td>
<td>0.825</td>
<td>0.90 (0.63, 1.29)</td>
<td>0.560</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2004-2005</td>
<td>1.04 (0.71, 1.51)</td>
<td>0.852</td>
<td>1.06 (0.73, 1.54)</td>
<td>0.777</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2002-2003</td>
<td>1.15 (0.79, 1.68)</td>
<td>0.459</td>
<td>1.26 (0.87, 1.84)</td>
<td>0.225</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1999-2001</td>
<td>0.85 (0.57, 1.27)</td>
<td>0.416</td>
<td>0.99 (0.66, 1.49)</td>
<td>0.995</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln(p)</td>
<td>0.43 (-0.51, -0.35)</td>
<td><strong>&lt;0.001</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p, shape parameter</td>
<td>0.65 (0.60, 0.71)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Harrell’s C</strong></td>
<td>0.41 (0.38, 0.53)</td>
<td><strong>&lt;0.001</strong></td>
<td>0.60 (0.56, 0.62)</td>
<td><strong>&lt;0.001</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Ref. = Reference category.
HR = Hazard Ratio, CI = Confidence interval
Harrell’s C shows coefficient
Bold p-value is significant
After adjusting for the other variables, the HR for the oldest age group was significant in the Cox model as well, meaning that the oldest group had a 49% higher risk of dying than the youngest group in both models if all the other variables are held constant. As the other age groups were not significant, the models do not estimate any monotonous increase in mortality risk with age.

Gender was significant compared to the univariable analyses with a p-value of <0.05 in both models. This means that females had a 23% lower risk of dying than males if all the other variables are held constant.

The Weibull model reports a Wald test for the null-hypothesis that $\ln(p) = 0$, meaning that the hazard is constant. We see that the $\ln(p)$ estimate was significant. Thus, the null hypothesis can be rejected. The shape parameter was <1, meaning that the hazard decrease with time.

4.4 Model validation – Schoenfeld residual and Harrell’s C-statistic

After using the Weibull and Cox regression model, PH assumption for each of the variables were investigated. Test results based on Schoenfeld residuals are shown in Table A.1 in appendices. The results showed that none of the variables violated the PH assumption (i.e. p<0.05), which implies that the proportional hazards assumption is fulfilled.

The result from the Harrell’s C-statistic was 0.41 (C.I. 0.38, 0.53) for the Weibull model and 0.60 (C.I. 0.56, 0.62) for the Cox model. This means that the Weibull model correctly identifies the order for the survival times for pairs of subject 41% of the time and the Cox model 60% of the times. The confidence interval for the Weibull model includes 0.5 which means that the model is no better at predicting an outcome than random chance. The result from the Cox model is a little bit higher than 0.5 with the 0.56 as the lowest confidence interval, meaning it is significantly different from 0.5.
4.5 Survival Curves

The survival curves show the proportion of the subject that survived, plotted against time. Figure 2-5 illustrates the survival curves of different groups represented from both the Weibull and the Cox model. The dotted lines represent the Cox survival curves and the solid lines represent the Weibull survival curves.

In Figure 2, survival curves for the different age groups are represented. The Weibull and Cox gave the same results in terms of the order of the graphs. There was a higher proportion that survived in the youngest age group, than in the oldest age group throughout the follow-up period. The curves overlapped the first year but after that, Weibull gave higher survival probability than Cox until the curves crossed after 6-7 years. It was after 6-7 years the difference between the survival probabilities from the two models appeared, where the curves from Weibull continued with a steady decline in survival probability over time and the curves from Cox began to flatten out and eventually became completely horizontal after 9-10 years.

Figure 2. Survival curves from Weibull and Cox model for age
Figure 3 shows the survival curves from Weibull and Cox for the two genders. The curves have the same pattern as for the age groups, where they overlapped the first year and the curves from Weibull gave higher probability from one year until the curves crossed after 6 years. Then the curves from Cox began to flatten out and Weibull had a steady decrease. After 13 years, the difference between the Weibull and the Cox curves was so big that the probability of surviving for females in the Weibull model was actually lower than for males in the Cox model. There was over 10% difference between the models when the follow-up period is over (0.56 and 0.47 for females and males in Cox model, and 0.45 and 0.35 in Weibull model). This means that if time-dependent annual transitions probabilities would have been made from this study, the transitions probabilities would be quite similar towards the first 6-7 years, but the estimates would have been more different after that. This applies to age as well, and after 13 years, the estimates would differ with more that 10%.

**Figure 3.** Survival curves from Weibull and Cox model for gender
Figure 4 shows the survival curves from Weibull and Cox for the different education levels. The curves also show the same pattern as the age groups and gender considering the differences between Weibull and Cox. The difference between the Weibull and Cox curves after 13 years was over 10% for education as well and the same differences would appear for education in terms of making time-dependent annual transitions probabilities from the two models.

**Figure 4.** Survival curves from Weibull and Cox model for education
Figure 5 shows the survival curves from Weibull and Cox for the different income groups. For income, the Weibull and Cox did not give the exact same results in terms of the order of the graphs. Cox gave the exact same survival curve for people with income of 200 000 to 300 000 and 600 000+, where Weibull provided a slightly higher survival curve for people with 200 000 to 300 000 than for 600 000+. Apart from this, the models agree that the order of the groups, ranked by the proportion that survives from lowest to highest, was: 0 to 100 000, 100 000 to 200 000, 300 000 to 400 000, 600 000+, 200 000 to 300 000, and 400 000 to 600 000. The differences are as great for income as for the other variables.

Overall, the results indicate that if the models were to be used for making time-dependent transition probabilities in an economic evaluation, the probabilities would deviate quite much as a function of time since diagnosis depending on whether a Cox or a Weibull model was chosen for the analysis.

**Figure 5.** Survival curves from Weibull and Cox model for income
5 Discussion

This study attempted to estimate effects of socioeconomic status on overall mortality after diagnosed with CRC and to compare results from the Cox proportional hazard and Weibull models. To the best of my knowledge, this is the first study on the association between socioeconomic status and mortality after diagnosed with colorectal cancer with data from the 21st century in Norway. This is also the first comparison between the Cox and Weibull model in terms of discussing implications for making time-dependent transition probabilities.

5.1 Main findings

The regression analyses found that there is a difference on overall mortality after diagnosed with CRC, partly associated with SES. Cox and Weibull gave very similar results in terms of effects, as the hazard ratios and p-values were virtually identical in the multivariable models. The multivariable analyses found that a middle level of education was associated with a reduced risk of mortality, compared to a low level of education. The unadjusted results showed a significant reduction in risk for the highly educated, however, after adjusting for the other factors, the results were not significant.

Income was also associated with a reduced risk of mortality after diagnosed with CRC with significant results for the groups 200 000 - 300 000 and 400 000 – 600 000. Income and education are probably interrelated as people with higher education likely earn more. In addition, a plausible explanation of why there is no clear pattern over the groups for education and income, can be that some of them are small. As mentioned, the result for those with highest education was not significant in the multivariable analysis. However, the highest education group has half as many individuals compared to the group with middle education. Also, the significant income group of 200 000 - 300 000 is by far the largest group among those who have a little higher income.

Centrality and family status did not seem to have an effect, even though one-person family was significant in the unadjusted analysis. Of the remaining variables, an interesting finding was that gender was not significant in the unadjusted analyses but became significant after
adjustment with a 23% lower risk for females. This strengthens the assumption that SES is different between genders, and that SES are confounding factors for gender.

The result from the Harrell’s C-statistic was poor for both models with 0.41 (C.I. 0.38, 0.53) for Weibull and 0.60 (C.I. 0.56, 0.62) for Cox. This is not surprising, as the analysis did only include a few independent variables and is such a simple model. In addition, income was only available as categorical in the data due to restrictions. Still, the results indicated somewhat better predictive power of the Cox model than the Weibull.

The survival curves gave quite different results after 6-7 years even though the multivariable analyses gave virtually identical hazard ratios and p-values. Also, both showed roughly similar and poor predictive ability as measured by Harrell’s C, although Cox was significantly better than Weibull since the confidence intervals did not overlap. One should then think that choosing Cox over Weibull was indifferent also for predictions. Nevertheless, survival curves diverged over time. The survival curves from Cox flattens out when the survival curves from the Weibull continue to decline during the entire time period. The reason why the Weibull continues to decline is the shape-parameter being less than 1, indicating evidence of a decreasing hazard over time. Hence, if used in a Markov model, the time-dependent transition probabilities made with equation 11 will continue to decrease as the cycles elapse even when this is not necessary supported by the observed data. Although this makes sense as the mortality risk is highest shortly after diagnosis of CRC cancer (The Norwegian Directorate of Health, 2017a), the curves decrease much more than the curves from Cox. The results are an approximately 10% difference in survival at the end of the follow-up time for all groups between the two models. It is surprising that models that both have PH and have the same covariate effects yield so different survival curves. To my knowledge, this is not well reported in the literature, and a potentially unknown result worth studying further. However, taking uncertainty into account due to less participants being observed for the full 13 years, the big difference between the Cox model and the Weibull model may not be significant.
5.2 PREVIOUS RESEARCH

Multiple studies have tried to address any social gradient in colorectal cancer mortality all over the world. Overall, the findings from this study are consistently the same as findings in literature mentioned in the background chapter. Higher education and income is significantly linked to decreased mortality in general (Adams, Hurd, McFadden, Merrill, & Ribeiro, 2003; Dahl, Elstad, Hofoss, & Martin-Mollard, 2006) and for mortality after diagnosed with CRC (Kravdal, 2000; Elstad, Torstensrud, Lyngstad, & Kravdal, 2011). Effect of marital status has been found, where unmarried/people living alone has been significantly associated with higher mortality rates in general (Dahl, Elstad, Hofoss, & Martin-Mollard, 2006) and for mortality after diagnosed with CRC (Kravdal, 2000). How central one lives has been shown to have a positive impact on general survival in other countries, but not in Norway or Finland (Kravdal, 2006; Kravdal, 2000; Auvinen, 1992). However, a comparison between the results from this study and other studies can be challenging because of different measurements. One study using similar measurement for education, income, family status and classification of central residency in the analysis was identified.

A study on social inequalities in cancer from a Norwegian cohort conducted by Kravdal in 2000 considered twelve cancer sites where colorectal cancer being one of them (Kravdal, 2000). He looked at the impact on mortality from education, income, marital status (married, never-married, divorce/separated, widowed), and place of residence (eastern non-central, eastern central, southern/western non-central, southern/western central, Trøndelag/northern non-central, Trøndelag/northern central), but did not include all four in the same model. He controlled for age, period, sub-site, stage and histology while this thesis controlled for age, gender and period. When education was the only socioeconomic variable he included, it showed to have a significant effect on reducing mortality. However, when including education and income in the same model, the effect of education was reduced but remained significant. The results showed a significant negative social gradient for both education and income. The effect of education was also reduced in this thesis after all variables was included. When he included marital status, he only included it with education and not income. He got a significant result for a higher risk for non-married compared to married but concluded that marital status was not an important confounder because of only a quite small change in the education effect. The analyses in this thesis did not include marital status, but family status which are similar to each other. However, one-person family in the unadjusted analysis was the only significant result and included divorced/separated, widowed and
unmarried, which are the same category groups Kravdal used except “Married”. Divorced/separated and widowed was not significant in Kravdal’s study, and in addition was almost no different to being married. This can mean that it was the unmarried individuals in the one-person families that made the result significant in the unadjusted analysis. In addition, since it did not give a change in the educational effect when marital status was included, it supports the theory that family status is not a confounder for education.

He also included place of residence alone with education. Controlling for place of residence also left the education effect estimates unchanged, and it was only Eastern central that was significant with a HR of 0.92. This mainly corresponds with the results from this thesis, that place of residency/ how central one lives has little effect on mortality.

5.3 INTERPRETATION OF RESULTS
The results regarding education was as expected because of the previous research mentioned earlier which found that higher education improves survival in general health and the study by Kravdal which found that higher education reduce mortality after diagnosed with CRC. The educational effect remaining after controlling for other variables can be interpreted as the importance of cognitive resources in the broad sense, such as body knowledge, communication ability and, in general, coping with life challenges.

The findings that income is interrelated with other variables support previous research. Education can, and most likely will, have a strong influence on income (Dahl, Bergsli, & van der Wel, 2014), meaning that the lower education one has, the lower income. Furthermore, it was the other SES factors that caused the effect of family status to disappear, and not age at diagnosis, gender or year of diagnosis. This points to a connection between the SES factors, meaning that a SES index including education, income, and family status could possibly have been used. In addition, the results of a negative social gradient in education and income correspond to the general findings that better health and longer life is associated with higher socioeconomic status, compared to those with a low level of education and low income (Adams, Hurd, McFadden, Merrill, & Ribeiro, 2003). On the other hand, gender and income are also known to be interrelated as men earn more than women. The results for gender and income showed an even lower risk after adjusting for each other and the other variables. This
means that women who have an income of 200 000-300 000 or 400 000 – 600 000 have a lower risk than men who earn the same.

In the light of the results regarding family status and how centrally one lives, the socioeconomic variables seem to have an impact on mortality through economic and cognitive resources, in a broad sense, rather than through family and closeness to health care institutions. The latter being very positive as Norway have equal access and right to healthcare. It can therefore be assumed that the differences in mortality may be, at least partial, due to education and income inequality. However, the Norwegian health system is almost free of charge and income inequalities should not cause a difference regarding point of access. If this is the case, it affects the equal access to healthcare even though it is almost free of charge. On the other hand, the effect of income inequalities on mortality is not necessary due to health care costs but may be due to other factors, for example health-risk behaviours or that people with higher SES lives longer in general.

Looking at age at diagnosis and year of diagnosis in the descriptive analysis, one can see how wrong and meaningless results can be obtained if survival analysis is not used in studies with different follow-up times. In the descriptive analysis, it looks like the oldest individuals lives longer than the youngest individuals do, which is of course wrong. However, given the results from the survival analysis, year at diagnosis was not significant in either univariable or multivariable analyses. This indicates that there have not been much revolutionary treatments for CRC during the period. Of course, these interpretations are limited due to short follow-up period in the recent years. In addition, the analysis shows an increasing mortality risk with age, which is not surprising.

If differences in mortality may be, partial, due to education and income inequality, a reduction of social inequality should lead to minor differences in mortality. To find out how to possible minimize social inequalities, one might look at the possible explanations for social inequality in health. The mobility explanation is based on health status determining social class, and not the other way around. This means that if all individuals had good health, there would not be any social classes. However, there are, for example, some generic factors that cannot be controlled for and thus it is impossible to achieve a society where everyone has equal health.

Another explanation is the materialist approach, which explains health inequalities through differences in socioeconomic positions. Despite a well-functioning health care system with equal access for all, social inequalities in mortality are found in this study. The theory that if
good material conditions such as clean water and healthy food, sanitation, and, most important, equal access to health care, were available for all, social inequality in health could be reduced (Elstad, 2005), is probably right. However, since these material conditions are available for all in Norway already, other explanations need to be investigated.

The cultural behaviour approach and the psycho-social perspective states that people with lower SES have higher health-risk behaviour and that they experience more negative life events. This means that if health-risk behaviour is reduced in those with low SES and that those with low SES maybe get more help from, for example, government organisations, the social inequalities might get reduced and leads to minor inequalities in mortality.

In addition, considering the life-course perspective, which says that health is also determined by past living conditions and events and is not just a result of current conditions and lifestyle choices (Elstad, 2005), start helping the younger in families with low SES may make them more resourceful and reduce the social inequalities later in life. However, these explanations and suggestions are very complex and would need extensive research.

The difference between the Cox and Weibull model in survival curves may not be significant. However, regardless of significance, the estimated cost-effectiveness in economic evaluation of some health intervention is often strongly influenced by the choice of survival curve. This is because the estimated expected costs and health effects for, for example, education level are functions of the expected time patients survive. Seemingly minor changes in curve fits can indeed often have an important impact on cost effectiveness, especially if considerable extrapolation is necessary (Hoyle & Henley, 2011). Although the difference may not be significant, the curves indicate that, based on point estimates, there may be situations where a point estimate for cost per life-year states that a treatment is cost-effective from a Markov model with transition probabilities based on Cox, but not cost-effective based on Weibull. Since one will often be led by point estimates, this may affect whether to obtain more information or not for further cost-benefit analyses. In addition, since in economic evaluation it is most common to estimate the survival curves by fitting a parametric survival model, the Weibull model would most likely been chosen if the data used in this study was to be used in a cost-effectiveness analysis. However, since the Cox model was shown to have a (barely) better fit and the point estimates in the time interval would have been the starting point, if the Weibull model was chosen it could yield wrong cost-effectiveness estimates that seem cost-effective. This reinforces the importance of comparing different models, and from the
literature you can get the impression that those who use Weibull do not consider Cox and vice versa which is a possible problem.

5.4 LIMITATIONS
A limitation of this study is that there are generally few variables to adjust for. As a result, the model only captures a small part of the variation in survival times and makes the predictive ability of the models low. Income could have been a continuous variable but had to be used categorical due to the data and could improve the predictive power. However, all the other variables that could have been included as continuous variables was tested in both ways. For year at diagnosis and classification of centrality, it did not matter if they were categorical or continuous, and education was only significant when it was included as a categorical variable. The limited sample size in some of the income groups and in the groups of highest education could be another limitation and might be the explanation for why there are no clear trends in the hazard ratios for education and income.

Not controlling for stage could also be a limitation, as a study from Finland found that controlling for stage at diagnosis of colon cancer diminished the difference in risk estimates between the highest and the lowest social class by approximately 50% (Auvinen, 1992). However, other studies which controlled for stage at diagnosis still found a significant beneficial effect of socioeconomic resources. For example, a study from Netherland (Schrijvers, Coebergh, Heijden, & Mackenbach, 1995) which had an index measure for SES where they controlled for urban residence, year at diagnosis, gender and age in addition to stage at diagnosis found a 4% increase in mortality after CRC for those with middle SES and a 25% increase in mortality for those with low SES. The effect of SES on mortality was in other words still considerably.

Another limitation could be that the model is not controlled for comorbidities, which is the occurrence of several different diseases or disorders at the same time in the same person. This could influence the results, as comorbidities are likely associated with both a higher mortality risk, and individuals of low SES may have a higher number of comorbidities than individuals of higher SES. In addition, a larger age span could have made the model more accurate as incidence is increasing in young people and declining in elderly globally (Holt, Kozuch, & Mewar, 2009) and because of the aging population.
In addition, there are several risk factors that could affect the mortality from CRC. None of them are controlled for in the analyses and might impact the certainty of the findings. However, lower levels of education and income has been associated with a significantly higher prevalence of health-risk behaviours, including smoking, being overweight, and physical inactivity (Lantz, et al., 1998). Even though they are not directly measured in this study, the use of education and income might adjust for some of the risk factors. Therefore, a reduced social inequality may lead to a reduction in health-risk behaviours, thus also reducing mortality. Regarding genetic and medical conditions associated for increased CRC risk, they are impossible to adjust for in this study. However, it may be assumed that genetic and medical conditions are independent of socioeconomic status and thus would not change the estimates found in this study. Endoscopic screening would perhaps change the results a little because of higher participation rates among high SES groups, as probability of attending a screening program in Norway is affected by the individual’s socioeconomic status (Dahl, Bergsli, & van der Wel, 2014). However, this study could not adjust for screening as the results is to be published in another article. Additionally, it is uncertain whether screening has any effect after a given illness.
6 CONCLUSION

This thesis show that socioeconomic status seems to have an impact on overall mortality after diagnosed with CRC as education and income had a significant effect on mortality. The fact that a central place of residence does not seem to have an impact in mortality, is very positive. Norway is a welfare state with a public health care system with equal access and equal rights to healthcare, and how central one lives should not have an impact on that.

It was surprising that similar hazard ratios from a Cox model and a Weibull model nevertheless gave some deviations in estimated survival over time in these data. This highlight the importance of comparing survival models not only between hazard ratios and predictive power, but how the choice of model impacts the survival over time.

There is still need for more research on the effect of socioeconomic status on overall mortality after diagnosed with colorectal cancer. Further studies including possible risk factors and other variables mentioned would most likely give more accurate results. In addition, the society is constantly changing and thus these results may be time limited. For example, a national screening program may soon be introduced. If differences in mortality are partially due to education and income inequality, a reduction of social inequality should lead to reduced differences in mortality. Efforts to reduce social inequalities should therefore be made.
REFERENCES


### A.1 Schoenfeld Residuals

| Table A1. Test of proportionality of hazards on Schoenfeld residuals |
|---------------------|---------------------|---------------------|
| Factor             | Group               | chi2    | p-value |
| Age at diagnosis   |                     |         |         |
| 50 - 59            | Reference           | 6.00    | 0.40    |
| 60 - 69            | 0.40                | 0.52    |
| 70 - 79            | 0.10                | 0.75    |
| Gender             |                     |         |         |
| Male               | Reference           | 0.33    | 0.57    |
| Female             |                     |         |         |
| Education          |                     |         |         |
| Low                | Reference           | 0.31    | 0.58    |
| Middle             | 1.51                | 0.22    |
| High               |                     |         |         |
| Income (NOK)       |                     |         |         |
| 0-100 000          | Reference           | 0.01    | 0.92    |
| 100 000 - 200 000  | 0.01                | 0.92    |
| 200 000 - 300 000  | 1.05                | 0.31    |
| 300 000 - 400 000  | 0.01                | 0.92    |
| 400 000 - 600 000  | 0.78                | 0.38    |
| 600 000 →         | 0.34                | 0.56    |
| Family status      |                     |         |         |
| Married with children | Reference      | 0.85    | 0.36    |
| Married without children |         | 1.39    | 0.24    |
| One-person family  |                     |         |         |
| Other families     |                     |         |         |
| Classification of centrality |  |         |         |
| Most central municipalities | Reference | 0.56    | 0.46    |
| 2. most central municipalities |         | 0.00    | 0.98    |
| 3. most central municipalities |         | 0.63    | 0.43    |
| 2. least central municipalities |         | 0.14    | 0.71    |
| Least central municipalities |         | 0.30    | 0.59    |
| Year of diagnosis  |                     |         |         |
| 2010-2011          | Reference           | 12.99   | 0.97    |
| 2008-2009          | 0.04                | 0.84    |
| 2006-2007          | 0.00                | 0.98    |
| 2004-2005          | 0.02                | 0.88    |
| 2002-2003          | 0.05                | 0.82    |
| 1999-2001          | 0.74                | 0.39    |
| Global test        |                     |         |         |
A.2 REK`s Decision Letter

Eline Aas
Universitetet i Oslo

2013/83 Screening i et helseøkonomisk perspektiv S-02113

Forskningsansvarlig: Universitetet i Oslo

Prosjektleader: Eline Aas

Vi viser til søknad om prosjekttjedring datert 08.12.2017 for ovennevnte forskningsprosjekt. Seknden er behandlet av sekretariatet i REK sør-øst på delegert fullmakt fra REK sø-øst A, med hjemmel i helseforskningsloven § 11.

Vurdering

REK har vurdert følgende endringer i prosjektet:
- Nye medarbeidere. Student Ilge Vermundsen Skappel, Universitetet i Oslo, professor Tor Iversen, Universitetet i Oslo og førstemannen Magnus Løberg, Universitetet i Oslo søkes knyttet til prosjektet.
- Mathy Verbaan er ikke ansatt ved Universitetet mer, så hun skal ikke jobbe videre på prosjektet og Mette Kalager jobber ikke aktivt med databehandling.

REK har vurdert søknaden og har ingen innvendinger til de endringer som er beskrevet.

Vedtak

Komiteen godkjenner med hjemmel i helseforskningsloven § 11 annet ledd at prosjektet videreføres i samvær med det som fremgår av søknaden om prosjekttjedring og i samvær med de bestemmelser som følger av helseforskningsloven med forskrift.

Dersom det skal gjøres ytterligere endringer i prosjektet i forhold til de opplysninger som er gitt i søknaden, må prosjektleader sende ny endringsmelding til REK.

Av dokumentasjonshensyn skal opplysningene oppbevares i 5 år etter prosjekttdutt. Opplysningene skal deretter slettes eller anonymiseres.

Opplysningene skal oppbevares avidentificert, dvs. et skilt i en mållinje- og en datafil. Forskningsprosjektets data skal oppbevares forsyvlig, av personopplysningsforskriften kapittel 2, og Helsedirektoratets veiledere for «Personvern og informasjons Sikkerhet i forskningsprosjekter innenfor helse- og omsorgssectoren».

Prosjektet skal sende sluttmelding til REK, se helseforskningsloven § 12, senest 6 måneder etter at prosjektet er avsluttet.
Klageadgjøring


Med vennlig hilsen

Knut Ruyter
Avdelingsdirektør
REK sett 1 og 2 sekretariat

Anne S. Kvalø
Seksjonsleiar

Kopi til: n.kvaloeu@medisin.uio.no, Universitetet i Oslo, medisinsk fakultet ved overveiende administrative ledelse, postnumeral@medisin.uio.no, Universitetet i Oslo ved overveiende administrative ledelse, universitetsdirektor@uio.no