UiO : Universitetet i Oslo

FACULTY OF MEDICINE

DEPARTMENT OF HEALTH MANAGEMENT AND HEALTH ECONOMICS

Regional differences in diagnostic stages and survival in colorectal and lung cancer:

Estimating costs and health gains of earlier diagnosis using Norwegian register data

Author:

Ida Laurendz Kommandantvold

Supervisor:

Eline Aas

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

May 2018

Regional differences in diagnostic stages and survival in colorectal and lung cancer:

Estimating costs and health gains of earlier diagnosis using Norwegian register data

2018

An analysis of regional differences in diagnostic stage and survival colorectal and lung cancer

Ida Laurendz Kommandantvold

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

Trykk: Reprosentralen, Universitetet i Oslo

Abstract

Background: An estimated one out of three individuals in Norway will at some point receive a cancer diagnosis. Previous research has identified a geographical difference in incidence of cancer across the different Norwegian counties. Earlier analysis has also reviled that a higher degree of cancer severity is usually associated with increased treatment costs. In addition, healthcare spending on cancer in Norway has increased substantially over the last decade.

Objective: The main objective of this thesis was to investigate stage-specific costs and survival for colorectal and lung cancer in Norway, in order to investigate potential health and financial gains by earlier diagnosis, specifically by regional assessment of *best practice*.

Methods: Data from the Cancer Registry of Norway was used to statistically test for significant difference between several prognostic factors and their effect on survival time. Variations related to any significant difference in survival, were then assessed by both parametric and non-parametric estimations. Using the dataset from the CRN, regional variations was investigated to assess *best practice* and incremental health gains were calculated using predictions from two different survival estimation methods. Analyzing potential cost saving scenarios using stage-specific costs for the two cancers were possible through a pathway model exploring the cost of lung cancer treatment, and a previously published decision-analytic model for colorectal cancer, both using a Norwegian perspective.

Results: The main non-parametric estimation reveled a regional variation in median survival time of between 6.19-9.23 and 48.24-66.59 months for lung and colorectal cancer patients, respectfully. The incremental health gain between the highest and lowest grade of cancer severity was 44.78 and 168.8 months for lung and colorectal cancer, respectfully. By applying a scenario of best regional practice, the results yielded a potential per patient and yearly cost saving for lung and colorectal cancer of NOK 12 269 and NOK 11 202, and approximately NOK 37 mill and NOK 50 million, respectfully.

Conclusion: There exists regional difference in the prognostic staging distribution and survival, which could indicate a disruption within the Norwegian principle of equal access and services. There is also substantial health and financial gains to be achieved from earlier diagnosis. Additional research should focus on gathering more detailed documentation and reporting on cancer treatment, socioeconomic variables, and supplier-demand variations. Furthermore, additional probabilistic decision-analytic modeling could investigate the uncertainty of cancer related costs, and also explore other financial scenarios for different sub-groups.

Acknowledgement

I would first like to take this opportunity to thank my supervisor Eline Aas, and Associated Professor Knut Reidar Wangen, at the Department of Health Management and Health Economics at the University of Oslo, for their advice and feedback. I would also like to express my gratitude to the Norwegian Cancer Society for placing their confidence in me by giving me the amazing opportunity to conduct research for this master thesis.

To the Department of Health Management and Health Economics, thank you for a great learning experience, and to my fellow students for a fun learning environment. A special thanks goes to Heidi M. Nilsen, whose valuable discussions of the thesis, and most of all friendship throughout these past two years has provided me with unforgettable memories.

Gratitude and a special thanks to my parents, Geir and Anne-Katrine, for their never ending support, encouragement and belief in me. None of this would have been possible without them.

Finally, I would like to give a big thanks to Svenn Alexander Hansen, the best partner in crime a person could ask for.

Table of content

List of Figures

List of Tables

Abbreviations

CRC	Colorectal cancer
TNM	Tumor-Node-Metastasis (Classification system)
NSCLC	Non-small-cell lung cancer
SCLC	Small-cell lung cancers
Hdir	Norwegian Directorate of Health
GP	General practitioner
IBD	Inflammatory bowel disease
NOK	Norwegian kroners
BMJ	British Medical Journal
SSB	Statistics Norway
CRN	Cancer Registry of Norway
CT	Computed Axial Tomography
RHA	Regional Health Authorities
KM	Kaplan-Meier
PH	Proportional hazard
AIC	Akaike's Information Criterion
BIC	Bayesian Information Criterion
DRG	Diagnostic Related Group

1 Introduction

Malignant cancers are the second leading cause of death in Norway, and a total of 32 827 new cases were reported in 2016. The incidence of cancers is expected to increase, due to increased life expectancy, new detection methods, and screening initiatives [1]. The diagnosis of cancer has a detrimental impact on patients' quality of life and the outcome is often negative, indicating short survival. Furthermore, the staging of cancer has a major impact on treatment options and survival. Accurate staging of cancer is vital towards treatment options and prognosis.

Previous research has identified a geographical difference concerning incidence of cancer across the different Norwegian counties [2]. Differences in treatment between the regions have also been identified, and quality improvements regarding equal access to health care have been much discussed [2]. The overarching aim of the Norwegian health services is to provide equal access, disregarding both geographic and demographic background [3]. Therefore, and identification of regional difference in the prognostic staging distribution and survival probability could indicate a disruption within this principle.

Cancer is also a resource demanding illness to assess, treat and monitor. Earlier analysis has also reviled that a higher degree of cancer severity is usually associated with increased treatment costs [4]. In Norway, health care spending related to cancer has increased with the growing number of cases, in conjunction with the increase in expensive new treatments and other factors.

This thesis will investigate regional variations in survival by diagnostic factors, and health care costs of colorectal and lung cancer in Norway. These two forms of cancers were chosen for investigation due to their different attributes and impact on epidemiology, survival and cost:

- In 2014, more Norwegians died of lung cancer than of any other cancer, making it the leading cause of cancer mortality. This is mostly attributed to the absence of clinical symptoms and effective screening program, often causing diagnosis to be set late in the cancer progression. While new treatments have made an impact on life-expectancy, long-term survival for lung cancer patients continues to be poor [5].
- After an initial skewed distribution of colorectal cancer, with a predominance of men, it today affects both genders almost equally. The rise of this cancer during the last decades, mostly attributed to environmental factors, makes colorectal cancer the most diagnosed cancer in Norway for both genders [6].

This thesis is inspired by the British report by Cancer Research UK "Saving lives, averting costs" [7], which investigated the costs and effects of earlier diagnosis. There exists extensive research into stage-specific cancer survival for lung and colorectal cancer, which is within the scope of this thesis. Norwegian research also exists, but often not as extensive as international publications. Norway however, has an extensive population-based registry, the Cancer Registry of Norway (CRN) which allows for extensive research into cancer related topics.

The main objective of this thesis is to analyze the regional survival attributes for colorectal and lung cancer across regions in Norway. Health care related costs associated with each stage of cancer will be estimated in order to investigate cost of lung, and colorectal cancer in a Norwegian setting. A lung cancer treatment pathway model to assess stage-specific lung cancer cost will be developed, and together with previous research of stage-specific costs in Norway for colorectal cancer, these two costing models will be used to examine potential financial gains by achieving earlier diagnosis. Data from the CRN will be investigated in order to identify potential health gain by earlier diagnosis, and to assess regional variations in survival time. To contribute to the subject of registry survival analysis, an adjudication between the most used survival distributions will be performed with regards to best model fit.

Chapter 2 will provide an introduction to the two cancers in question, specifically the risk factors, etiology and epidemiology. Information on clinical management and treatment will be provided, reflecting the current practice in the Norway. The Norwegian health care system will be explained briefly to set the backdrop for regional analysis. In addition, national initiatives will be presented, explaining current strategies in cancer management and enablers for earlier diagnosis. Finally, the financial impact of cancer in Norway will put the decision problem further in context.

Chapter 3 will provide an overview of the dataset from the Cancer Registry of Norway, including an explanation of the variables used in the main analysis and how the data was processed to ensure data quality. The complete listing of data used in the costing pathway for lung cancer will be provided together with other important data and considerations.

In Chapter 4 the concept of survival analysis will be introduced and explained. Interpretation of the most used methodological approaches will be outlined, together with the consideration of other statistical models and how they might be used to estimate the effect of one or more factors that may predict survival. The design and scope of the modelling exercise used to

estimate the costs of lung cancer is also presented. Key components of the overarching study design will be described together with explanation of the analysis and scenarios. Throughout Chapter 4 there will be given brief explanations of the theoretical framework to support the analysis.

Chapter 5 will present the results from the various statistical analysis, and the adjunction between the different survival models. The regional differences will be highlighted, and the results from the health and cost gains scenarios will be presented.

In Chapter 6 the findings from the analysis will be interpreted, discussed and compared to previous research on stage-specific survival and costs. Interpretation of the study will be discussed, together with possible policy implications and impact on future research. Limitations are also addressed as to guide readers and for generalizability.

Chapter 7 will give an abbreviated summary of the main findings of the study. The conclusion will also recommend the need for future research and possible areas of interest for further analysis. The thesis` contribution to the important discussion of regional variations will also be highlighted.

2 Background

Cancers can develop at any point during an individual's lifetime, and an estimated one out of three individuals in Norway will at some point receive a cancer diagnosis [1]. Both sexes are affected at an almost equal rate, but there is are some cancers that are considered gender specific. There are a number of both known and unknown risk factors with the development of cancer, but it is most often a function of increased age. Due to increased life expectancy, general lifestyle trends, and new methods of detection and screening initiatives, the incidence of cancers is only expected to grow [1]. Malignant cancers are the second most leading cause of death in Norway, with more than 10 000 deaths in 2015, and cancer in the colon, rectum, lung, prostate and female breast accounted for approximately 50% of the cancer mortality [2]. More than 30 000 new cases were reported in 2015 and at the end of 2016, 262 884 Norwegians were alive after being diagnosed with cancer, being either still in treatment, remission or considered disease-free [1].

2.1 Cancer

2.1.1 Risk factors

A common risk factor for both cancers is tobacco smoking. The smoking prevalence in three Norwegian counties from 1999-2017 is depicted in Figure 1 and shows a steady decline over the last 18 years. In 1999, the largest percentage of everyday smokers were found in Finnmark with 45 %, while in 2017 the largest portion was found in Østfold with 17 %. The lowest percentage over the last 10 years is found in Oslo, with 9 % everyday smokers between the ages of 16-74 in 2017. This change in smoking prevalence is contributed to multiple factors, such as general tobacco trends and effective legislation.

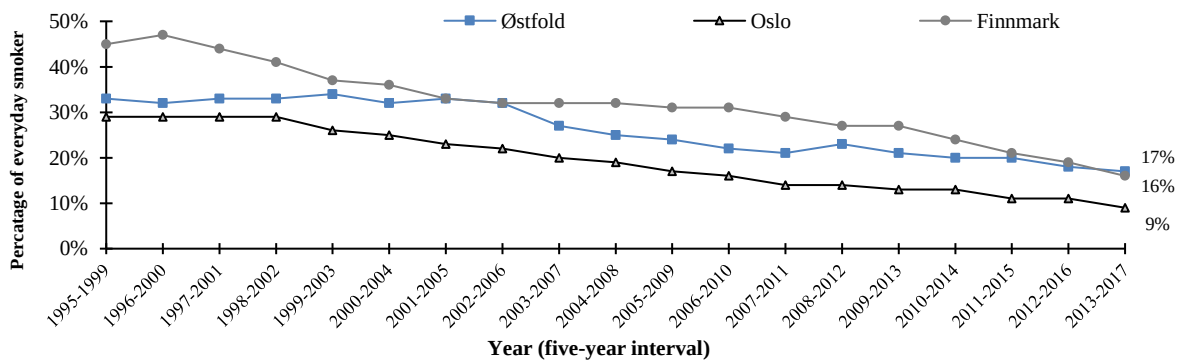


Figure 1: Smoking prevalence in three Norwegian counties by time-period. Percentage (%) of everyday smokers 16-74 years old. Source: SSB table 07662.

Lung cancer

The lung is a complex organ composed of a great number of cell types with specialized functions to support the main task of the lung, which is gas exchange. Efficient passage of oxygen and carbon dioxide between the blood and surrounding environment occurs against exposure of toxic gases and fine particles as well as infectious agents. The primary risk factor for lung cancer is tobacco smoking, which accounts for approximately 85% of all lung cancer diagnoses. The second most leading cause of lung cancer is radon exposure [8]. Other risk factors include indoor and outdoor air pollution, exposure to hazardous chemicals and asbestos [9].

Colorectal cancer

The occurrence of colorectal cancer (CRC) is contributed to both nonmodifiable and environmental risk factors. Nonmodifiable risk factors refers to those attributes that an individual cannot control with regards to development of CRC, and includes age and genetic/hereditary susceptibility [6]. Approximately 5-10% of CRC are a direct consequence of recognized hereditary conditions [6]. While some cases of CRC are attributed to nonmodifiable factors, it is widely considered to be an environmental disease and includes cultural, social and lifestyle factors. Nutritional practices strongly influence the risk of colorectal cancer. Excessive alcohol consumption, a high fat diet and extensive meat consumption, in combination with a low consumption of fruit and vegetables is especially liked with higher risk [6]. Physical inactivity and obesity are also risk factors associated with CRC. While tobacco smoking is associated most commonly with the development of lung cancer, it is also harmful to the colon and rectum and it has been estimated that 12% of CRC deaths are attributed to tobacco smoking [10].

2.1.2 Etiology

Cancer is a group of disease involving abnormal cell growth with the potential to invade or spread to other parts of the body [11]. Oncogenes are cancer-susceptible cells and arise when a single cellular lineage receives multiple mutations in contradiction to a normal single mutation. Whenever a cell is divided, another cell has to perish to ensure that the total number of cell remains the same. Cancer occurs when the equilibrium between cell birth and cell death is altered [12].

Lung cancer

Tobacco smoking and other types of hazardous exposures typically manifests as chronic obstructive pulmonary disease and emphysema, as a result of structural damage. This leads to morphological changes in the bronchial epithelium, progressing from basal cell hyperplasia to metaplasia, severe dysplasia to carcinoma *in situ*, and finally, carcinoma [13]. This series of pathophysiological changes is the primary reason for the development of lung cancer. Non-small-cell lung cancer (NSCLC) and small-cell lung cancers (SCLC) makes up the two subtypes of lung cancer, representing 85% and 15% of new cases each, respectively [5]. Due to the asymptomatic occurrence of lung cancer, diagnosis is often set at late in the cancer progression, when the cancer has metastasized.

Colorectal cancer

The most common form of colorectal cancer is adenocarcinoma, representing 98 % of cases. The corresponding 2 % consists of lymphoma, adenosquamous and squamous cell carcinoma [14]. Colorectal cancer has a well understood trajectory and dynamic progression. The disease begins as a benign adenomatous polyp, a growth that develops on the mucous membrane that lines the large intestine. The polyp develops into an advanced adenoma with high-grade dysplasia and then progress to an invasive cancer. The clinical behavior of CRC results from interactions at many levels and the multifactorial colorectal carcinogenesis drives initiation, promotion and progression. This dynamic process involves interactions between environmental influences, individual cancer susceptibility and accumulated somatic changes in the colorectal epithelium [15].

2.1.3 Prognostic staging

In order to understand cancer progression and survival, a staging system is often utilized. As the cancer progresses over time from initial etiology and pathophysiology, a staging system can differentiate between intermediate measures and survival rates that are distinct between groups. The most common notation system that describes the stages of cancer is the Tumor-Node-Metastasis (TNM) Classification of Malignant Tumors (TNM) [16]. The TNM staging system is used in all solid tumors, using the size and extension of the primary tumor, its lymphatic involvement and the presence of metastases to classify the progression of cancer. The staging parameters are summarized in Table 1.

Table 1: Staging parameters which composes the TNM Classification of Malignant Tumors[a] and the Dukes classification system[b] for colorectal cancer.

Tumor (T): Size or direct extent of the primary tumor	**Node (N):** Degree of spread to regional lymph nodes	**Metastasis (M)**: Presence of distant metastasis	**Dukes:**
T1 Contained within the organ with a diameter <3 cm	**N0** No regional lymph nodes metastasis	**M0** No distant metastasis	**A** Invasion into, but not through the bowl wall
T2 Between 3 cm and 5 cm in diameters	**N1** Regional lymph node metastasis present; tumor spread to closest or small number of regional lymph nodes	**M1** Metastasis to distant organs (beyond regional lymph nodes)	**B** Invasion through the bowl wall penetrating the muscle layer, but not involving lymph nodes
T3 Between 5 cm to 7 cm in diameter	**N2** Tumor spread to an extent between N1 and N3		**C** Involvement of lymph nodes
T4 The tumor has a diameter >7 cm	**N3** Tumor spread to more distant or numerous regional lymph nodes		**D** Widespread metastases

[a)] Tumor-Node-Metastasis (TNM) 8th edition [16]. [b)] Dukes classification system [17].

Lung cancer utilizes the TNM staging system, while CRC often use the TNM classification system in combination with more a specialized staging system, the Dukes classification system (Table 1) [17]. However, the TNM system is mostly used by physicians to record the anatomical extent of disease, and is often condensed into categories such as group or *stage*, most commonly by the American Joint Committee on Cancer (AJCC). The stage groups, also called Prognostic Stage Group, is adopted with the intention that categories within each group are more or less homogeneous in respect to survival, and that survival rates are distinct between the groups [18]. The combination and interaction of TNM staging system and the Prognostic Stage Group are summarized in Table 2. These staging groups are utilized worldwide, and clinicians in Norway handle cancer progression according to the TNM and AJCC staging system.

Table 2: The combination and interaction of TNM staging system[a], the Prognostic Stage Group (AJCC)[b] and the Dukes classification system[c].

AJCC stage	**TNM stage**			**Dukes**
Stage 0	Tis	N0	M0	
Stage I-A	T1	N0	M0	A
Stage I-B	T2	N0	M0	
Stage II-A	T3	M0	N0	B
Stage II-B	T4	N0	M0	
Stage III-A	T1-2	N1	M0	C
Stage III-B	T3-4	N1	M0	
Stage III-C	Any T	N2	M0	D
Stage IV	Any T	Any N	M1	

[a)] Tumour-Node-Metastasis classification system 8th edition. [b)] American Joint Committee on Cancer (AJCC) staging system. [c)] Dukes classification system for colorectal cancer. [17, 18]

2.1.4 Epidemiology

Lung cancer

Lung cancer accounted for approximately 9% of all cancer cases in Norway in 2016, and the mean age at diagnosis was 70.5 years. There has been an exponential increase in lung cancer incidence over time [19]. The rise in lung cancer incidence is attributed to two main causes. The first is an aging population, as the lung cancer development is positively correlated with age. The second is the increased of women diagnosed, which has compensated for the stabilization of men being diagnosed. The incidence of lung cancer for women and men is expected to be comparable in the future, as smoking habits have become similar between genders [19]. Figure 2 depicts the increase in lung cancer cases from 2000 to 2016. The total number of new cases was 3 206 in 2016.

Colorectal cancer

In Norway, a rapid increase in the incidence of CRC have been observed over the last decades, with the exponential growth occurring as far back as the 1960´s [20]. However, the rates for CRC appears to be stabilizing, especially among the younger generation [21]. According to global estimations by country, incidence rates in Norwegian women now rank second worldwide, although mortality rates for both genders are substantially lower in comparison with other countries. The incidence in Norway was 4 634 in 2016 (Figure 2) and the prevalence was 21 142. Deaths by CRC in 2014 was 1 147 [22].

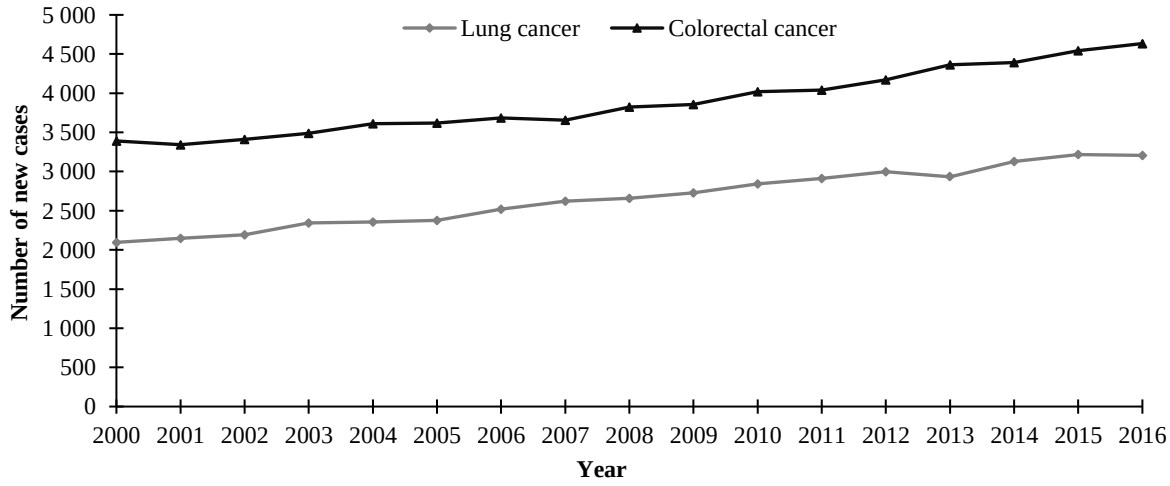


Figure 2: Number of diagnosed patients each year with lung or colorectal cancer in Norway between 2000 and 2016.

2.1.5 Diagnostic stage distribution in Norway

Determination of the correct diagnostic stage is one the most important factors for both an educated and correct decision regarding relevant treatment options, and with regards to predicting long-term survival for cancer patients. The CRN utilizes a crude condensed staging system for both lung cancer and colorectal cancer [23]. A descriptive table of coherency between the TNM staging system and the crude condensed system by the CRN is found in Table 6.

Lung and colorectal cancer

Figure 3 represents the distribution of diagnostic stage among newly diagnosed patients for lung and colorectal cancer in Norway from 2000-2016. Between 2000 and 2016 the total number of cases of cancer in the lung and colon/rectum diagnosed in localized disease was 16.4 % and 19.7 %, respectively. While the proportion of patients diagnosed with metastatic disease was 22.5 % for colorectal cancer and 44.8 % for lung cancer.

There is also a substantial number of cases which had an unknown stage at time of diagnosis. The reason for this could be both incomplete reporting from the hospital side, or that the reporting clinician does not have the necessary information at time of registration about the diagnostic stage of the patient.

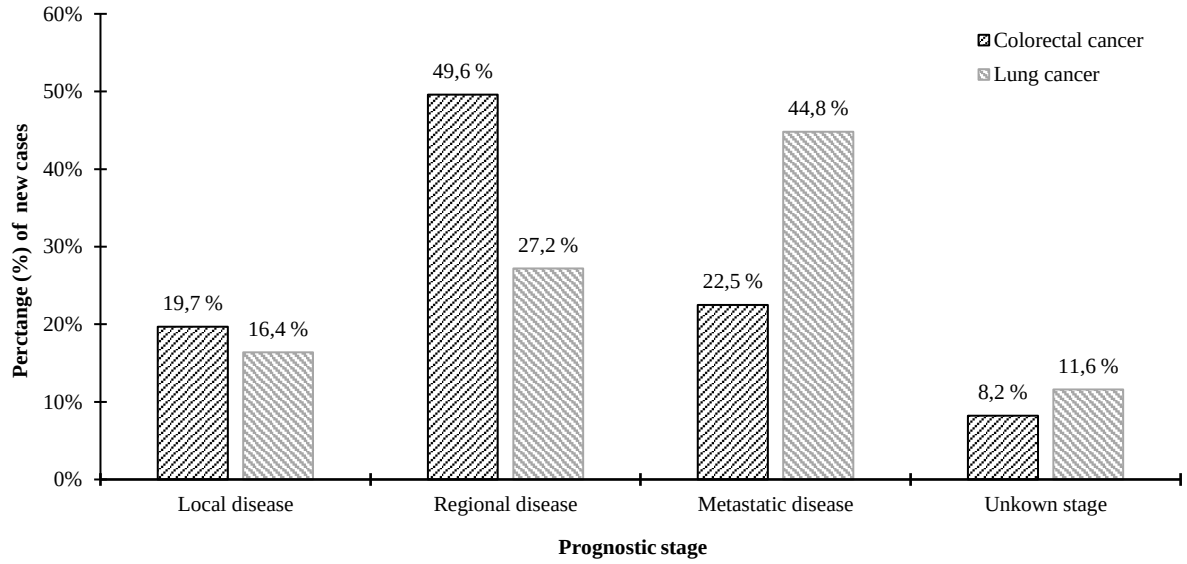


Figure 3: Distribution of diagnostic stages in Norway between 2000-2016 for lung and colorectal cancer. Percentage of new patients diagnosed in each stage, out of total number of new cases.

2.2 Clinical management

The clinical management and treatment options are determined by the national guidelines on cancer, requisitioned by the Norwegian Directory of Health (Hdir) [24]. The guidelines are continually revised to reflect the current best practice. The guidelines are roughly based on British Medical Journal Best Practice – Clinical Evidence (BMJ Best Practice) and adjusted to Norwegian conditions by expert panels and specialized physicians within each field.

2.2.1 Diagnostics

Lung cancer

Suspicion of lung cancer should be an overall assessment of the patient, including age and known risk factors (e.g. smoker). Symptoms include coughing, breast/shoulder pain, dyspnea, weight-loss, abnormalities in the thorax area, hoarseness, and finger-clubbing. On the basis of a clinical assessment, risk factors, and/or symptoms, the general practitioner (GP) will requisition a Computed Axial Tomography (CAT/CT) of the thorax from the specialized hospital in the region [25].

Colorectal cancer

Suspicion of CRC should be an overall assessment of the patient, age and known risk factors (e.g. diet, smoker, family history and various chronic bowl syndromes). Symptoms includes visible blood or mucus in the feces, discovery of tumor or polyps by a proctoscopy, or persistent change in regular bowl movement (> 4 weeks) for patients over 40 years of age. Polyps or dysplasia in the colorectal region is usually asymptomatic. The GP will under suspicion of CRC requisition a colonoscopy from the regional hospital [24].

2.2.2 Treatment strategies by diagnostic stage

After initial diagnostic procedures, the process of correctly identifying which stage the cancer has progressed is crucial. This could be done with either an MRI, CT, bronchoscopy and x-ray tests. Identification of cancer progression and staging is performed in the secondary healthcare system, primarily at first contact with the secondary health services, then again through biopsy after surgery (if surgical intervention).

Lung cancer

Current treatment options for lung cancer consists of surgery, neo/adjuvant therapy, chemotherapy, radiotherapy and immune therapy, often in combinations. Treatment alternatives are based mainly on stage, but other factors such as the patient's overall health, lung function and certain traits of the cancer itself are also important [26]. The complete recommended treatment guidelines by stage of diagnosis for lung cancer is presented in Appendix 9.2 Table A 1.

Colorectal cancer

Current treatment options for CRC consists of surgery, neo/adjuvant therapy, chemotherapy, radiotherapy and immune therapy, sometimes in combinations. Treatment alternatives are based mainly on stage and location in the colon or rectum, but other factors such as the patients overall health and certain traits of the cancer itself are also important [27]. The complete recommended treatment guidelines by stage of diagnosis for lung cancer is presented in Appendix 9.2 Table A 1.

2.2.3 Follow-up

Lung cancer

To identify possible recurrence or new primary tumor, regular controls after curative treatments is performed every 3-6 months the first 2 years, then replaced by yearly follow-up. The follow-ups consist of general anamneses, clinical examination, X-ray of the thoracic area and blood samples. These controls could be performed at the local hospital in coordination with the patients GP. The most important prognostic factor following lung cancer treatment is quitting tobacco products. It is estimated that about 40 % of lung cancer patients who regularly smoked before diagnosis quits after treatment [28].

Colorectal cancer

Patients diagnosed in stage I (Dukes A) which received surgery/resection of the colon are not routinely checked for metastasis, but are considered for a colonoscopy after 5-10 years if other risk factors such as age and overall health gives reason for this. Patients who have undergone local surgery, colonoscopy with removal of cancerous polyp, will be checked for positive

margins (the removed tissue is examined to identify if there are cancerous cells on the edges of the removed section) every six months for five year [29]. All patients diagnosed with stage II or III (Dukes B or C) who received curative treatment are routinely checked based on age and overall health [29]. In all patients who received curative treatment, a follow-up schedule is determined by the national guidelines and is presented in Table 3:

Table 3: Recommended guideline for follow-up of colorectal cancer where curative resection or oncologic treatment of recurrent or metastasis could be relevant. Recommended tests by months after surgery.

Months after surgery	1	6	12	18	24	30	36	48	60
CEA[a]	x	x	x	x	x	x	x	x	x
CT liver/abdomen		x							x
UL-liver with contrast			x	x	x	x	x	x	
Low dose CT of the thorax			x		x		x	x	x
Colonoscopy									x
Examination of the rectum/perineum[b]	x	x	x	x	x	x	x	x	x

[a)] Carcinoembryonic antigen (tumour marker test). [b)] Only applicable for cancer in the rectum. Source: [29]

2.2.4 Stage-specific survival

Lung cancer

A major international population-based study included nearly 60 000 patients from Norway and other selected western countries. The one-year overall survival for all stages in Norway was 38.4 % in the diagnostic period 2004-2007. When differentiating between prognostic stages the one-year overall survival was 74.9 % for localized disease, 47.4 % for regional disease, 18.3 % for distant disease and 42.2 % for unknown stage. When adjusting for age, the results were nearly identical [30, 31].

Colorectal cancer

A population-based study with Norwegian registry data included over 17 000 patients and calculated one-year net survival by stage diagnosed between 2000 and 2007. The one-year net survival for patients diagnosed with localized disease was 93.3 %, for regional disease it was 87.7 % and for patients diagnosed with distant disease the one-year net survival was 38.5 %. For patients with an unknown stage the one-year net survival was 65.4 % [31].

2.3 Healthcare provision in Norway

I Norway the government is responsible for providing health care to the population, in accordance with the fundamental concepts of equal and universal access, decentralization and free choice of provider. Primary health and social care is the responsibility of the municipalities, with the Norwegian ministry of health playing an indirect role through legislation and funding mechanisms. In specialized care, the ministry acts with a direct role through ownership of hospitals and its provision of directives to the boards of regional health authorities (RHA) [3].

2.3.1 Primary care

A patient's first contact during a medical situation is usually with the primary health care service and the GP. The GP acts as the gatekeeper for the secondary health care services, and it is often the GP who assess whether a patient is need for more specialized health services. The national guideline explicitly covers the GPs responsibly in referring the patient to specialized hospitals if there is a need for specialized treatment or any indication of cancer.

All individuals residing in Norway has a right to a GP through the national GP scheme, which entails all individuals being registered on a specific GP list. Individuals could either be registered at a specific GP, or at a GP office/practice (not have a specific doctor). Individuals registered at a GP office/practice might have a different doctor at each visit. In addition, these GP practices also have a high turnover of practicing GPs and doctors. The arrangement is voluntary, however about 99.6% of the Norwegian population are registered through the GP scheme, which assures a good coverage and access to primary health care services [3]. The GP has primary responsibility for the patients overall health, diagnostics and treatment.

2.3.2 Specialist, inpatient and somatic care

For medical episodes which require specialized treatment, patients are to be referred from the primary health care service, most often the GP. The secondary health services ensure the provision of specialized diagnostic services, treatment and follow-up for each patient. As increasing knowledge about different illnesses and disease have arisen, hospitals have become more specialized [3]. Due to this progression, all patients now have free choice of provider and can choose which hospital they would like to be treated, regardless of geographical affiliation.

There are four regional health authorities (RHA), which are responsible for supervising specialist inpatient somatic and psychiatric care. The ministry provides the RHAs´ budget and issues an annual document instructing the RHAs as to overarching aims and priorities. It is up to each RHA to determine how to distribute their resources among the regional facilities. The four RHA are comprised of 47 regional facilities within specialized health services [3]. Medical services of the highest complexity are provided in university hospitals and each of the four regions has at least one university hospital located in a large city connected to a university. Highly specialist care is therefore concentrated in urban (i.e. more densely populated areas) and people living in rural areas have to travel longer distances for access to certain specialists.

2.3.3 Palliative care

A number of official reports and national plans linked to cancer care and national cancer strategy to promote the gradual development of palliative care in Norway. The latest strategic policy document is the National Action Program for Cancer Palliative Care published in 2007. Palliative care services are provided at all levels of care. Patients staying at home receive services provided by the GPs and with municipal home care services. All main hospitals have multidisciplinary palliative care teams providing ambulatory services in hospital departments and surrounding municipalities. Specialist palliative care for patients with complex needs are centered in inpatient units in larger hospitals [3].

2.4 The cost of cancer

Cancer related costs make up almost 13 % of the total expenditures in the Norwegian health care system, summing up to over 14.5 billion Norwegian kroner (NOK) in 2014. The secondary health services and somatic hospitals accounted for around 90 % of cancer related costs. Primary care costs accounted for approximately NOK 500 million and costs for over the counter pharmaceuticals added up to about NOK 1100 million [32].

2.4.1 Stage-specific costs

While the association between earlier diagnostic stages at diagnosis have been widely recognized with positive correlation with regards to survival prospects, the influence of stage at diagnosis and its coherency with costs of resources employed in patient care and management remains somewhat unclear. Although the vast majority of patients will receive the same diagnostic procedures in the initial phase of hospitalization, after staging has been performed and the treatment pathway has been determined, the different stages and corresponding treatments will deviate from each other.

Lung cancer

Previous literature on stage-specific costs have placed special emphasis on phases of care, monthly costs and location of costs. The most recent study of stage-specific NSCLC costs using a state-transition model was published in 2004 in France, with cancer registry data for new cases of lung cancer between 1998 and 1999 [33]. The study separated lung cancer cases by five groups; *operability* (*local* and *regional*), *inoperability* (*local* and *regional*) and *distant*. Identification of costs were done through in individual study with 428 patients, with a male-female sex ratio of 4.66. The costs were gathered over an 18-month period, and the costs ranged from 20 691 $ (USD) for distant lung cancer to 27 794 $ for regional inoperable lung cancer. The mean cost for locally operable lung cancer was estimated at 25 050 $, with 27.2 % of the costs taking place in postoperative monitoring, 50.2% within the first-line treatment, 7.8% in second-line treatment, 3.6% in remission and 11.2% in palliative care [33]. The study did not include cost of diagnostic procedures, overhead costs or indirect costs.

A recent publication on using retrospective, descriptive analysis on resource use and a direct medical cost to analyze stage-specific lung cancer costs was conducted in Spain [34]. Their retrospective analysis estimated a mean cost per patient diagnosed with NSCLC ranging from 13,218 Euros in stage III to 16,120 Euros in stage II. The main cost components in mean cost per patient in NSCLC were chemotherapy (29.5%) and surgery (22.8%). The surgical and inpatient care cost represented 58.9% in stage I, while in stage II it was 45.9%, decreasing its relative weight as disease stage progressed. In addition, the relative chemotherapy cost increased in more advanced stages from 5.2% in stage I to 45.7% in stage IV [34].

Colorectal cancer

A 2014 German study estimated CRC treatment costs based on health insurance data from Germany [35]. The study focused on phase of care and estimated that mean costs in CRC patients peaked in the initial phase of care (€ 29,400) and the end-of- life phase (€ 64,600). In addition, the cost estimates ranged from € 15,000 to € 21,300 for early stage disease and from € 29,800 to € 35,000 for later stage disease [35].

A 2017 Spanish study conducted a retrospectively collected demographic data, clinical data and resource use of a sample of 529 patient. The estimated total cost per patient was € 8644 for stage I, € 12 675 for stage II, € 13 034 for stage III, and € 24 509 for stage IV. They estimated a total annual cost for colorectal cancer, when extrapolated to the whole Spanish health system, a total of € 623.9 million [36].

A recent study of stage-specific costs of CRC was analysed from a Norwegian perspective. The study analyses the cost and clinical pathway for CRC, using parameters from an observational study at the largest university hospital in Norway with a total of 2 049 patients [4]. The study combined parameters from several different sources compiling a model-based estimate for analysis of both costs and survival. The model estimated a cost range from € 23 386 (2011 Euros) for stage I CRC to € 61 396 for stage IV. However, some limitations were identified with the observational data ranging from 1993-2010, which could lower survival estimates explained by older and less effective treatments [4].

2.4.2 Economic burden on society

Increasing knowledge about cancer biology and the effect of competing treatments is resulting in better options of preventive measures and treatment [37]. While radical surgery was previously the predominant and perhaps the only life-prolonging treatment option for cancer patients, the rise of various combination treatments, chemotherapy, and lately the introduction of immunotherapy, has increased survival time remarkably. Especially in the last decade, innovative pharmaceutical treatments have dissipated highly effective treatments, often at a high monetary cost [7].

A large study of cancer costs in Norway from 2014 highlighted the burden of cancer, from both the perspective of patient, health care sector and society [32]. While the cancer related costs have increased in the secondary health services by 24.5 % from 2011 to 2014, the primary health services have increased by over 30 % (Appendix 9.2 Table A 2). This could be seen as a positive shift in demand on the primary health services. The desire for the primary health services to control access and strains on the secondary health services have been the wanted effect since the national coordination reform in 2001. However, as more strain is put on the primary health services, an increase in funding by the state must adjust for this shift in recourse utilization [32]. The study also investigated costs and resource use for specific cancer, including lung and colorectal cancer to investigate their economic burden on society. In Figure 4 the Diagnostic Related Group (DRG) related costs in somatic hospitals for lung and colorectal cancer are depicted to emphasize the magnitude of costs related to these two cancers. These cost estimates does not include costs to the primary health care sector, patients related costs or other societal costs [32].

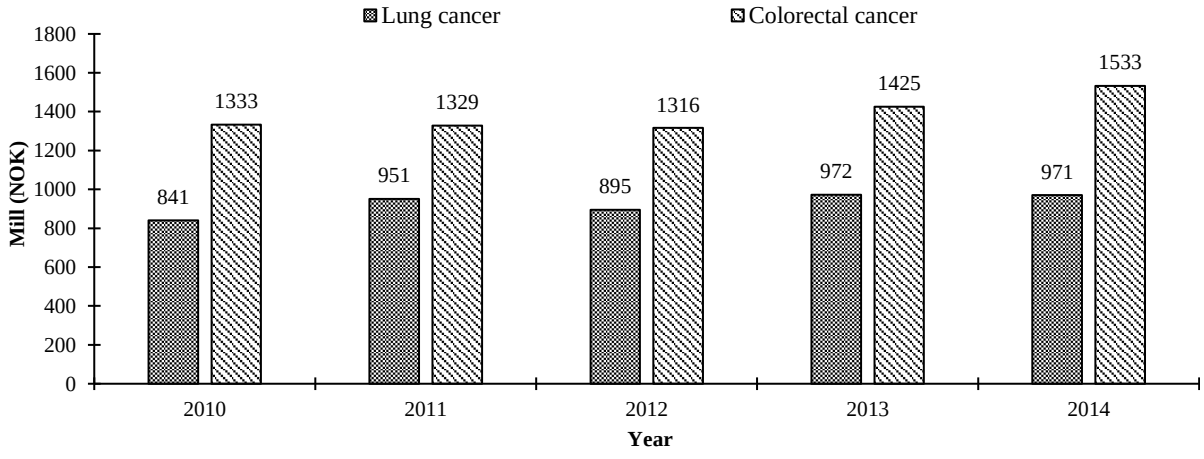


Figure 4: DRG-costs in somatic hospitals (nominal cost in mill NOK) for lung and colorectal cancer from 2010-2014 [32].

2.5 Identification of regional variations

2.5.1 Access

Access to health services has a major impact for both individuals and the population. Access to emergency health services contributes to substantial life-expectancy for the individual, and a major effect through somatic services on population level health measured in quality of life [3].

"*The values underlying the principles for the Norwegian health care sector are rooted in the following fundamental view: Each individual has an inviolable intrinsic value regardless of gender, religion, socioeconomic status, level of functionality, relationship status, and place of residence or ethnic background. The population must have equal access to health care services. Similar cases must be treated in the same way" Norwegian Ministry of Health and Care Services, 2017 [38]*.

Through various legislation, these fundamental principles ensure the population access to health care services when they are required.

Regional difference in access to primary healthcare

Table A 3 in Appendix 9.2 and Figure 5 represents a review of the GP scheme in 2016, conducted on behalf of the Norwegian Directorate of Health, identifying trends and tendencies, which could indicate regional differences in access and utilization of primary care services [39]. Figure 5 depicts numbers of GP with diagnostic, rehabilitation and treatment responsibility per 10 000 inhabitants in each county in Norway in 2016. The highest share of GPs is found in Finnmark, with 13.6 per 10 000 inhabitants. Hordaland however had almost half that, with 7.5 GPs per 10 000 inhabitants. In addition, a general population survey conducted in 2010 revealed that the proportion of population reported having longer waiting times to the GP was higher in Norway compared to other European countries [3]. However, there are also distributional differences of doctors within the RHA in each health region, which are presented in Table A 4 in Appendix 9.2. Per 2018, Helse Nord RHA had the highest concentration of doctors in specialized health care with approximately 31 per 10 000 inhabitants, while Helse Vest RHA had the lowest with 21 per 10 000 inhabitants. There was a total of 475 doctors with specialization within oncology in Norway in 2017, distributed among the four RHAs (47 regional facilities) [39].

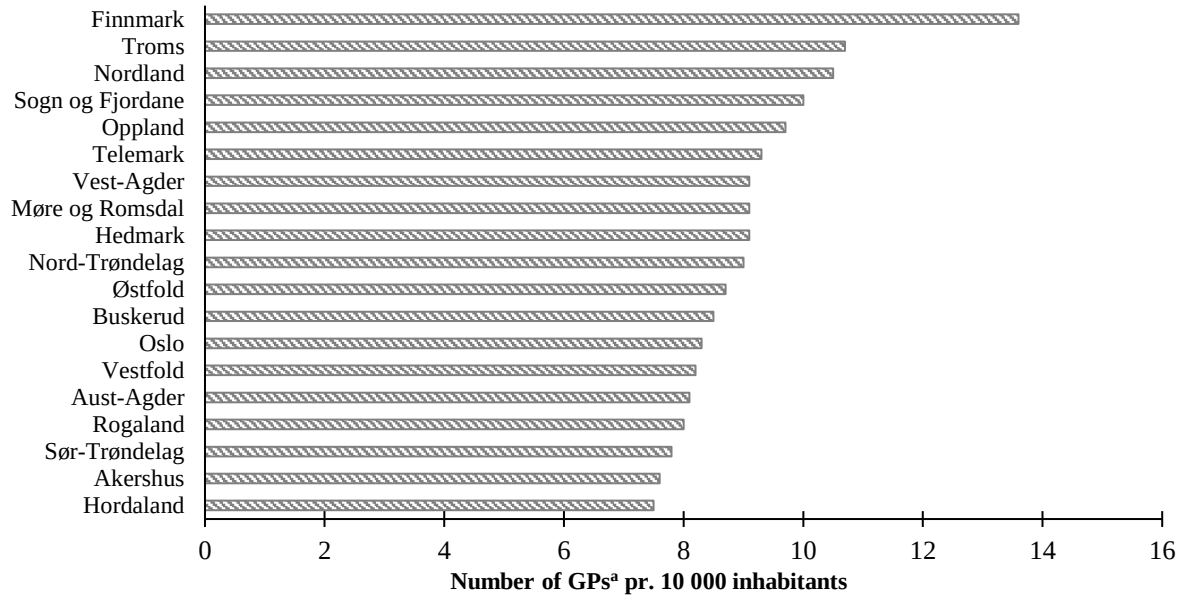


Figure 5: Number of [a)] General Practitioners (GPs) in full time equivalents, with diagnostic and treatment responsibility pr. 10 000 inhabitants in each county in Norway in 2016. Source: SSB table 03807.

2.5.2 Regional difference in prognostic stage distribution

As before mentioned, the determination of a correct stage is one the most important factors regarding relevant treatment options and with regards to long-term survival for cancer patients. Regional differences in prognostic stage at point of diagnostic could therefore be an indicator for regional differences with respect to survival. It is a desire to diagnose the highest percentage of numbers in the lowest grades of severity (e.g. localized) and lowest percentage in the higher grades (e.g. distant). Another interesting factor is the amount of cases being registered as unknown stage.

Lung cancer

In Figure 6, the "worst" and "best" counties are depicted, representing the respective edges of the scale with regards to diagnostic stages. Sogn og Fjordane has both the lowest share of patients diagnosed in localized disease and the highest share diagnosed in distant disease. Sogn og Fjordane also has the highest percentage of patients diagnosed with regional disease. The county with the highest share of patients diagnosed in localized disease is Troms, with 19.3 %. Telemark has the lowest share of patients diagnosed with distant disease, with 40.2 %. Nord-Trøndelag only has 7.1 % of patients diagnosed with unknown diagnostic stage, while Hedmark has the highest share with 18.4 %. In Table A 10 in Appendix 9.5 all counties and their respective proportion of patients diagnosed in each stage of lung cancer.

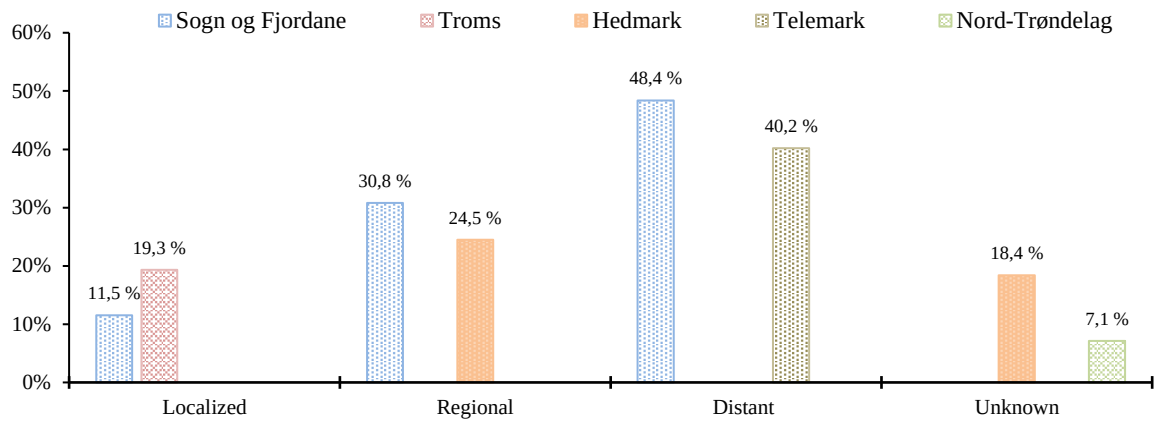


Figure 6: Proportion (%) of lung cancer patients diagnosed cases in each diagnostic stage by average in Norway, and the highest and lowest share of new cases in each diagnostic stage by county.

Colorectal cancer

As in Figure 6, Figure 7 represents the case of colorectal cancer, and the six counties depicted in the figure represents the respective edges of the scale. Telemark has both the highest share of patients diagnosed in localized and lowest for regional disease. The county with the lowest share of patients diagnosed in localized disease is Aust-Agder, with 16.6 %. Oppland has the highest percentage of patients diagnosed with distant disease (25 %), and Hordaland has the lowest percentage (20.7 %). Akershus has the highest share of patients diagnosed with regional disease and the lowest share of patients with unknown stage. Akershus only has 6.4 % cases with unknown diagnostic stage, while Oslo has the highest share with 10.2 %. Compared to patients with lung cancer, a substantially higher proportion of colorectal cancer patients are diagnosed with regional disease. In Table A 11 in Appendix 9.5 all counties and their respective proportion of patients diagnosed in each stage of colorectal cancer.

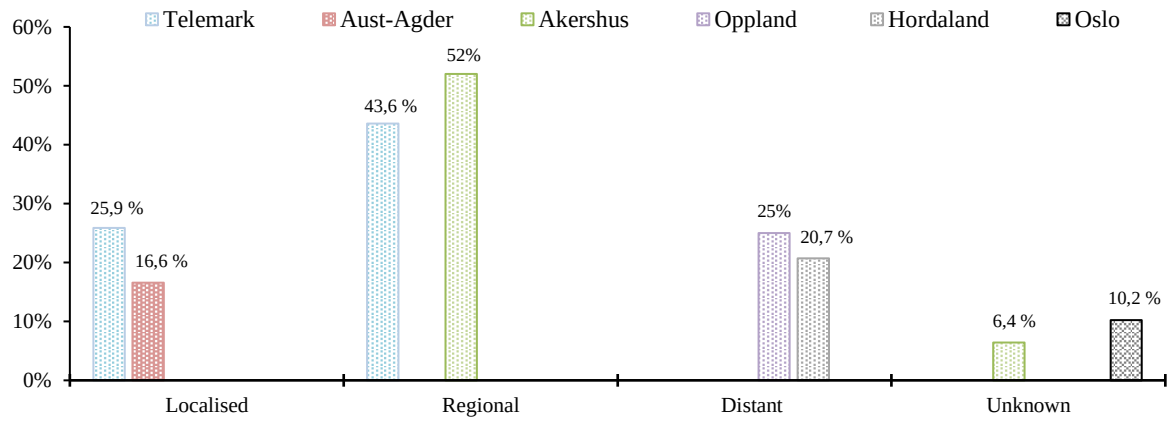


Figure 7: Proportion (%) of colorectal cancer patients diagnosed cases in each diagnostic stage by average in Norway, and the highest and lowest share of new cases in each diagnostic stage by county.

2.5.3 National strategies and initiatives for cancers

In 2013, a new national strategy against cancer was constructed. “Together against cancer – National cancer strategy 2013-2017” aimed at facilitating for an increased focus on cancer treatment. Five main target areas were defined; (1) A more user oriented cancer care; (2) Norway is to be a pioneer in good patient progress; (3) Norway is to be a pioneer for preventive measures for cancers; (4) More patient will survive and live longer with a cancer diagnosis; (5) The highest quality of life for both patients and dependents [40].

The overarching aim of these national strategies is to ensure equal and quick treatment for all citizens across Norway and eliminate any possible geographical variation, in line with the fundamental principles of the Norwegian healthcare system. As a result of the above mentioned national strategy the standardized patient trajectory was developed, and the national guidelines and the national screening programs was emphasized more heavily.

National guidelines

The national guidelines were established as an instrument to secure good quality services and coherent treatment pathway, making correct prioritizations, and avoiding unwanted variation of services. These national guidelines are an expression of what is considered best (current) practice, and expert panels and service recipients are actively involved in the preparation and construction of these guidelines [25]. Hdir is the governmental authority responsible for constructing and revising national guidelines concerning all cancers and contains professional guidelines on diagnostics, treatment and follow-up. The national guidelines contains normative timelines from assessment and treatment of cancer in the secondary health services. It states that a patient shall be considered within 5 working days, clinical assessment within 10 days, and treatment start within 20 days after a referral from the GP [40].

Standardized patient trajectory/pathway

In 2014, the Norwegian government assigned Hdir with the task of developing a standardized treatment pathway for cancers. These, together with the national guidelines was meant as a tool for GPs and was implemented through 2015. The aim was to clarify the reason behind cancer suspicion, for the GP to establish a diagnosis and referral to specialized care.

The standardized treatment pathways are a package patient course that includes all encounters with the health care system, from referral to follow-up. Prolongation times are determined on the basis of expert opinion on acceptable waiting time between the various measuring points in the patient trajectory. These pathways were adapted from a Danish model, which indicated positive experiences from both patients, dependents and GPs. I Norway, as in Denmark, the regional health authorities publish quarterly reports. This gives each hospital the opportunity to identify where in the treatment pathway unwanted issues arise and make the appropriate adjustments [41].

Colorectal cancer screening

After extensive international and national research on the effectiveness of colorectal cancer screening, and a pilot project conducted in 3 counties with more than 140 000 individuals, the Norwegian Ministry of Health ultimately proposed an introduction of a national colorectal screening program. It is set to be nationally introduced with the first invitations to participate being sent out through 2019. The invitation will be sent out to individuals, both men and women, the year they turn 55 years old. Initially the test will be a fecal immunochemical test (FIT) which tests for hidden blood in the stool, which can be an early indicator for colorectal cancer [20].

Lung cancer screening

To date, there is no national screening initiative for lung cancer in Norway. The overall assessment by the Ministry of Health, Hdir and expert panels is that a national lung cancer screening pilot/program should be awaited until more and better documentation is available. Especially the identification of an appropriate target population is the main. This is necessary to yield the best outcomes for the individual patient, the respective level of service and assessment of resources. A study requisitioned by the national health authorities is expected by the end of 2019 and will provide knowledge on whether a national lung cancer screening program is effective with consideration to both cost and effect [40].

3 Data, materials and inputs

3.1 Cancer Registry of Norway

To estimate survival and stage distribution, unlinked registered data from The Norwegian Cancer Registry (CRN) were collected for cancer in the lung and colon/rectum. The first section will provide a description of the CRN dataset used in the survival analysis together with assumptions made for the collected variables and eventual coding procedures. The CRN is a national cancer registry established in 1951 and is one of the most extensive national cancer registries in the world. Whenever an individual is diagnosed with cancer, either while alive or postmortem, the doctors are bound by legislation to report it to the CRN [2].

Data was received from CRN for all individuals diagnosed in the period 2000-2016 with the following ICD-10 codes as main diagnosis: C18 - Malignant neoplasm of the colon; C19 - Malignant neoplasm of the rectosigmoid junction; C20 - Malignant neoplasm of the rectum; C34 - Malignant neoplasm of the bronchus and lung. All cases for patients with ICD-10 C18, C19 and C20 were condensed to a joint colorectal cancer variable. Other than the main cancer diagnosis, variables used in the main analysis are presented in Table 4:

Table 4: Variables received for the Cancer Registry of Norway included in the main analysis

Variable	Type of variable
Patient ID	Numerical – Continuous
Gender	Categorical – Dichotomous
County of residence	Categorical - Nominal
Year of birth	Numerical – Continuous
Date of diagnosis	Interval – Continuous
Metastasis	Categorical – Ordinal
Date of death	Interval – Continuous
Cause of death	Categorical - Nominal

3.1.1 Patient population

Ensuring data quality

For the variables of interest presented in Table 4 numerous steps were taken to ensure the quality, accuracy and coding of the data, which is described below.

The data was checked by ID records, in order to identify whether any patients had multiple records. If there were multiple records for identical patient ID, they were visually checked as to whether it was a simple duplication or if they differentiated with consideration to other variables. If the record was simply duplicated, they were removed from the analysis.

From the total cases originally diagnosed with ICD-10 C34 and C18-C20, cases notified to CRN postmortem were excluded from the survival analysis as survival time for these patients would effectively be zero. This was also the case for patients who had delayed diagnosis after death, which can occur whenever the CRN and the Norwegian Death Registry are linked and updated. The survival time for these patients were negative and removed from the analysis.

Preparation

To ensure patient anonymity the variables for date of diagnosis and date of death were given in month/year format, not including the specific date. This however created a survival time of zero for those patients who died within the month they received the diagnosis and would therefore indicate death on the same day the patient received the diagnosis, which is not probable for most of these patients (e.g. not postmortem). To ensure that that survival time for these patients would be included in the analysis a date variable was included. These patients should implicitly have a survival time of more than one day, yet less than 30 days. This was done for all patients in the dataset as to not differentiate between those with survival time <1 month and >1 month. Diagnosis date was therefore assumed to be the 1st day of the month, and date of death on the 15th of each month, potentially adding two weeks of survival time to every patient.

Staging procedure

The CRN utilizes a crude condensed staging system for both lung cancer and colorectal cancer, based upon the parameters in the TNM system. This entails that the analysis will be performed with the condensed CRN staging system, expressed by Table 5. Survival and incidence rates by stage for cancer of the lung and colon/rectum types was identified through coding of CRN data by the above mentioned variables and parameters in Table 5 [42].

Table 5: CRN[a] parameters for determination of diagnostic stage.

Diagnostic stage	Parameters
Localised disease	No direct growth into surrounding tissue/organs, lymph nodes or organ metastasis Micro-invasive growth, carcinoma initial infiltration
Regional metastasis	Lymph node metastasis within the primary anatomic section Microscopic growth into surrounding tissue structure
Distant metastasis	Lymph node metastasis outside the primary anatomic section Organ metastasis within the primary anatomic section Organ metastasis outside the primary anatomic section
Unknown	Identified metastasis, unknown location Unknown propagation at time of diagnosis

[a)] Source: *[42]*

Population characteristics

Selected patient characteristics for patients diagnosed with lung cancer are presented in Table 6. There was a total of 45 278 patients diagnosed with lung cancer between 2000 and 2016. Table A 5 in Appendix 9.3 provides a more detailed summary of the patient characteristics, separated by diagnostic time-period and incidence in each county.

Table 6: Patient characteristics. Patients diagnosed with lung cancer in the time period 2000-2016 by gender, age-group and diagnostic stage.

	Males		**Females**		**Total**
	N	%	N	%	N
Age					
0-49	697	48.5	739	51.5	1 436
50-59	2 956	51.9	2 744	48.1	5 700
60-69	7 509	56.6	5 762	43.4	13 272
70-79	9 126	58.5	6 481	41.5	15 608
80-89	5 023	59.8	3 372	40.2	8 396
90 and over	476	55	390	45	866
Stage					
Localised	3 902	52.5	3 524	47.5	7 426
Regional	7 314	59.4	4 996	40.6	12 310
Metastatic	11 570	57	8 721	43	20 293
Unknown	2 999	57.2	2 244	42.8	5 244
Total	**25 787**	**57**	**19 488**	**43**	**45 278**

Patient characteristics for patients diagnosed with colorectal cancer are presented in Table 7. There were a total of 66 059 patients diagnosed with CRC between 2000 and 2016. Table A 6 in Appendix 9.3 provides a more detailed summary of the patient characteristics, separated by diagnostic time-period and incidence in each county.

Table 7: Patient characteristics. Patients diagnosed with colorectal cancer in the time period 2000-2016 by gender, age-group and diagnostic stage.

	Males		Females		Total
	N	%	N	%	N
Age					
0-49	1 656	49.3	1 726	50.7	3 362
50-59	3 834	53.9	3 279	46.1	7 113
60-69	8 437	56.2	6 569	43.8	15 006
70-79	10 750	52.9	9 588	47.1	20 338
80-89	7 574	43.9	9 672	56.1	17 246
90 and over	1 008	33.9	1 966	66.1	2 974
Stage					
Localised	6 520	49.8	6 520	50.2	12 999
Regional	16 434	50.1	16 346	49.9	32 780
Metastatic	7 728	52.1	7 097	47.9	14 825
Unknown	2 665	49	2 770	51	5 435
Total	**33 259**	**50.3**	**32 800**	**49.7**	**66 059**

3.2 Probabilities

The lung cancer pathway modelling required transitions probabilities estimated from patient-level data. Such data would typically come from some longitudinal study (e.g. trial or cohort study) [43]. The following section explains the collection of transition probabilities for the lung cancer pathway modelling;

Lung cancer transition probabilities

Probabilities for the lung cancer model was derived from various meta-analysis and systematic reviews [37, 44-61]. To establish model parameters, systematic searches were performed in order to identify relevant publications. This was done through the PubMed, Cochrane Library, Elsevier and EBSCO. To identify relevant publications through a strategic search, the following keywords were used: “stage-specific”, “survival”, “clinical pathway”, “Recurrence/remission”, “treatment”, “palliative” in combination with “lung cancer”. Furthermore, studies were assessed based on relevance, publication date, appropriateness to a Norwegian setting/comparability and the overall data strength (e.g. sample size and relevant follow-up).

Norwegian publications were considered most relevant for the model input. Where no Norwegian publications identified for the required parameters, only international comparable publications were considered.

Decision tree

Table A 7 in Appendix 9.3 summarizes parameters from the treatment specific decision tree and are mutually exclusive probabilities to establish movement through the different primary treatment strategies/pathways and entry into the different Markov models. These are presented as non-time-dependent transition probabilities.

State-transition Markov model

Transition probabilities for the deterministic model were derived mainly from meta-analysis and systematic reviews. Traditional strategies in lung cancer treatment have been vigorously studied, resulting in large number of available systematic reviews and meta-analysis (e.g. surgery and radiation). However, certain model inputs were derived from single studies. This was most relevant for the palliative treatment probabilities, as there is a major turnover in what is currently best practices, and the continuing introduction of new life-prolonging pharmaceuticals. The majority of publications had survival as the main endpoint, and studies which focused on stage-specific survival was of particular interest. In addition, studies which included probabilities of distant and local recurrence, and remission after secondary treatment were identified in order to establish the movement through the state-transition model.

Due to the one-month cycle length of the state-transition model, all probabilities related to transition through the model were converted to monthly transitions (unless already presented as monthly probabilities). Probability is the measure of the likelihood that an event will occur. The event can occur at any point in time and is quantified as a number between 0 and 1. While a probability is restricted in this interval, a rate is expressed as on a scale from 0 to infinity and is the instantaneous potential for an event to occur [43]. Conversion of probabilities and rates were performed by the standard method proposed by Drummond et.al [62]. A probability presented with an inappropriate time horizon can be recalculated by first converting the probability to a constant instantaneous rate. This is done by deriving the following expression:

$$r = -\frac{[\ln(1-p)]}{t} \qquad (1)$$

Where r is the rate, p is the probability over time period t. It is then possible to convert the calculated rate to a probability over a given time period. This presents the probability as a function of rate:

$$p = 1 - exp(-rt) \qquad (2)$$

Table A 8 in Appendix 9.3 summarizes parameters from the Markov model and their monthly transition probabilities, derived as previously mentioned. The presented probabilities were used to estimate rates and conditional probabilities, and to establish transitions between health states in the model.

3.3 Costs

Lung cancer costs

Cost estimates was derived mainly from DRG-weights and the corresponding DRG unit-price, in addition to various literature sources. Systematic searches were performed in order to identify relevant publications. This was done through the PubMed, Cochrane Library, Elsevier, EBSCO and Oncolex[1]. To identify relevant publications through a strategic search, the following keywords were used: "stage-specific", "stage" "prognostic", "costs", "clinical pathway", "recurrence/remission", "treatment", "palliative", in combination with "lung cancer" or "NSCLC". All cost parameters from the pathway model is presented in Table A 9 in Appendix 9.3.

For costs identified through the DRG system, the standard method of multiplication of the DRG-weight and corresponding DRG unit-price was performed. All calculations performed on DRG used the 2018-unit price, set at NOK 43 428 for somatic services:

$$= 43\,428 \times DRGweight \qquad (3)$$

Adjustment for costs estimates

All costs which were not in NOK were adjusted by the purchasing power parity (PPP) method, which is a method of that states that the exchange rate between two countries is equal to the ratio of the currencies respective purchasing power. PPP exchange rates help costing, but excludes potential profits, which is not considered a major issue for health service unit costs. The PPP exchange rate calculation is expressed as:

[1] Oncolex is a Norwegian encyclopedia for oncology health personnel (www.oncolex.org). The website is continuously updated for both treatment and costs for cancer in Norway.

$$\frac{S_1}{S_0} = \frac{(1 + I_y)}{(1 + I_x)} \qquad (4)$$

Where S_0 is the spot exchange rate at the beginning of the time period and S_1 is the spot exchange rate at the end of the time period. I_y is the expected annualized inflation rate for country y, which is considered to be the foreign country and I_x is the expected annualized inflation rate for country x, which is considered to be the domestic country. For Norwegian costs estimates, inflation adjustments were performed.

3.4 Ethical considerations

Data approval was necessary from the CRN in order to conduct the analysis using their respective dataset. The data was received with anonymized patient IDs, and the dataset was not linked with any other registries. However, since the data consisted of health registry data (Helseregisterloven § 29), a notification was necessary from The Norwegian Data Protection Authority (DPA). Approval from both DPA and CRN was received in February 2018 (Appendix 9.1).

Data extraction from the CRN was funded by the Norwegian Cancer Society. The funding source had no effect on the choice of methodology, interpretation or presentation of the data, and are the sole responsibility of the author.

4 Analysis and methods

The theoretical framework behind survival analysis will initially be described, followed by the methods applied to estimate survival. The application of statistical tests and regression methods, separated by non-parametric and parametric survival models are then described. First, the survival estimates will be described using univariate and non-parametric approach. Multivariate data analysis and the concept of parametric survival models will then be described and the corresponding reasoning behind the application of these analysis.

The rationale behind the assessment of potential health gains by assessment of prognostic factors included in the univariate and multivariate data analysis will then be described.

Modelling of the clinical pathway for lung cancer treatment will then be described, with presentation of the model structure and important modelling concepts. The main purpose of the modelling exercise is to estimate the distribution of costs between the different stages. Health care related costs associated with each stage of cancer for colorectal cancer in Norway will be estimated from previous research, with regards to cost variation between diagnostic stages.

In order to examine financial gains, the method used for weighting stages and the corresponding costs will be presented, together with the other important adjustments. The cost estimates will provide the framework for the main sub-analysis to examine potential economic gains by achieving earlier diagnosis.

Finally, some important key assumptions will be addressed.

4.1 Survival analysis

When there is a desire to measure a clinical outcome, such as time until a specific event, survival analysis is a useful tool. Survival analysis explores groups which are suspected to vary with consideration to, for example, survival. Registry data is suitable for this type of analysis, if the desired event is recorded to have occurred at a specific time-point. There are several types of survival analysis that express the outcome of interest in the dataset and the main difference is what is considered the end-point is of interest [63]. In *overall survival,* the event of death is under interest. In analysis with overall survival as the main end-point there is no separation of between death due to the disease or other causes. Due to possible misconceptions and reporting-

error, overall survival is the main endpoint in this analysis, because death caused by disease related factors (comorbidities due to cancer or cancer treatment) might not be included when only assessing disease-specific survival. When comparing survival distributions for two or more groups, such as survival for different stages of cancer, both parametric and non-parametric methods can be applied.

First, some important and general concepts encountered in survival analysis will be explained, in order to comprehend the function behind survival analysis. Furthermore, the different methods and approaches will be discussed for consideration to the statistical analysis.

4.1.1 The survival and hazard function

When estimating survival data, we generally describe and model in terms of two related functions, *survival* and *hazard* [63]. The survival function or probability, denoted *S*(*t*), estimates the probability that a patient survives from the time point of origin, the cancer diagnosis, to a specified time point, time *t*.

$$S(t) = \mathrm{P}(T > t) = 1 - F(t) \qquad (5)$$

The time variable is fundamental to survival analysis because survival probabilities for different values of *t* provides crucial summary information from time to event data. The *T* is assumed to be a continuous random variable with the probability density function *f(t)* and cumulative density function $F(t) = Pr\{T<t\}$, giving the probability that an event has occurred by duration *t*. The T could be thought of as the *waiting time* until the occurrence of an event. Then the event of interest is *death* and the waiting time is therefore *survival time*. The complement of this gives the probability of surviving for a period of time greater than *t*, expressed as *1-F(f)* [64]. These values describe directly the survival experience by the patient within the dataset.

The hazard function, denoted *h(t)* when the hazard is a rate, is the probability that an individual who is under observation at time t has an event at that time, conditional on having survived to time *t* [43]. It represents the instantaneous event rate for an individual who has already survived to time *t*, formally expressed as:

$$h(t) = \lim_{\delta t \to 0} \frac{P(t \leq T < t + \delta t | T \geq t)}{\delta t} \qquad (6)$$

The numerator is the conditional probability that the event will occur in the interval [*t*, *t*+δt), given that it has not occurred before, and the denominator is the width of the interval. Dividing one by the other we obtain a rate of event occurrence per unit of time. Taking the limit as the width of the interval goes down to zero, we obtain an instantaneous rate of occurrence. The conditional probability in the numerator may be written as the ratio of the joint probability that *T* is in the interval [*t*, *t*+δt) *and* $T \geq t$ (which is, of course, the same as the probability that t is in the interval), to the probability of the condition $T \geq t$. The former may be written as $f(t)\delta$, while the latter is *S(t)* by definition [63]. Dividing by δ and passing to the limit using the exponential density function, gives the hazard rate function expressed as:

$$h(t) = \frac{f(t)}{S(t)} \qquad (7)$$

In words, the rate of occurrence of the event at duration *t* equals the density of events at *t*, divided by the probability of surviving to that duration without experiencing the event. The hazard function is of interest because it provides insight into the conditional failure rates and provides a vehicle for specifying a survival model.

4.1.2 Censoring

Censoring is an important concept in survival analysis. It is a form of missing data problem, which is common when analyzing survival data. Censoring occurs when the total survival time for the patient cannot be correctly determined [63]. When patients whose time of diagnosis is known, but who are still alive after the observation period ends, they are said to be *right-censored*. Left censoring occurs when the patient has exhibited the event (e.g. death) at the start of the study and information about when the patient experienced the event is unclear. It is common to encounter right-censored data when analyzing survival time in register data.

If censoring occurs during the observation period, such as a patient dropping out or there is a lost to follow-up, it will affect the survival rates. In national register data, this is usually due to movement out of the country. Censored events during the observation period that coincide with an event are usually considered to fall immediately after the event. Censoring removes the patient from the denominator (displayed as patient at risk of dying) [63]. However, the patient adds information through the time of observation until censored, and therefore contributes to the survival probability as long as possible.

4.2 Univariate data analysis

4.2.1 Non-parametric survival estimation

The Kaplan-Meier (KM) estimation and the log-rank test are the standard non-parametric methods of survival analysis, and by definition assumes no specific distribution of parameters [63].

In preparing Kaplan-Meier survival analysis, each patient is characterized by three variables: 1) their serial time (time spent in the observation period), 2) their status at the end of their serial time (whether or not the event (say death) occurred) and 3) their corresponding group. For the survival time probabilities and curves to be constructed, patients are arranged according to serial time, regardless of time at diagnosis during the observation period. By constructing the data in this approach, all patients begin at the same time point and are followed through time until an event [63]. The two outcomes that can occur is 1) the patient has experienced the event of interest (e.g. death) or 2) they are censored.

Within the patient group under observation, a number of patients (k) experience an event in the follow-up period at distinct time-points $t_1<t_2<t_3<\ldots<t_k$ [63]. As the event *death* is assumed to occurs independently of one another, the probabilities of surviving from one time-interval to the next may be multiplied together to five the cumulative survival probability. This yields the $S(t)$ from Equation 5 and is constant between times of events, and therefore the estimated probability is a step function that changes value only at the time of each event. This estimator allows each patient to contribute information to the calculations for as long as they are event-free (e.g. still alive) [63].

When analyzing Kaplan-Meier survival curves, occurrence of the event of interest and unit of measurement is essential, along with shape of the curve. Curves that have wide steps with sudden drops indicates a small number of observations. While curves that have a large number of small *steps* signifies a large number of observations (e.g. patients). In national register datasets, there is often a large number of observations, indicated by a large number of small *steps* [63]. The Kaplan-Meier survival curve provides a useful summary of the data that can be used to estimate median survival. The median survival is often used as survival analysis is often characterized by a large skewness, and the median is generally a better measure of central location than the mean [63].

4.2.2 Non-parametric tests for survival comparison

The log-rank test is used for comparison of survival between different stages of cancer. This method calculates at each event time, for each group (e.g. stage), the number of event one would expect since the previous event if there was no difference between the groups. These values are then summed up over all event times (E_i) to give the total number of events in each group (i) [63]. The log-rank test compares observed number of events (O_i) for each group (i), to the expected number by calculating the test statistics:

$$X^2 = \sum_{i=1}^{g} \frac{(O_i - E_i)^2}{E_i} \qquad (8)$$

This value is compared to a chi-square (χ^2) distribution with (g-1) degrees of freedom, where g is the number of groups. The p-value may be computed to calculate the statistical significance of the difference between the survival curves. The log-rank test was used for comparison of survival between stage of cancer, county of residence, gender and age. A hazard ratio based on the log-rank test however, is only applicable when just two groups are compared (e.g. gender)

4.3 Multivariate data analysis

When comparing several prognostic factors in terms of survival, it is often sensible to adjust for patient-related factors, known as covariates or confounders, which could potentially affect the survival time of a patient [65]. Multiple prognostic factors can be adjusted for using multivariate modelling. The principle strength of these statistical models is their ability to asses several covariates simultaneously [43].

4.3.1 Semi-parametric survival estimation

The Cox proportional hazard (PH) model is the most commonly used multivariate approach for analyzing survival time data in medical research. The Cox PH model is a regression model which describes the relation between the event, as expressed by the hazard function and a set of possible covariates [65]. It takes a “semi-parametric” approach, assuming no set distributional shape, which is an advantage of the Cox PH model. The Cox PH model is expressed as:

$$h(t) = h_0(t) \times exp\{b_1 x_1 + b_2 x_2 + \cdots + b_p x_p\} \qquad (9)$$

Where the hazard function *h(t)* is determined by a set of p covariates (x_1, x_2,…, x_p). The impact of these covariates is measured by the size of their respective coefficients (b_1, b_2… b_p). The initial term h_0 is referred to as the baseline hazard, and is the value of the hazard if all the x_i are equal to zero. The t serves as a reminder that the hazard may, and probably will, vary over time.

4.3.2 Parametric survival estimation

Parametric PH models are a class of models similar to the Cox PH, both with regards to interpretation and the overall concept. The most obvious distinguishing feature between parametric models is the shape of the hazard they assume the data follows. There exists a substantial number of probability distributions, but for this analysis one of the most important properties is a continuous distribution supported on semi-infinite intervals (0,∞) and their appropriateness to survival analysis. When assigning a parametric distribution to the data, there are several considerations and assumptions which needs to be checked [65].

Some parametric models popularly used in survival analysis have been immediately excluded from analysis and includes the Log-Normal, Log-logistic, and Generalized Gamma models which are preferable when the hazard rises to a peak before decreasing, which is not applicable in this analysis [65]. This leaves two distributions which are appropriate within this framework, Exponential and Weibull. To adjudicate the best model fit, the appropriateness of the models will be tested by several methods:

1. Informal assessment: Plotting the smoothed empirical hazard or cumulative hazard against those estimated by the models.

2. By log(-log(survival)) survival plots

3. By Akaike's Information Criterion (AIC): A statistic that trades off a models likelihood against its complexity, given by:

$$AIC = -2 \ln L + 2(c + a) \qquad (10)$$

Where L is the models likelihood (log-likelihood), c is the number of covariates and s is the number of ancillary parameters. A lower value of the AIC suggests a better model.

4. By Bayesian Information Criterion (BIC): a statistic closely related to AIC, under the assumption that data distribution is in the exponential family, given by;

$$BIC = -2 \ln L + k\ ln(n) \qquad (11)$$

Where L is the models likelihood (log-likelihood) and k is the number of free parameters to be estimated. If the estimated model is a linear regression, k is the number of regressors, including the intercept. A lower BIC value suggests a better model fit.

Both the AIC and BIC penalizes the number of estimated parameters and rewards the likelihood function (goodness of fit), discouraging overfitting opposed by the goodness of the fit [65].

The most distinguishing feature between the two parametric models is the hazard function. The main drawback of parametric models is the need to specify the distribution that most appropriately mirrors that of the actual survival times. However, where a suitable distribution can be found, the parametric model is more informative than the Cox model. It is straightforward to derive the hazard function and to obtain predicted survival time from parametric models [65]. Additionally, the appropriate use of these models offers the advantage of being slightly more efficient, as they yield more precise estimates (e.g. smaller standard errors).

4.4 Estimating health gains

When assessing the impact of diagnostic factors it is important to have accurate and complete detail on the factors under investigation. The main variation in estimated and predicted median survival time by stage, sex, age, and county of residence at time of diagnosis was of interest. These variations were also estimated with regards to final choice of statistical model. With the KM estimator and the statistical models previously explained calculating expected survival though variation of the diagnostic factors, it is possible to calculate the incremental difference between the different sub-groups. However, since there are regional variations with consideration to survival, important steps were taken in order to estimate health gains by achieving *regional best practice* most correctly. The non-parametric estimation and the parametric predication of median survival time was then calculated in groups, representing the counties above and below the national average median survival time to represent "worst" and "best" regions. The best regions accounted for the top 30 % of counties with the longest median

survival time, while the worse regions accounted for the bottom 30 % of counties with the shortest median survival time. This was only the case for the regional/best practice health gains, while the health gains by overall stage was performed for the three main diagnostic stages.

4.5 Modeling the clinical pathway of lung cancer

The structure of the model was designed to reflect the complexities of the natural progression of lung cancer, prognosis and treatment strategies. The model structure was based on relevant literature, protocols and national guidelines. A decision tree and a series of Markov model was developed to estimate long-term costs of lung cancer at a population level. The decision tree structure was used to simulate short term medical costs for each stage of cancer for the primary treatment strategy, until a disease-free state was achieved. The Markov models were used to simulate long-term and/or post-remission and palliative treatment outcomes.

4.5.1 Decision tree

The decision tree was constructed for the purpose of estimating short-term medical costs, allowing for consideration of prognostic factors (e.g. stage) and establish movement through the different primary treatment strategies/pathways and entry into the different Markov models. The construction of a decision tree allowed for costs which are mutual for all stages to be incorporated in a standard method. As initial diagnostic procedures are similar across all stages (to establish staging etc.), the main difference in cost will be on which treatment is performed.

Treatment alternatives are represented as a decision node (square), while chance nodes (circles) represents a point in the decision tree where chance determines which event will occur. In a decision tree, for all branches emanating from a chance node must equal to 1.0, as the sum of the events must occur [62].

The decision tree structure (Figure A 1 in Appendix 9.4) begins with four branches divided according to stage (I-IV), each with further splits according to the different primary treatment strategies. The primary treatment strategies represent the dominant strategies according to the national guidelines. The main difference between the treatment strategies consists of two considerations, whether the tumor is found to be resectable (surgery vs. radiation therapy and/or chemotherapy) and whether or not the intent is curative (curative treatment vs. palliative treatment). This design allowed the model to capture the different costs before a disease-free

state was achieved. After the primary treatment, the response to treatment was included, mainly whether or not the patient had sufficient response to the primary treatment. Patients achieving sufficient response were considered disease-free and entered the Markov model according to the corresponding stage. If the patient indicated insufficient response to the primary treatment strategy, it was assumed to enter the Markov model directly Model A2 (stage IV).

4.5.2 Markov models

A series of Markov models was, as before mentioned developed to estimate long-term medical costs. Five different Markov models was created with the intent to reflect the movement between the different stages of lung cancer, as this has a major impact on both costs and probability of recurrence, post-recurrence remission and mortality. The state-transition Markov model is presented with the intent of visualizing the complete pathway of lung cancer states and outcomes, according to stage, after initial primary treatment. Each cycle length is one month, and the patient could only be in one health state per cycle. Model A1 represent patients moving from stage I-III branches. A2 represent patients moving from stage IV branch. In Model A1 and A2 full circles represents health states and dotted circles are events. Absorbing states are represented by a dot-dash circle. Model B represents the palliative treatment/end-of-life movement and full circles represents a health state and dotted square is an event. Each line represents transitions with different probabilities.

In Figure 8, the transitions between health states and/or events, related to each stage (A1 and A2) and palliative treatment of lung cancer is depicted. Movement into the Markov model is made clear by the endpoints after primary treatment from the decision tree (Figure A 1 in Appendix 9.4). Based on the natural progression of lung cancer, the model consists of four possible health state scenarios for patients in stage I to III: patients remaining in disease-free after primary treatment (Disease-Free); patient entering remission after either local or distant recurrence (Remission); Patient receiving palliative treatment after either local/distant recurrence or from remission after local/distant recurrence (Palliative treatment); patients dying (Death). The model also consists of two mutually exclusive events: local recurrence after initial disease-free state (Local Recurrence) and distant recurrence after initial disease-free state (Distant Recurrence). For patients diagnosed in stage IV, a disease-free state is not achievable, only entry into remission.

At the initial time period, patients entered the Markov model from the decision tree into the "Disease Free" health state (except stage IV patients), depending on whether they achieved a sufficient response to the primary treatment. However, as before mentioned, patients achieving insufficient response to primary treatment was assumed to enter the Markov model directly into Model A2. Patients which entered into the "Disease-Free" health state could either remain, indicated by circled arrow, or experience one of the two mutually exclusive events "Local Recurrence" or "Distant Recurrence". Patients could not remain in these health states, and immediately transitioned to either "Remission", "Palliative Treatment (Model B)" or "Death". The purpose of differentiating between the locations of recurrence is its important impact on achieving remission after recurrence and possible movement to palliative treatment. There are also significant monetary differences between the treatment of local and distant recurrence.

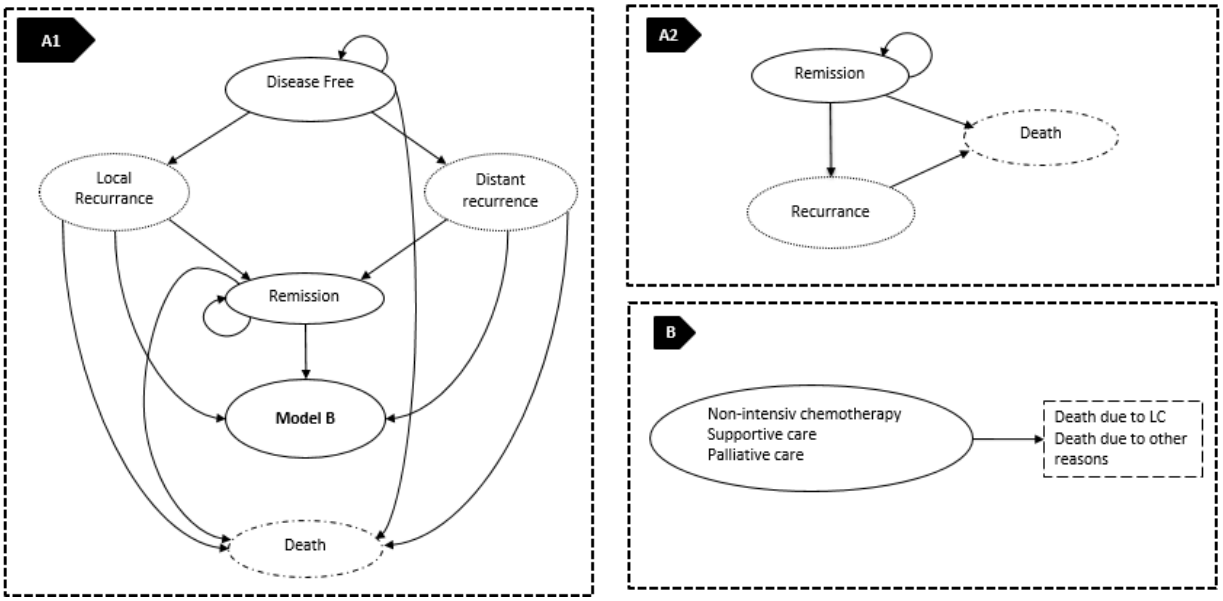


Figure 8: State transition Markov model of lung cancer progression after primary treatment, separated by stage. Model A1 represent patients moving from stage I-III branches. A2 represent patients moving from stage IV branch. In Model A1 and A2 full circles represents health states and dotted circles are events. Absorbing states are represented by a dot-dash circle. Model B represents the palliative treatment/end-of-life movement and full circles represents a health state and dotted square is an event.

It is important to note that in the visualization of the Markov model in Figure 8, although stage I-III have been combined to a joint model, transition probabilities inside the model are divided into three different semi-models. All patients, regardless of which health state they reside in, have the probability of death at given any cycle. There has been no attempt to separate the probability of death given cancer and death by other causes. This is due to the fast progression of lung cancer, a clinical indication of lung cancer as the main driver for death, and the use of overall survival (e.g. death by any cause) in the survival analysis.

4.6 Estimating cost-savings

It is important to note that the method applied trough the modelling exercises describes costs starting at the point of diagnosis and estimate immediate and long-term costs (incident costs). This variant of the incidence approach is to divide cancer care into initial, continuing, and terminal care phases and apply these phase specific cost estimates to survival probabilities [66]. Stage-specific cost variations has been well established through previous research. However, there has been little focus on potential cost savings related to earlier diagnosis. Previous studies on cost variation has so far indicated a skewed distribution, with cancer patient cost of care in the first and last year in active treatment being sizably higher than intermediate measuring times In many cases it is also less costly, from both a patient pathway and a total healthcare perspective, when the cancer is diagnosed early [67].

4.6.1 Costs by stage of diagnosis

The developed lung cancer pathway model estimates costs over a time horizon of five years for stages I trough IV from a health care payer, not including patient's out-of-pocket costs and indirect costs. Furthermore, the previous literature on stage-specific costs for colorectal which will be utilized in this thesis also separated costs according to the TNM classification system. This yields a complication with regards to the CRN staging system (Table 5) used in the dataset for lung and colorectal cancer, as the crude staging system only differentiate between three diagnostic stages. The method for resolving this complication in this thesis is to combine the costs for stage I and II and estimating a 60 % of total costs of these two stages to represent *local disease*. Costs associated with stage III will represent *regional disease*, and 90 % of the total costs associated with stage IV will represent *metastatic disease*. This is will most likely provide a *conservative* estimate. There is however a significant number of studies that does combine stage I and II cancer costs, and differentiating rather on stage IA compared to IIB [33, 67, 68].

4.6.2 Adjusting for regional variations

Incidence rates by stage for cancer of the lung and colon/rectum types was identified through coding of CRN data by the variables in Table 4 and the coding procedure from Table 5. Aggregated numbers, as well as diagnostic stage variation between counties was utilized, which will indicate regional differences in stage distribution. However, since there are regional variations with consideration to population and stage distributions, numerous steps were taken

in order to estimate cost savings by earlier diagnosis most correctly. To determine which county had the most *favorable* stage distribution, the mean stage for all counties was calculated in order to fully capture the significance of all stages, and not just the county which has the highest proportion of patients in localized disease. The mean was then calculated from the counties above and below the national average to represent "worst" and "best" regions. To adjust for the unequal number of patients in each county, together with the calculation of county stage distribution above and below nation average, a weighting method was used. The total number of patients in each was divided by the total number of patients in each county for all years in the diagnostic period of 2000-2016, representing the *population weight*.

After the above described calculations, the best regions accounted for the top 30 % of counties with the lowest mean stage, while the worse regions accounted for the bottom 30 % of counties with the highest mean stage. The mean stage each county was then weighted according to their corresponding population weight for the cost estimation.

4.6.3 Cost-savings estimates

Potential cost-saving estimates will be calculated using different scenarios with consideration to distribution of stage, both absolute and relative cost-saving per patient and national expenditures. For all scenarios, calculations of per patient cost-savings using the methodology described in the above presented chapters (4.6.1-4.6.2) and will first be presented, followed by the total cost-savings per year.

National average

To determine the comparator of nation average, the per patient cost of patients diagnosed with either lung or colorectal cancer was calculated:

$$C_{average}^{CL} = \left(Y_{Local} \times R_{national\ average\ Local\ disease}\right) + \left(Y_{Regional} \times R_{National\ average\ Regional\ disease}\right) + \left(Y_{Distant} \times R_{National\ average\ Distant\ disease}\right) \quad (12)$$

Equation 12 yields the average cost per patient diagnosed with either lung or colorectal cancer using the national average stage distribution where $R_{\text{National average Local disease}}$ represents the average national incidence rate for local disease, $R_{\text{National average Regional disease}}$ represents the

average national incidence rate for regional disease and $R_{\text{National average Distant disease}}$ represents the average national incidence rate for distant disease. C^{CL}_{Average} is the average cost per patient diagnosed with either lung or colorectal cancer. Y_{Local} represent the total cost for localized disease, Y_{Regional} represent the total cost for regional disease and Y_{Distant} represent the total cost for distant metastasis.

(A) All patients are diagnosed in the next lowest severity/stage by national stage distribution

To estimate potential cost saving from early detection (Stage I/localized disease vs. Stage IV/metastatic disease), the total costs for the different stages and incidence rates by stage will be used. Here the scenario is that all patients diagnosed (incidence) with regional disease is allocated to localized disease, and distant metastasis are allocated to regional disease. This scenario is quite extreme, and are only produced for an experimental purpose. Calculation of the estimated per-patient *cost-saving* from early detection by the national average for cancer in the lung and colon/rectum are given by:

$$S^{CL}_{total} = \left(Y_{Regional} - Y_{Local}\right) \times R_{Regional} + \left(Y_{Distant} - Y_{Local}\right) \times R_{Distant} \qquad (13)$$

R_{Regional} represents the average national incidence rate for regional disease and R_{Distant} represents the average national incidence rate for distant metastasis. S^{CL}_{total} is the total monetary cost saving for per patient in this scenario. Y_{Local} represent the total cost for localized disease per patient, Y_{Regional} represent the total cost for regional disease per patient and Y_{Distant} represent the total cost for distant metastasis per patient.

(B) 50 % reduction in each diagnostic stage of national average

It is also possible to estimate potential cost saving from early detection by halving the number of patients in the later stages and allocating these to the "previous" and less severe stage. The total costs for the different stages and incidence rates by stage will be used. Calculation the total estimated per-patient cost-saving by applying a scenario of halving the numbers in the more severe states for cancer in the lung and colon/rectum are given by:

$$S^{CL}_{total} = \left(Y_{Regional} - Y_{Local}\right) \times R_{Regional \times 0.5} + \left(Y_{Distant} - Y_{Local}\right) \times R_{Distant \times 0.5} \qquad (14)$$

$R_{Regional}$ represents the average national incidence rate for regional disease and $R_{Distant}$ represents the average national incidence rate for distant metastasis. S^{CL}_{total} is the total monetary cost saving for each patient. Y_{Local} represent the total cost for localized disease per patient, $Y_{Regional}$ represent the total cost for regional disease per patient and $Y_{Distant}$ represent the total cost for distant metastasis per patient. Here the scenario is that 50 % of patients diagnosed with regional disease is allocated to localized disease, and 50 % of patients with distant metastasis are allocated to regional disease.

(C) Worst counties vs. the national average

It is also possible to estimate potential cost saving by assuming that the county with the highest incidence rates achieves the national average stage distribution. The average total costs for the different stages and incidence rates by stage will be used. Calculation the average total estimated per-patient cost-saving by applying a scenario of achieves the national average stage distribution compared to the worst county for cancer in the lung and colon/rectum are given by:

$$\begin{aligned} S^{CL}_{total} = {} & \{(Y_{Local} \times R_{WL}) - (Y_{Local} \times R_{NL})\} \\ & + \{(Y_{Regional} \times R_{WR}) - (Y_{Regional} \times R_{NR})\} \\ & + \{(Y_{Distant} \times R_{WD}) - (Y_{Distant} \times R_{ND})\} \end{aligned} \quad (15)$$

R_{WL} represents the county with the lowest incidence rate for localized disease and R_{NL} represents the national average incidence rate for localized disease. R_{WR} represents the county with the highest incidence rate for regional disease and R_{NR} represents the national average incidence rate for regional disease. R_{WD} represents the county with the highest incidence rate for metastatic disease and R_{NL} represents the national average incidence rate for distant disease. S^{CL}_{total} is the total monetary cost saving for each patient. Y_{Local} represent the total cost for localized disease per patient, $Y_{Regional}$ represent the total cost for regional disease per patient and $Y_{Distant}$ represent the total cost for distant metastasis per patient. Here the scenario is that the worst county achieves the national average stage distribution.

(D) National average vs. best practice

Some counties might have a more positive stage distribution than others, giving reason for the next scenario, which is cost-saving by achieving the best county stage distribution compared to the national average stage distribution. The average total costs for the different stages and

incidence rates by stage will be used. Calculation the average total estimated per-patient cost-saving by applying a scenario of achieves the best county stage distribution compared to the national average for cancer in the lung and colon/rectum are given by:

$$S_{total} = \{(Y_{Local} \times R_{WL}) - (Y_{Local} \times R_{NL})\} + \{(Y_{Regional} \times R_{WR}) - (Y_{Regional} \times R_{NR})\} + \{(Y_{Distant} \times R_{WD}) - (Y_{Distant} \times R_{ND})\} \quad (16)$$

$R_{Regional}$ represents the average national incidence rate for regional disease and $R_{Distant}$ represents the average national incidence rate for distant metastasis. S_{total} is the total monetary cost saving for each patient. Y_{Local} represent the total cost for localized disease per patient, $Y_{Regional}$ represent the total cost for regional disease per patient and $Y_{Distant}$ represent the total cost for distant metastasis per patient. Here the scenario is that the national average achieves the best county stage distribution.

For the estimation of yearly cost savings, the number of patients and stage distribution was stratified by diagnostic period. The per patient cost savings as mentioned in Chapter 4.6.3 was then performed on the stratified time-periods in order to quantify the increase in incidence over time, and hence the increase in possible cost savings. Since the regional variations were already adjusted for, a simple multiplication was possible (e.g. since the average per patient cost saving were regardless of stage, and rather a mean per patient cost saving using each scenario). The total yearly cost savings was calculated only using the two final scenarios, as these represents the most achievable schemes.

4.7 Key assumptions

Both with regards to the survival analysis, costing pathway for lung cancer treatment and the estimation of health and financial gains, a series of assumptions was made in order to perform the analysis:

- Both the transition and cost parameters in the pathway for lung cancer treatment are based on the treatment of NSCLC. This was done since the CRN dataset do not differentiate lung cancer by NSCLC or SCLC, and previous literature identifies 95 % of lung cancer cases as NSCLC.

- The target population for the lung cancer pathway modelling are all patients with lung cancer as the only diagnosis. Other treatments not directly towards the lung cancer diagnosis are not included (comorbidities from the cancer).

- The duration of the decision three is 1 month, and a five year time horizon with monthly cycle-lengths were applied for the state-transition mode lung cancer pathway model.

- Since the diagnosis date was assumed to be the 1st day of the month, and date of death on the 15th of each month, there was a potential of adding two weeks of survival to every patient.

- The analysis uses overall survival as the main end-point, therefore there were no separation of between death due to the disease or other causes

- The top and bottom 30 % includes 6 counties in each group, out of a total of 20 also including those treated diagnosed in Norway but without registered address.

- The cost weighting performed due to the condensed CRN staging system was performed before calculation of the various scenarios.

- The per patient cost saving were calculated across all stages. Hence, it is important not to be misunderstand the (applied) per patient cost saving with *per stage* cost savings.

4.8 Software

Survival analysis, all tests and regressions were performed in STATA 15.1.1. Modelling of the clinical pathway of lung cancer were calculated in Microsoft Excel 2016. Figures were created in both STATA 15.1.1 and Microsoft Excel 2016.

5 Results

In the results section below, findings using the four described survival analysis methodologies described in Chapter 4.1-4.3 is presented. The results are based on the patient population for both cancers described in Chapter 3.1.1. The survival analysis results begin with the univariate and non-parametric survival estimations, followed by both semi- and parametric models and corresponding regression. The survival analysis results conclude with the adjudication between the different survival models. The results of the calculation of health gains by earlier diagnosis and achieving best regional will then be presented. Finally, the result from the lung cancer costing model, and application of the stage distribution and cost saving scenarios outlined in Chapter 4.6.3, the per patient and yearly financial gains by earlier diagnosis is presented.

5.1 Survival

5.1.1 Univariate data analysis – Non-parametric results

Overall survival

The KM survival curve for patients diagnosed with lung cancer from 2000 to 2016 is depicted in Figure 9. The steep decline in the early years indicates poor prognosis from the disease. Specifically, of the 45 275 patients diagnosed with lung cancer, less than 40 % are still alive 12 months after receiving the diagnosis. The KM survival curve for patients diagnosed with colorectal cancer from 2000 to 2016 is depicted in Figure 10. Of the 66 039 patients diagnosed with colorectal cancer, 76.5 % are still alive 12 months after receiving the cancer diagnosis. The 95% confidence limits of the survivor function is also displayed, together with and number at risk. The limits are extremely thin due to the high number of patients in the cohort, but it does show indications of widening towards the end. This is a usual occurrence due to the number of censored and patients still being alive at the end of the time-period. Table 8 summarizes the one-, 3- and 5-year absolute survival for all patients diagnosed with lung cancer. As observed both in the KM curve in Figure 9 and Table 8, the survival probability after diagnosis of lung cancer stabilizes after the first three years since diagnosis. Median survival time for all patients diagnosed with lung cancer in the study period was 7.24 months. Table 9 summarizes the one-, 3- and 5-year absolute survival for all patients diagnosed with colorectal cancer. Median survival time for all patients diagnosed with colorectal cancer during the study period was 57.39 months (4.8 years).

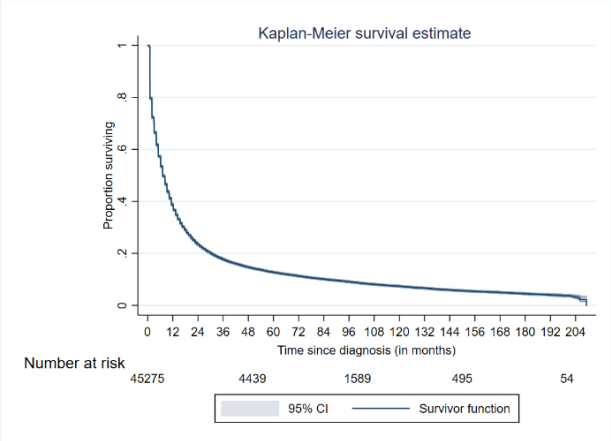


Figure 9: Kaplan-Meier survival curve (survivor function) for patients diagnosed with lung cancer, in months since diagnosis. Including total cohort and patients at risk at 50 month intervals (50-, 100-, 150- and 200-months). CI=confidence interval.

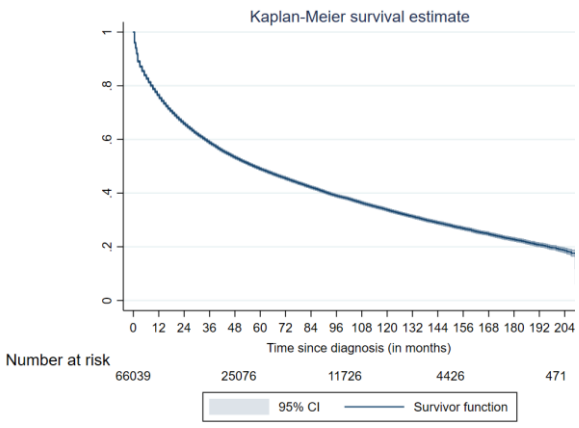


Figure 10: Kaplan-Meier survival curve (survivor function) for patients diagnosed with colorectal cancer, in months since diagnosis. Including total cohort and patients at risk at 50 month intervals (50-, 100-, 150- and 200-months). CI=confidence interval.

Table 8: One-, 3- and 5-year absolute survival for patients diagnosed with lung cancer in the diagnostic period 2000-2016. CI= confidence interval

1-year absolute survival (95% CI)	3-year absolute survival (95% CI)	5-year absolute survival (95% CI)
38.68 (38.22-39.13)	17.60 (17.23-17.97)	12.76 (12.42-13.10)

Table 9: One-, 3- and 5-year absolute survival for patients diagnosed with colorectal cancer in the diagnostic period 2000-2016. CI= confidence interval

1-year absolute survival (95% CI)	3-year absolute survival (95% CI)	5-year absolute survival (95% CI)
76.52 (76.19-76.85)	58.70 (58.31-59.10)	49.04 (48.63-49.46)

Overall survival by age

The KM survival curve comparing age-groups for patients diagnosed with lung cancer from 2000 to 2016 is depicted in Figure A 2 in Appendix 9.5. There are visual differences in median survival between the age groups (10-year age intervals). There are clear indications that higher age is associated with shorter median survival, both by the KM curve and Table 10. The corresponding KM survival curve for age-groups for patients diagnosed with colorectal cancer from 2000 to 2016 is depicted in Figure A 3 in Appendix 9.5. Table 10 provides information about the survival difference between the age-groups for patients with lung cancer. While the median survival time for patients in the age-group 0-49 years is 14.34 months, the median survival for patients who are 90 years or older is only 1.04 months.

Table 10: Difference in survival for age-groups diagnosed with lung cancer from 2000-2017. 10-year age interval. CI= confidence interval. Kaplan-Meier estimation and log-rank test for equality of the survivor function.

	Age-group					
	0-49	**50-59**	**60-69**	**70-79**	**80-89**	**90+**
Number of deaths (O_i)	1 024	4 635	10 870	13 448	7 883	844
Median survival time (months) (95% CI)	14.34 (5.2-100.5)	11.25 (4.1-33.8)	10.24 (3.1-29.7)	7.21 (2.05-21.49)	3.09 (1.04-11.25)	1.04 (1.01-4.14)
Expected number of deaths (E_i)	1 739.7	5 980.2	12 780.5	12 824.2	5 044.7	334.5

$\chi^2 = 3551$, 5 d.f, $p<0.001$

Table 11 provides the KM estimator and log-rank test about the survival difference between the age-groups for patients diagnosed with colorectal cancer. While the median survival time for patients in the age-group 50-59 years is 178.4 months, the median survival for patients who are 90 years or older is 8.2 months. Patients who are diagnosed with lung or colorectal cancer at an older age can expect shorter survival time than patients diagnosed at an early age. The 95% confidence interval also supports this conclusion. A test of difference between the median survival times in the groups is indicative of a difference in survival for both cancer ($p<0.001$)

Table 11: Difference in survival for age-groups diagnosed with colorectal cancer from 2000-2017. 10-year age interval. CI= confidence interval. Kaplan-Meier estimation and log-rank test for equality of the survivor function.

	Age-group					
	0-49	**50-59**	**60-69**	**70-79**	**80-89**	**90+**
Number of deaths (O_i)	1 123	2 692	6 366	11 596	13 032	2 637
Median survival time (months) (95% CI)		178.4 (30.8-214.5)	118.9 (25.6-209.14)	63.6 (15.4-150.7)	28.7 (5.2-82.9)	8.2 (1.5-28.8)
Expected number of deaths (E_i)	2 331.1	4 996.4	9 747.6	11 738.9	7 791.5	840.6

. $\chi^2 = 10560$, 5 d.f, $p<0.001$

Survival by gender

The KM survival curve comparing gender for patients diagnosed with lung cancer from 2000 to 2016 is depicted in Figure A 4 in Appendix 9.5. There are significant differences in median survival between the genders, as visualized by both Figure A 4 and Table 12, and the test of difference between the median survival times in the groups is conclusive of a difference in survival ($p<0.001$). The number of deaths observed among females and males were 15 973 and 22 731, respectfully. When only two groups are compared, the log-rank test is testing the null hypothesis that the ration of the hazard rates in the two groups are equal to 1. The HR of 0.824 indicates that there is 18.2% less risk of death at any point in time among females compared to males. Overall, there is a strong indication that females have longer median survival time compared to males after lung cancer diagnosis.

Table 12: Difference in survival for males and females diagnosed with lung cancer from 2000-2017. CI= confidence interval. Kaplan-Meier estimation and log-rank test for equality of the survivor function.

	Females	**Males**
Number of deaths (O_i)	15 973	22 731
Median survival time (months) (95% CI)	9.19 (2.09-28.63)	7.14 (2.05-18.49)
Expected number of deaths (E_i)	17 815.6	20 888.4
Hazard ratio	0.824	
Log-rank test	$\chi^2 = 370$, 1 d.f, $p<0.001$	

The KM survival curve comparing gender for patients diagnosed with colorectal cancer from 2000 to 2016 is depicted in Figure A 5 in Appendix 9.5. The test of difference between the median survival times in the groups is indicative of a difference in survival ($p<0.001$). The HR of 0.934 indicates that there is 6.6% less risk of death at any point in time among females compared to males, as seen in Table 13. Overall, there is a strong indication that females have longer median survival time compared to males.

Table 13: Difference in survival for males and females for diagnosed with colorectal cancer from 2000-2016. CI= confidence interval. Kaplan-Meier estimation and log-rank test for equality of the survivor function.

	Females	**Males**
Number of deaths (O_i)	18 263	19 183
Median survival time (months) (95% CI)	62.52 (13.31-174.3)	53.32 (13.34-158.92)
Expected number of deaths (E_i)	18 905.2	18 540.8
Hazard ratio	0.934	
Log-rank test	$\chi^2 = 44.4$, 1 d.f, $p<0.001$	

Survival by diagnostic stage

The KM survival curve comparing diagnostic stages for patients diagnosed with lung cancer from 2000 to 2016 is depicted in Figure 11. The corresponding KM survival curve comparing diagnostic stages for patients diagnosed with colorectal cancer is depicted in Figure 12. The number of deaths observed among patients diagnosed with lung cancer between the different diagnostic stages, localized, regional, metastatic and unknown stage were 4 325, 10 105, 19 620, and 4650, respectfully. Using the log-rank method, the expected number of deaths in each diagnostic stage is summarized in Table 14. The number of deaths observed among the different diagnostic stages localized, regional, metastatic and unknown stage for colorectal cancer patients were 4 923, 15 487, 12 805, and 4231, respectfully (Table 15). Overall, there is strong indications that a lower severity of diagnostic stage is associated with longer median survival time for both cancers ($p<0.001$).

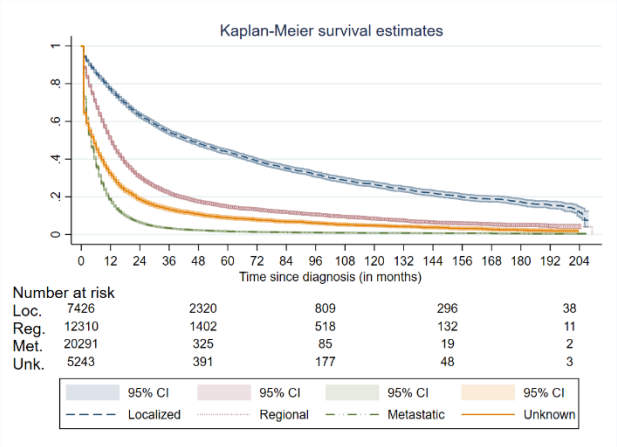


Figure 11: Kaplan-Meier survival curve for lung cancer since time of diagnosis by diagnostic stage. CI= confidence interval.
χ^2 = 11 747.9, 3 d.f, p<0.001

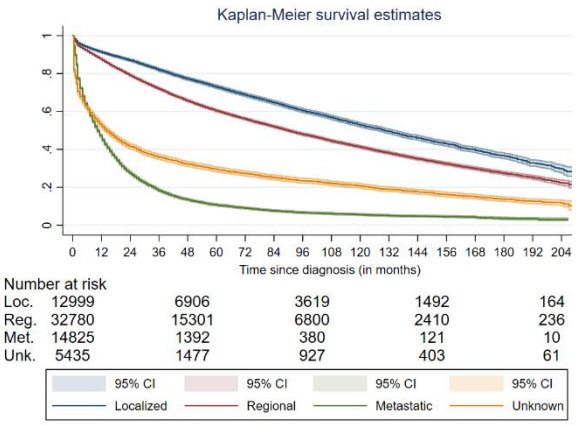


Figure 12: Kaplan-Meier survival curve for colorectal cancer since time of diagnosis by diagnostic stage. CI= confidence interval.
χ^2 = 21 230.3, 3 d.f, p<0.001

Table 14: Difference in survival for each diagnostic stage for patients diagnosed with lung cancer from 2000-2016. CI= confidence interval. Kaplan-Meier estimation and log-rank test for equality of the survivor function.

	Diagnostic stage			
	Localized	**Regional**	**Metastatic**	**Unknown**
Number of deaths (O_i)	4 325	10 105	19 620	4 650
Median survival time (months) (95% CI)	45.07 (13.37-126.15)	12.29 (5.05-30.79)	4.11 (1.04-9.23)	5.15 (1.01-16.44)
Expected number of deaths (E_i)	10 841.7	12 506.1	11 584.4	3 767.9

χ^2 = 11 747.9, 3 d.f, p<0.001

Table 15: Difference in survival for each diagnostic stage for patients diagnosed with colorectal cancer from 2000-2016. CI= confidence interval. Kaplan-Meier estimation and log-rank test for equality of the survivor function.

	Diagnostic stage			
	Localized	**Regional**	**Metastatic**	**Unknown**
Number of deaths (O_i)	4 923	15 487	12 805	4 231
Median survival time (months) (95% CI)	130.2 (54.36-210.12)	90.21 (30.75-192.81)	10.31 (3.09-26.68)	14.35 (1.55-85.09)
Expected number of deaths (E_i)	9 334.9	21 167.1	4 514.5	2 429.5

CI= confidence interval. χ^2 = 21 230.3, 3 d.f, p<0.001

Survival by county

The one-, 3 and 5-year absolute survival for both cancers in each county is presented in Table A 12 and Table A 13 in Appendix 9.5. Due to the high number of observation in the KM curves for regional survival, making the graphs difficult to visually interpret, only median survival time for all counties are presented in Table A 14 in Appendix 9.5 for lung cancer patients and Table A 15 in Appendix for colorectal cancer patients. Table A 14 and Table A 15 in Appendix 9.5 includes number of deaths observed among colorectal and lung cancer patients for the

different counties in Norway and the expected number of deaths in each county. There are significant differences in median survival among the Norwegian counties (p<0.001) for both patient groups. For visual purposes the county with the shortest median survival is indicated by the italic font, while the county with the longest median survival time is indicated by the bold font. For patients diagnosed with lung cancer residing in Oppland the median survival time was 6.19 months, while patients from Vest-Agder has a median survival time of 9.23 months. For colorectal cancer patients residing in Oslo the median survival time was 48.17 months, while patients from Akershus has a median survival time of 66.59 months. These counties represent the shortest and longest median survival time for each cancer.

Survival by county & stage

Further investigations into the counties with the shortest and longest median survival time is presented in Figure 13 (lung cancer) and Figure 14 (colorectal cancer). Patients diagnosed with lung cancer and localized disease in Vest-Agder had a median survival time of 59.7 months, patients with the corresponding disease progression in Oppland has a median survival time of only 37.1 months. For patient diagnosed with colorectal cancer and localized disease in Akershus the estimated survival time was 144.9 months, while patients with the same diagnostic stage in Oslo, had a survival time of 128.1 months.

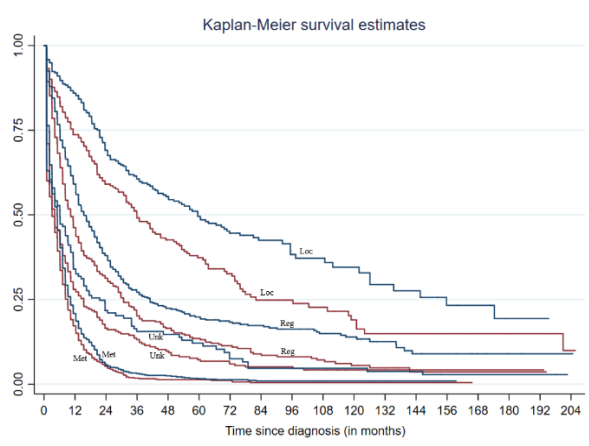


Figure 13: Kaplan-Meier survival curves for lung cancer patients residing in Vest-Agder (Blue) and Oppland (Red) at time of diagnosis, by localized (Loc), regional (Reg), metastatic (Met) and unknown (Unk) stage.

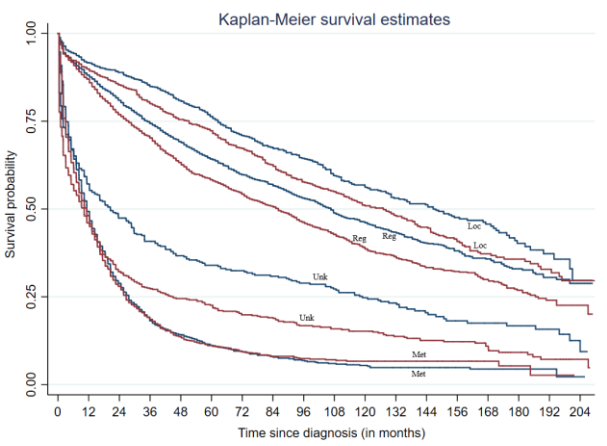


Figure 14: Kaplan-Meier survival curves for colorectal cancer patients residing in Akershus (Blue) and Oslo (Red) at time of diagnosis, by localized (Loc), regional (Reg), metastatic (Met) and unknown (Unk) stage.

5.1.2 Multivariate data analysis – Parametric and semi-parametric

Cox semi-parametric regression estimates – Lung cancer

To determine the independent predictors for patient's survival after the lung cancer diagnosis, the Cox's regression model was studied. The proportional hazard assumption was investigated for each variable using the *log(-log(s(t))* against *log(t)* plot. Correlation between ranking of individual failure time and the Schoenfeld residuals was also considered. The Schoenfeld residual test showed that the PH assumption was held for factors stage (p<0.001), age-group (p<0.001) and county (p =0.0028), and violated for gender (p=0.089*)*. However, since previous tests for gender indicated significance, it was included in the analysis. The univariate and multivariate Cox regression model is presented in Table A 16 in Appendix 9.5. Through the diagnostic period (2000-2016) men exhibited a 19% increased risk of dying (95 % CI 1.16-1.21) compared to women. The Cox regression predicted a median survival time of 10.1 months for women and 7.2 months for men. Patients between 70-79 years exhibited a 95% increased risk of dying (95% CI 1.83-2.08) compared to those younger than 49 years. In addition, those aged 90 years and over had an increased risk of dying of 3.62 times higher (95% CI 4.22-5.07) compared to patients younger than 49 years. The Cox regression predicted a median survival time for patients younger than 49 years of 15.1 months and 8.1 months patients between 70-79 years.

Patients with regional and metastatic disease at time of diagnosis exhibited a 2.26 (95% CI 2.17-2.34) and 4.32 (95% CI 5.14-5.51) times higher, respectfully, increased risk of dying compared to those diagnosed with localized disease. Patients with unknown stage at time of diagnosis exhibited a 2.19 times higher risk of dying (95% CI 3.05-3.32) compared to those diagnosed with localized disease. The Cox regression model predicted a median survival time of 37.3 months for patients diagnosed with localized disease. Patients with regional, metastatic, and unknown stage at time of diagnosis had a median survival time of 11.5, 3.1 and 6.2 months, respectfully. For the county covariate, Østfold represented the reference group. Patients who resided in Vest-Agder exhibited a 22 % (95% CI 0.73-0.83) reduction in the risk of dying compared to Østfold. Vest-Agder was the county with the highest reduction in risk among the Norwegian counties with a predicted median survival time of 9.8 months. The Cox regression predicted the shortest median survival time of 6.2 months exhibited in Oppland.

Comparison of the Cox, Weibull and Exponential regression models – Lung cancer

The aim was to derive a prognostic model, focusing on the four covariates which were also investigated in the Cox regression. The Weibull PH regression is presented in Table A 17 in Appendix 9.5 and the Exponential PH regression is presented in Table A 18 in Appendix 9.5. The regression outputs for all three models shows the effect size (given as hazard ratios), 95 % CI, and statistical significance for each of the covariates in relation to overall survival. Each factor is assessed through separate univariate regressions (left-hand side columns). However, the aim is to describe how the factors jointly impact on survival, and so all four factors were incorporated into the multivariate models (right-hand side columns). There are strong indications that males, increased age, severity and diagnostic stage impaired survival to a varying degree (all statistically significant) in all three regressions. County of residence at time of diagnosis was also statistically significant (to a varying degree), the county factors was however less significant for the Weibull PH model. All the covariates predicted by the three different models yielded the results visualized in Figure A 6 – A 9 in Appendix 9.5. Figure A 6 represents the prediction on survival according to age-groups. Figure A 7 represents the three different models prediction on survival according to gender. Figure A 8 represents the three different models prediction on survival according to county, and Figure A 9 represents prediction on survival according to diagnostic stages.

Cox semi-parametric regression estimates – Colorectal cancer

The Cox regression model was investigated in order to predict colorectal cancer patients' survival, and the independent predictor's significance. The proportional hazard assumption was investigated for each variable using the *log(-log(s(t))* against *log(t)* plot. Correlation between ranking of individual failure time and the Schoenfeld residuals was also considered. The Schoenfeld residual test showed that the PH assumption was held for factors stage ($p<0.001$), age-group ($p<0.001$) and county ($p=0.0036$), and violated for gender ($p=0.056$). The univariate and multivariate Cox regression model is presented in Table A 19 in Appendix 9.5. Predictions from the multivariate model are as follows; through the diagnostic period from 2000 to 2016, males had a 17.2 % increased risk of death (95 % CI 1.15-1.19) compared to females, with a survival time of 50.3 and 63.6 months, respectfully. Patients between 60-69 years of age at time of diagnosis exhibited a 48.6 % increased risk of dying (95 % CI 1.39-1.58) compared to patients that were younger than 49 years when diagnosed with colorectal cancer, with a predicted median survival time of 110.1 and 166.2 months, respectfully.

Patients with regional and metastatic disease at time of diagnosis exhibited a 41.9% increased (95% CI 1.37-1.47) and 6.61 times higher (95% CI 6.97-7.46), respectfully, increased risk of dying compared to those diagnosed with localized disease. Patients with unknown stage at time of diagnosis exhibited a 216% increased risk of dying (95% CI 3.03-3.29) compared to those diagnosed with localized disease. Median survival time predicted by the Cox regression for patients with localized, regional, metastatic and unknown stage were 143.9, 100.2, 10.8 and 33.1 months, respectfully. According to the Cox multivariate regression, patients that were registered as residing in Hedmark had the greatest risk of death compared to residents in Østfold. However, the covariate was not significant. Akershus was the county with the highest reduction in risk, and covariate was significant. The Cox regression predicted the shortest median survival time of 58.2 months exhibited in Oslo, and the longest in Akershus with 68.9 months.

Comparison of the Cox, Weibull and Exponential regression models – Colorectal cancer

As with the lung cancer data set, the aim was to derive a prognostic model with the corresponding four covariates. The Weibull PH regression is presented in Table A 20 in Appendix 9.5 and the Exponential PH regression is presented in Table A 21 in Appendix 9.5. The regression output for the three regression models shows the effect size (given as hazard ratios), 95 % CI, and statistical significance for each of the covariates in relation to overall survival. Each factor is assessed through separate univariate regressions (left-hand side columns). However, the aim is to describe how the factors jointly impact on survival, and so all four factors were incorporated into the multivariate models (right-hand side columns). There are strong indications that males, increased age, severity and diagnostic stage impaired survival to a varying degree (all statistically significant) in both models. County of residence at time of diagnosis was also statistically significant (to a varying degree). All the covariates predicted by the three different models yielded the results visualized in Figure A 10 – A 13 in Appendix 9.5. Figure A 10 represents the prediction on survival according to age-groups. Figure A 11 represents the three different models prediction on survival according to gender. Figure A 12 represents the three different models prediction on survival according to county, and Figure A 13 represents prediction on survival according to diagnostic stages.

Adjudicating between different parametric models

When the variables are selected to be in the model, it is important to evaluate how well the model represents the data. A survival model is adequate if it represents the survival patterns in the data to an acceptable degree. The survival pattern for all models were assessed by the goodness of fit method and the Cox-Snell residual plot. No apparent trend was identified to violate the validity of the models and no covariate has been incorrectly omitted. Identification of the correct parametric model can be investigated by the log-likelihood, AIC, and BIC statistics previously explained. The assessment of the adequacy of the two parametric models, Weibull and Exponential for the lung cancer dataset are presented in Table 16:

Table 16: Log likelihood, Akaike's Information criteria (AIC) and Bayes Information criteria (BIC) of three different distributions fitted to the full model – Lung cancer dataset

Model	Log likelihood (LL)	No. of covariates (c)	No. of ancillary parameters (a)	AIC	BIC
Exponential	*-75 107.87*	30	1	150 273.7	150 526.6
Weibull	-73 488.89	30	2	*147 038.5*	*147 299.9*

AIC = -2*ln(LL) + 2*number of estimated parameters. BIC = -2*ln(LL) + ln(number of observations)*number of estimated parameters.

The highest maximum log-likelihood is achieved by the Exponential PH model, which is an preferable result. The Weibull PH achieves the smallest AIC and BIC values, which indicated a better fit than the Exponential PH model. A conclusive result is therefore not achieved, but the model which met the most criteria's is the Weibull PH model, indicating the best parametric model for the data. The assessment of the adequacy of the two parametric models, Weibull and Exponential for the colorectal cancer dataset are presented in Table 17. The adjudication results yield the same results as in the lung cancer dataset, with Weibull as best model fit.

Table 17: Log likelihood, Akaike's Information criteria (AIC) and Bayes Information criteria (BIC) of three different distributions fitted to the full model – Colorectal cancer dataset

Model	Log likelihood (LL)	No. of covariates (c)	No. of ancillary parameters (a)	AIC	BIC
Exponential	*-89 428.69*	30	1	178 915.4	179 179.2
Weibull	-87 648.54	30	2	*175 357.1*	*175 630*

AIC = -2*ln(LL) + 2*number of estimated parameters. BIC = -2*ln(LL) + ln(number of observations)*number of estimated parameters.

Assessing model appropriateness

Identifying the model which best fit the data could be difficult when comparing both non-parametric and parametric models. Since the likelihood computed in a Cox model is a partial likelihood, it cannot be compared with fully parametric models which compute log-likelihood. However, the results from the regression output from the Cox or parametric PH models may be compared directly, as the model types are merely different approaches to assessing the same quantity. Both the best fitted parametric model (Weibull PH) and the Cox PH model yields the same indications as to the significance of the prognostic factors.

5.2 Estimated health gains

Only the KM estimation and the Weibull PH results will be addressed, as they represent the standard non-parametric estimation and the parametric model which indicated the best fit for the dataset. The incremental health gains (in survival months) by overall stage and regional/best practice using the method described in Chapter 4.4, will be presented. In addition, using the incidence rates from the dataset, possible life-years saved by earlier diagnosis with overall diagnostic stage allocations will be mentioned.

5.2.1 Kaplan-Meier and Weibull estimation

By stage

The KM estimation from Table 14 yielded that patients diagnosed with regional disease at time of diagnosis can expect to live on average 8.18 months longer than patients with metastatic disease, while patients diagnosed with localized disease can expect to live on average 40.96 months longer than patients diagnosed with metastatic disease.

Predictions from the fully fit Weibull model and is presented in Table 18 with the median survival time for each diagnostic stage of lung cancer, together with the incremental gain in months between the stages. The Weibull PH model predicted a median survival time for patients with localized disease of 49.79 months, 44.78 months longer than patients with metastatic disease at time of diagnosis. The Weibull model predicted slightly longer median survival times for all stages compared to the KM estimator.

Table 18: Difference in survival between patients diagnosed localized, regional, metastatic or unknown diagnostic stage of lung cancer at time of diagnosis, from 2000-2016, using the fully fitted Weibull PH model. CI=confidence interval

Diagnostic stage	Median survival time[a] Weibull estimation (95% CI)	Incremental Δ Weibull estimation
Unknown	9.03 (8.91-9.15)	
Metastatic	5.01 (4.98-5.04)	-4.02
Regional	16.26 (16.14-16.39)	11.25
Localized	49.79 (49.29-50.28)	33.53

[a]in months

The corresponding results for the predictions from the fully fit Weibull with the median survival time for each diagnostic stage of colorectal cancer as, together with the incremental gain in months between the stages is presented in Table 19:

Table 19: Difference in survival between patients diagnosed localized, regional, metastatic or unknown diagnostic stage of colorectal cancer at time of diagnosis, from 2000-2016, using the fully fitted Weibull PH model. CI=confidence interval

Diagnostic stage	Median survival time[a] Weibull estimation (95% CI)	Incremental Δ Weibull estimation
Unknown	32.28 (31.58-32.98)	
Metastatic	15.62 (15.47-15.78)	-16.66
Regional	117.85 (117.01-118.68)	102.23
Localized	184.42 (182.36-186.47)	66.57

[a]in months

Using the results from the KM estimation (Table 15), patients diagnosed with regional disease at time of diagnosis can expect to live on average 79.9 months longer than patients with metastatic disease, while patients diagnosed with localized disease can expect to live on average 119.89 months longer than patients diagnosed with metastatic disease. The Weibull PH model predicted a median survival time for patients with localized disease of 184.42 months, 168.8 months longer than patients with metastatic disease at time of diagnosis. The predictions from the Weibull model is substantially larger compared to the KM estimator, especially for predications of median survival time for localized and regional disease.

Furthermore, if all patients diagnosed with lung cancer and metastatic disease in the diagnostic period of 2012-2015 (5 114 patients) were instead diagnosed with regional disease, it would result in gain of 4 794 life-years (over a 4 year period) using the Weibull prediction, and 3 486 life-years using the KM estimator. The reciprocal result for all patients diagnosed with lung cancer with metastatic disease in the diagnostic period of 2012-2015 (3 795 patients) would result in gain of 32 330 life-years, using the Weibull prediction and 25 268 life-years using the KM estimator (over a 4 year period).

By county

Using the top and bottom 30 % to represent the respective outer edges of best regional practice for the KM estimation yielded that patients with lung cancer residing in the top 30 % had a median survival time of 8.88 months (e.g. mean of the median survival time), while the bottom had a survival time of 6.83 months. This entails that patients residing counties among the top 30 % can expect to live on average an estimated 2.05 months longer than patients diagnosed from bottom 30 % and 1.64 months longer than the national average, using the KM estimator.

Table 20: Difference in survival between the county with longest and shortest median survival time, together with the national average for patients diagnosed with lung cancer from 2000-2016. CI=confidence interval

County	Median survival time[a] Weibull estimation (95% CI)	Incremental Δ Weibull estimation
Bottom 30 %	13.28 (12.58-13.99)	
Norway	15.89 (15.71-16.05)	2.61
Top 30 %	19.18 (18.18-20.19)	3.29

[a]in months

Predictions om median survival time from the fully fit Weibull model and is presented in Table 20 with the median survival time for the respective outer edges of survival time, together with the incremental gain in months between the two counties and the national average. The Weibull PH model predicted a median survival time for patients residing in the top 30 % of 19.18 months, 5.9 months longer than patients residing in the bottom 30 % at time of diagnosis, and 3.29 months longer than the Norwegian average. The Weibull model predicted a longer median survival times for all stages compared to the KM estimator. However, the incremental gain in months between the two methods are in close proximity.

By county – Colorectal cancer

The corresponding results for incremental gains for the Weibull survival model predictions in survival time for the county with longest and shortest median survival time and the national average for patients diagnosed with colorectal cancer is presented in Table 21:

Table 21: Difference in survival between the county with longest and shortest median survival time, together with the national average for patients diagnosed with colorectal cancer from 2000-2017. CI=confidence interval

County	Median survival time[a] Weibull estimation (95% CI)	Incremental Δ Weibull estimation
Bottom 30 %	93.49 (91.19-95.8)	
Norway	100.96 (100.22-101.69)	7.74
Top 30 %	114.29 (111.72-116.87)	13.3

[a]in months

Using the top and bottom 30 % to represent the respective outer edges of best regional practice for the KM estimation yielded that the top 30 % had a median survival time of 63.24 months, while the bottom 30 % had a median survival time of 49.36 months. This entails that patients residing in the top 30 % at the time of diagnosis can expect to live on average an estimated 13.88 months longer than the bottom 30 %, and 5.85 months longer than the national average. The Weibull PH model predicted a median survival time for patients residing in top 30 % of 114.29 months, 20.8 months longer than patients residing in the bottom 30 % at time of diagnosis and 13.3 months longer than the Norwegian average. The predictions from the Weibull model is substantially larger compared to the KM estimator for all counties, as with the predictions for difference between counties for survival for patients with lung cancer. However, the incremental gain in months between the two methods are in close proximity.

5.3 Estimated cost-savings

5.3.1 Stage-specific costs

The per patient direct hospital costs of localized, regional and metastatic lung cancer are presented in Figure 15, depicted together with the equivalent result for colorectal cancer. The cost of direct treatment cost for lung cancer are calculated over the duration of the model (5 years) presented in Chapter 4.5 and weighed according to the reasoning explained in Chapter 4.6.1. Sub-results from the clinical pathway modelling is presented in Figure A 14 in Appendix 9.5. The main cost component was *treatment* for all stages, followed by *diagnostic procedures*, *pre-treatment*, then *others*. The cost component other consisted of surgical complications as a result of the main treatment and at-home palliative care.

The total lifetime CRC costs from the Joranger et.al [4] model and article estimated costs as following for stage I-IV; 23 386 €, 33 501 €, 49 894 € and 61 396 €, respectfully. These estimates were given in 2011 euros, and by Formula 4 (PPP-calculation) and weighting according to the reasoning explained in Chapter 4.3.1, the costs for the aggregated stages are presented in Figure 19. It is important to note that the time-horizon in the lung cancer pathway model presented in Chapter 4.5 is 5 years while the estimations from Joranger et.al [4] estimated the lifetime cost of colorectal cancer with a time-horizon of 10 years.

The national average stage distribution for lung cancer yields an average cost per patient of NOK 279 069, excluding costs of patient with unknown stage at diagnosis calculated using Formula 12. When assigning the number of patients with unknown diagnostic stage according to the weight of the known stage distribution, the average cost per patient is NOK 311 441. The national average stage distribution for colorectal cancer yields an average cost per patient of NOK 372 257, excluding costs of patient with unknown stage at diagnosis calculated using Formula 12. When assigning the number of patients with unknown diagnostic stage according to the weight of the known stage distribution, the average cost per patient is NOK 413 206. The average cost per diagnostic stage is depicted in Figure 15.

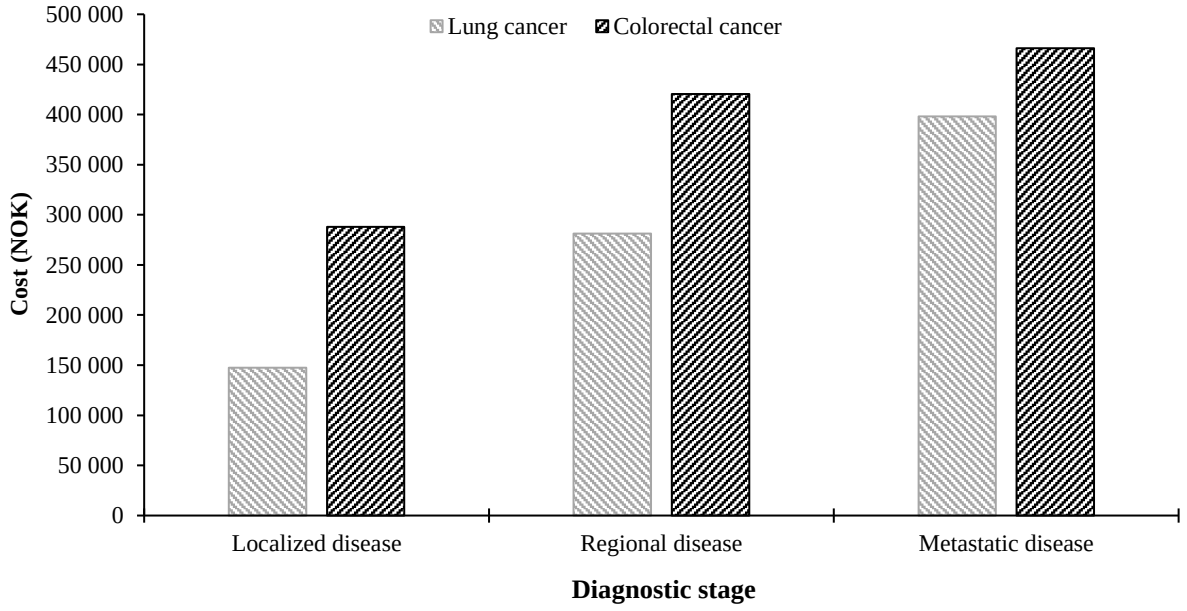


Figure 15: Direct hospital costs (NOK) of lung cancer over a 5-year time horizon (grey), and the total lifetime cost (NOK) of colorectal cancer (black), by localized, regional and metastatic disease.

5.3.2 Cost-saving estimates

Per patient cost savings

The per patient cost saving, calculated by all four scenarios are depicted in Figure 16. The per patient cost saving were calculated across all stages, which entails that they are adjusted for stage, and apply only to the various scenarios. Hence, it is important not to misunderstand the per patient cost saving with per stage cost savings.

While scenarios A and B are quite extreme, they do illustrate the significance of the treatment costs associated with a higher degree of cancer severity and diagnostic stage. Scenario C on the other hand is quite reasonable considering the counties which have a higher proportion of patients in the more severe stages could possibly achieving the national stage distribution. In this scenario there is a possibility of achieving an average per patient cost saving of NOK 8 617 for lung cancer patients, and NOK 21 583 for colorectal cancer patients. Scenario D is considered best practice, and the estimated average per patient cost saving in this scenario yields a possible financial gain of NOK 12 269 for lung cancer patients and NOK 11 202 for colorectal cancer. For possible cost savings by earlier diagnosis of colorectal cancer, the monetary value is higher in scenario C, while for lung cancer the cost savings are higher in scenario D. This is because of the large amount of patients diagnosed with regional disease, and the small cost variation for colorectal cancer between regional and distant disease presented in Figure 15. While for lung cancer the cost variation between the diagnostic stages are higher, and the variation between scenario C and scenario C is that there is a greater distance from the top 30 % to the national average, than the bottom 30 % to the national average.

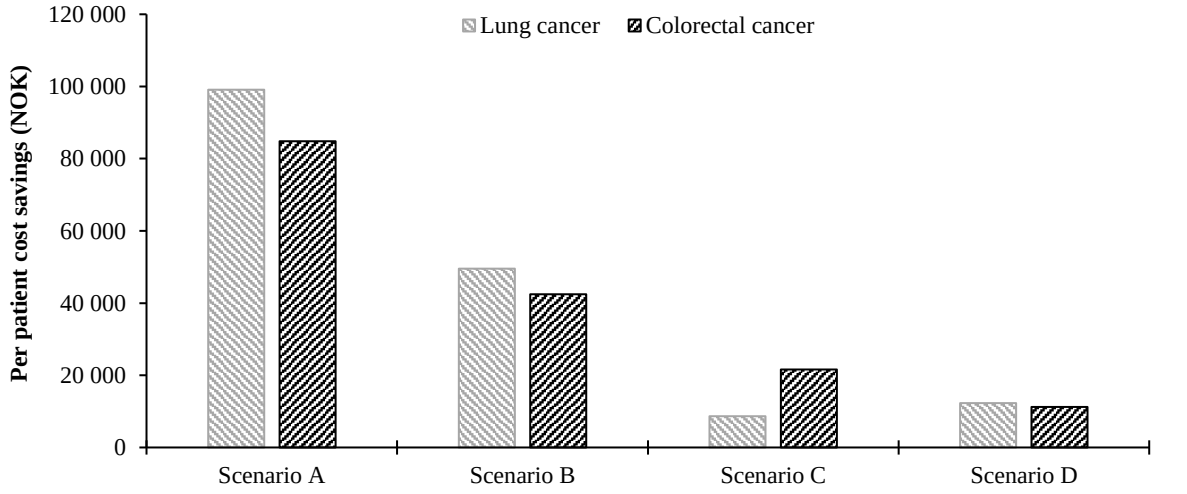


Figure 16: Per patient cost savings (NOK) according to four allocation scenarios for lung and colorectal cancer.

Yearly cost savings

Since scenarios A and B are quite extreme, they have not been included in the yearly cost saving estimation. However, from Figure 16, illustration of the significance of the treatment costs associated with a higher degree of cancer severity and diagnostic stage are comprehended. Scenario C on the other is quite reasonable considering the counties which have a higher

proportion of patients in the more severe stages could possibly achieving the national stage distribution. Scenario D could be interpreted as best practice with the desire of achieving the stage distribution represented by counties with a desirable stage distribution than the national average.

For while the average per patient cost savings appear small for the two most "achievable" scenarios, lung cancer and colorectal cancer are among the most frequently diagnosed cancers. Hence, with 4 634 new cases of colorectal cancer and 3 206 new cases of lung cancer in 2016, there is a potential of extensive yearly cost savings. Due to the increase in incidence of both cancer, it is of interest to examine the yearly cost savings by diagnostic period. Since the previous calculation already accounted for the mean cost saving per patient, simple multiplication, stratified by diagnostic period. With the increase in incidence, the possible cost savings increase correspondingly, as visualized in Figure 17 and Figure 18.

When adjusting for 4-year consecutive intervals between 2000 and 2015, the increase in cost saving estimates increase over the diagnostic periods. Figure 21 depicts the increase in potential yearly cost savings over the four consecutive diagnostic periods for lung cancer. As visualized by Figure 17, scenario D yields the greatest cost saving scenario, with an estimated potential yearly cost saving of over NOK 37 559 099 for the diagnostic period of 2012-2015. From scenario C, the calculation yielded a yearly cost saving of NOK 26 448 754 for the diagnostic period of 2012-2015.

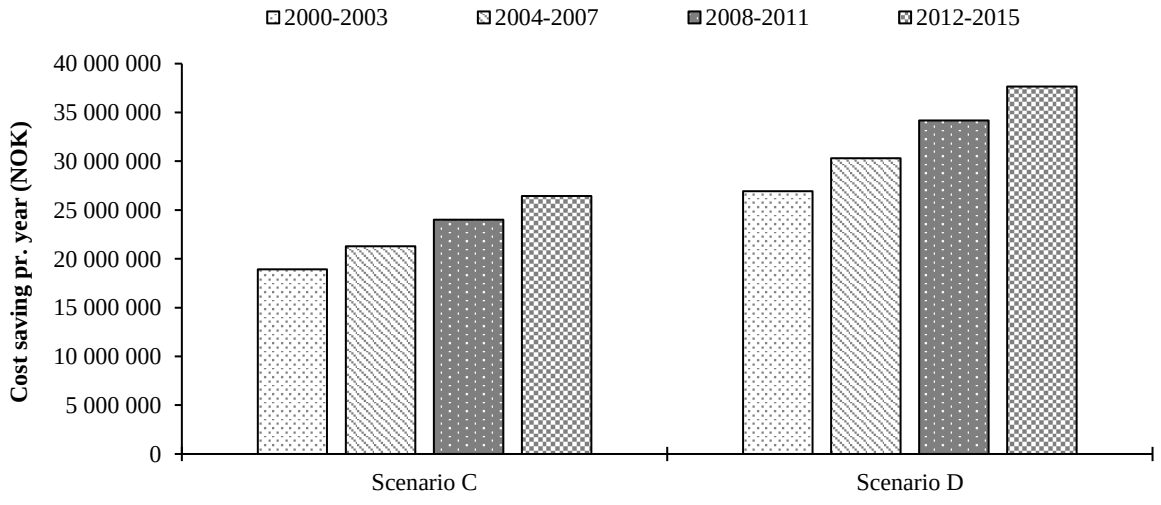


Figure 17: Yearly cost savings (NOK) stratified by diagnostic period (4-year consecutive intervals) according to allocation scenario C and D, using the average per patient cost saving for lung cancer.

For potential cost savings related to earlier diagnosis of colorectal cancer, scenario C; "*The counties which have a higher proportion of patients in the more severe stages achieving the national stage distribution*" yielded a total yearly cost saving of almost NOK 94 239 426 for the diagnostic period of 2012-2015. Also here there are strong indication that the increase in incidence yields larger cost saving potential. In the diagnostic period 2000-2003, the yearly cost saving was estimated to be NOK 73 578 518, which entails that with the increase in incidence there is a 28 % increase in potential cost savings compared to the diagnostic period of 2012-2015. Scenario D which could be interpreted as best practice, and yielded a potential cost saving estimation of NOK 48 910 820 for the diagnostic period of 2012-2015. Figure 21 depicts the increase in potential cost savings over diagnostic period for colorectal cancer. For scenario D, the potential cost saving was estimated to be NOK 38 187 686, NOK 40 797 757, NOK 44 079 949 and NOK 48 910 820 over the consecutive four-year diagnostic intervals.

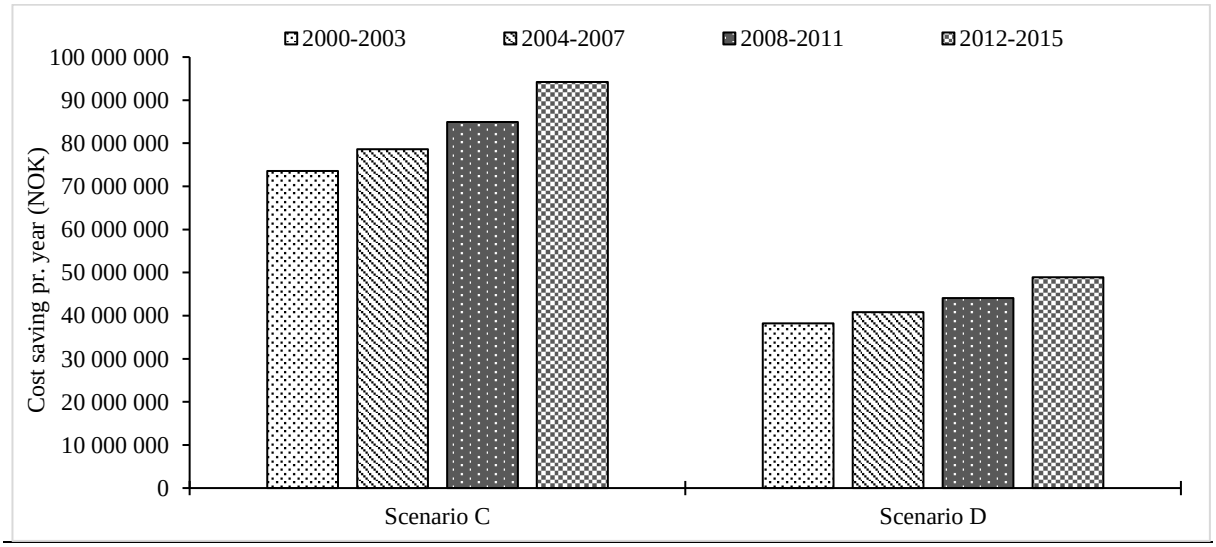


Figure 18: Yearly cost savings (NOK) stratified by diagnostic period (4-year consecutive intervals) according to allocation scenario C and D, using the average per patient cost saving for colorectal cancer.

6 Discussion

This study aimed at investigating and analyzing stage-specific survival, the regional differences in survival, and the stage-specific cost differences of patients diagnosed with lung or colorectal cancer in Norway. The main emphasis was placed on the regional diagnostic stage distribution and the corresponding survival, in order to investigate health gains by earlier diagnosis and regional best practice. As stage specific costs were included, an analysis of financial gain was possible by allocative scenarios. Different statistical survival models were analyzed as to examine which could best fit the data and predict survival over time. In addition, the most used survival distributions were compared, with the aim of informing future studies survival analysis using national registry data.

6.1 Main findings

This thesis found that there are significant differences in survival between both the diagnostic stages and the Norwegian counties for patients diagnosed with lung or colorectal cancer. Patient characteristics such as age and gender had a significant impact on survival time for both cancers, which is frequent for all cancer survival analysis.

The KM estimator and corresponding log-rank test for equality of the survival function for the difference between regional survival time in patients diagnosed with lung cancer yielded a highly significant result. Not unexpectedly, a lower grade of cancer severity (e.g. diagnostic stage) indicated longer median survival time between all diagnostic stages. More specifically, patients diagnosed with lung cancer and with localized disease had a median survival time of 45.97 months, while patients diagnosed with metastatic disease had a median survival time of 4.11 months, a difference of almost 41 months or 3 ½ years. For patients diagnosed with colorectal cancer and localized disease the median survival time was estimated to be over 130 months, while for patients with metastatic disease could expect a median survival time of 10.31 months. This entails a difference of almost 120 months, or 10 years.

The county with the shortest median survival time for lung cancer patients, as concluded by the KM estimator, was identified as Oppland with 6.19 months. The greatest median survival time was found in Vest-Agder with a median survival for all patients of 9.23 months. Oppland and Vest-Agder also had the shortest and greatest 1-year absolute survival with 34.12 % and 43.59

%, respectfully. This entails an almost 10 percentage point in difference, or a 27.7 % higher 1-year absolute survival in Vest-Agder. The reciprocal results for patients diagnosed with colorectal cancer concluded by the KM estimator yielded a significant difference in survival between all counties in Norway. For patients with colorectal cancer the county with the shortest median survival time was identified as Oslo with 48.17 months and the greatest median survival time was found in Akershus with a median survival for all patients with colorectal cancer of 66.59 months. This entails a difference in survival months of 18.42 months. In the case of colorectal cancer, there result was not similar when assessing 1-year absolute survival and the regional stage distribution. Finnmark had the greatest with 79.9 % of patient still being alive one year after diagnosis, while in Hedmark, after one year, 74.4 % of patients were still alive.

Both the construction of KM survival curves for different patient groups and the long-rank test to investigate differences are performed as univariate analysis, and only describe survival with respect to each factor ignoring the impact of any other. The multivariate survival analysis with proportional hazard assumption was therefore performed, both semi-parametric and parametric distribution considerations. The focus was on factors (e.g. covariates) gathered at the time of diagnosis.

The Cox regression model of the independent predictors for patient's survival after the lung cancer diagnosis estimated similar results as the non-parametric estimation for the covariates; gender, age-group and gender, and the mentioned covariates were all statistical significant in both the univariate and multivariate analysis. The county covariate (e.g. unordered categorical) were statistically significant when assessed alone, while when analyzed singly there were mixed results. In the univariate analysis Oppland yielded the highest HR of 1.023 (2.3 %) compared to the reference group (Østfold), while Vest-Agder had HR of 0.819 which entails a 22 % reduction in the risk of dying compared to Østfold through the diagnostic period 2000-1016. This corresponds to the non-parametric KM estimation. The results for the county covariate shifted in the multivariate analysis, where the reference county (e.g. Østfold) yielded the highest baseline hazard, while Oppland had the highest HR and Vest-Agder had the lowest when comparing across the remaining counties.

The reciprocal results with the Cox regression model for patients diagnosed with colorectal cancer also estimated similar results as the non-parametric estimation for the covariates; gender, age-group and gender, and the mentioned covariates were all statistical significant in both the univariate and multivariate analysis. As with the lung cancer dataset, the county covariate were

statistically significant when assessed alone, while when analyzed singly there were mixed results. In the univariate analysis Akershus yielded a HR of 0.879 which entails a 12.1 % reduction in the risk of dying compared to Østfold through the diagnostic period 2000-1016 and the result was significant. However, the Cox regression model yielded many insignificant results among the county covariates in both the univariate and multivariate analysis. This could indicate that this covariate do not have a significant effect on survival.

The Weibull PH and Exponential PH models predicted similar results as the Cox model when fitting the both the colorectal and the lung cancer dataset, and may therefore be interpreted in the same manner. The results derived from the Weibull PH model and Exponential PH model yielded the similar result with regards to the HR for the scale of difference between counties, and they also estimated a higher degree of significance for the county covariates than the Cox regression. This was the case for both of the datasets. The adjudication between the parametric models were therefore assessed on the basis of the Log-likelihood, AIC and BIC results, which identified the Weibull PH models as the preferred model for both cancers.

From the fully fit Weibull PH model, a median survival time for patients with lung cancer and localized disease of 49.79 months was predicted, 44.78 months longer than patients with metastatic disease at time of diagnosis. The Weibull PH model for the colorectal cancer dataset predicted a median survival time for patients with localized disease of 184.42 months, 168.8 months longer than patients with metastatic disease at time of diagnosis.

The Weibull PH predicted a median survival time for patients with lung cancer residing in the top 30 % (of the counties with the longest median survival time) of 19.18 months, 5.9 months longer than patients residing in the bottom 30 % at time of diagnosis, and 3.29 months longer than the Norwegian average. The reciprocal result for the colorectal cancer predicted a median survival time for patients residing in top 30 % of 114.29 months, 20.8 months longer than patients residing in the bottom 30 % at time of diagnosis and 13.3 months longer than the Norwegian average.

The per patient cost estimated by the deterministic costing model for NSCLC calculated over the duration of the model (5 years) the cost of localized, regional and metastatic disease of NOK 147 296, 281 344 and 398 194, respectfully. Direct hospital costs for metastatic disease is estimated to be over 2.7 times higher than localized disease. Among the different sub-groups, the treatment costs yielded the highest cost across all diagnostic stages, followed by costs

related to diagnostics, pre-treatment and other, respectfully. For the calculation of the total lifetime cost of colorectal cancer from Joranger et.al [4], the cost of localized, regional and metastatic disease was NOK 287 993, 420 605 and 466 230, respectfully.

Application of cost saving scenario C resulted in a possible average per patient cost saving of NOK 8 617 for lung cancer patients, and NOK 21 583 for colorectal cancer patients. Scenario D which could be considered best practice, estimated an average per patient cost saving of NOK 12 269 for lung cancer patients and NOK 11 202 for colorectal cancer.

Scenario D yielded the greatest cost saving scenario for lung cancer, with an estimated potential yearly cost saving of over NOK 37 559 099 for the diagnostic period of 2012-2015. From scenario C, the calculation yielded a yearly cost saving of NOK 26 448 754 for the diagnostic period of 2012-2015. For potential cost savings related to earlier diagnosis of colorectal cancer, scenario C yielded an estimated financial gain of NOK 94 239 426 per year for the diagnostic period of 2012-2015. For scenario D, the potential cost saving was estimated to be NOK 38 187 686, NOK 40 797 757, NOK 44 079 949 and NOK 48 910 820 over the consecutive four-year diagnostic intervals 2000-2003, 2004-2007, 2008-2011, and 2012-2015, respectfully.

Future research on several of the aspects in this thesis is highly encouraged, especially due to the lack of explanatory variables, such as treatment and other patient related factors. Further analysis examining the relation between county of residence and both stage of diagnosis and survival time is encouraged. In addition, the uncertainty surrounding the cost estimates should be examined further in order to fully capture the actual observed treatments and coherent costs.

6.2 Comparability to previous research

As explained in the introduction, the inspiration behind this thesis is the British report by Cancer Research UK “Saving lives, averting costs” [7], which investigated the costs and effects of earlier diagnosis in colon, rectum and NSCLC, among others. In England, the national stage I-IV distribution was: 13 %, 31 %, 32 % and 24 % (colon); 26 %, 22 %, 29 % and 23 % (rectum); 15 %, 8 %, 22 % and 55 % (NSCLC), respectfully. This was after assigning the proportion of unstaged patients were allocated to each stage according to the proportions observed with staged patients. The number of patients with unknown stage was 12 % for colorectal cancer and 15 % for NSCLC. Compared to the Norwegian stage distribution identified in this thesis, although aggregated to a three-stage system, indicates a slightly better national average than

that observed in England. In addition, the number of reported patients with unknown stage was substantially lower in Norway, with 8.2 % for colorectal cancer and 11.6 % for lung cancer.

Within England, there was a threefold variation in early stage distribution between the *best* and *worst* region for patients with colorectal cancer, and a fourfold for lung cancer. This is a significantly larger variation than what was found in this thesis between the Norwegian counties, suggesting less regional unwanted variation.

For comparison of costs to the British report, the PPP-calculation has been performed in order to more correctly compare results. The costs for stage I colon cancer estimated by the British report was £ 3 608 and £ 4 760 for rectal cancer. Stage IV costs were estimated to be £ 13 395 and £12 642, respectfully. This is substantially lower than the cost estimated by Joranger et.al [4] which was applied in this thesis. In addition, the costs estimated from the British report for lung cancer was £ 5 700 for stage I and £ 16 137 for stage IV. This is also substantially lower than the model estimation for lung cancer from this thesis.

The British report did not calculate the per patient potential cost savings. They did however, estimate the financial implications of achieving the same level as the current best stage distribution of in England from a yearly perspective. They estimated a yearly cost saving of this scenario of £ 36 444 378 for colorectal cancer, which is substantially higher than what was estimated in this thesis. When adjusting for the incidence of colorectal cancer in the UK, the cost saving results are smaller. This is due to the substantially higher cost of colorectal cancer treatment applied in this thesis. For lung cancer they estimated an increase of costs of £ 6 930 893. This is quite contradictory what was estimated in this thesis. Again, this could be due to the substantially higher cost of lung cancer treatment estimated by the pathway model, and applied in this thesis.

6.3 Interpretation of results

Over a long period of time there has been concerns regarding regional differences in several aspects within the different levels of health care in Norway. There has been major scientific evolvement, multiple national reforms, and the quality of care has risen over the last decades. However, regional differences still exist and are under much debate. A disruption of the principles on which the Norwegian healthcare rests could and should cause major concern.

The identification of regional differences in survival supports the disruption within the principle. With patients diagnosed with lung cancer residing in Vest-Agder having a 27.7% higher 1-year survival probability than patients in Oppland indicates a significant regional difference in survival with regards to the patient's county of residence at time of diagnosis. While patients diagnosed with colorectal cancer residing in Akershus at time of diagnosis can expect to live on average 14.34 months longer that patients living in Oslo.

The parametric Weibull PH model predicted a longer median survival times for all stages compared to the non-parametric KM estimator. However, the survival calculations of the incremental gains in months between the two methods are in close proximity. The non-parametric method is preferred because it does not require a certain distributional assumption. However, estimations of parametric methods are stronger than non-parametric methods. This is however reliant on the correct identification of distributional family. The Weibull PH was identified though model adjudication, however, since the Weibull PH was not compared to the semi-parametric Cox regression model, the Cox model could predict survival better than the Weibull PH.

There were no other prognostic factors other than age, gender and diagnostic stage at time of diagnosis included in the dataset from CRN. Although the findings from this thesis identified a clear difference in both diagnostic stage distribution and survival time between regions in Norway, other prognostic factors or demographic variables are expected to have a major impact. Prognostic factors which are not identified though the dataset are possible explanations, such as more precise stage explanation (TNM stage), overall health and possible comorbidities. Furthermore, the prognostic factors such as health before diagnosis and other comorbidities might be rooted in cultural and demographic factors such as education, income, regional dietary norms, relationship status, distance (urban/rural) to primary or specialized care, and many others. The risk of smoking was addressed in the background of this thesis, yet this is difficult to adjust for without information on the smoking history of the patients in the dataset. However, these are explanatory factors are believed to have a great impact on survival.

There were substantial health gains to be achieved by both regional best practice and a lower degree of cancer severity. The incremental difference in survival between the three diagnostic stages (overall) was remarkable, and emphasizes the significance of achieving an early diagnosis. The variation in stage distribution and survival time, both regionally and nationally, together with other main findings outlined in Chapter 6.1 implies significant regional

differences which are susceptible to several possible implications. The finding supports the debate of regional differences, that even though both treatment and quality is increasing, there still exists a geographical variation. The director of health at Hdir said as late as 24th of April 2018 that; "*Although there is an increasing usage of healthcare services, and the services as well as the treatments have improved, we are seeing large variation, both geographical and within the different levels of care*" [2].

The significance of access to primary care, and identification of regional differences as presented in Chapter 2.5.1 have previously been debated as a possible explanatory factor to variation in population health and health outcomes. Troms has the highest proportion of patients diagnosed with localized disease, and the second highest GP-inhabitant ratio. However, Sogn og Fjordane has the lowest proportion of patients diagnosed with localized disease and a high GP-inhabitant ration. This opens for other factors which could explain why some counties who seems to have a good GP-inhabitant ratio, ensuring good access, still diagnoses patients in a later stage of disease progression than counties with a lower ratio. For colorectal cancer, the number of GPs per 10 000 inhabitants and the percentage of patients diagnosed in each diagnostic stage presents an interesting interaction. Finnmark has the second highest percentage of patients with colorectal cancer diagnosed with localized disease (23.4 %), which is preferable, and they have the highest GP-inhabitant ratio. While Sør-Trøndelag has a low percentage with 17.9 % diagnosed with localized disease, and a low GP-inhabitant ratio.

However, there seems to be little coherency between the number of GPs per 10 000 inhabitants and survival time for both cancer. This is, among others, indicated by Akershus county, which has the longest median survival time for patients diagnosed with colorectal cancer and the second lowest number of GPs per inhabitant. Further displayed by Oppland county which has a considerably high number of GPs, yet has significantly shorter median survival time than Akershus. The GP-inhabitant ratio displays little significance on overall survival. The number of GPs per inhabitant and stage of disease could therefore indicate a positive connection to higher access to primary care, yet this does not seem to influence future survival time, which opens for other possible differences in a higher level of care.

[2] Translated 25th of April 2018 from: https://helsedirektoratet.no/nyheter/helsetjenestene-i-norge-bedre-kvalitet-men-fortsatt-for-store-forskjeller

There is high number of patients which stage is not recorded at time of diagnosis. As depicted and estimated throughout the analysis, the patients which are registered with unknown stage indicates a poor survival time, suggesting that some of those patients where stage is not recorded may have been diagnosed late. The lack of staging could interfere with survival time, as the specter of survival time between stages are very broad.

Regional variation in survival time could be explained by treatment variations in specialized health services, which is cause for concern. Since direct treatment of cancer is provided at only a few specialized hospitals, it is troublesome that the county of residence at time of diagnosis could influence survival time. The treatment strategies outlined in Chapter 2.2.2 are based on the national guidelines and not actual observed treatments through the CRN dataset, so it is not possible to identify any potential occurrence of treatment variation between the counties and/or hospitals. Since the dataset does not include hospital of primary (or any) treatment, it is not possible to analyze hospital variations.

Due to the extent of patients with unknown stage registered to the CRN and the lack of other explanatory variables such as treatment and hospital of care mentioned above, there is need for further work on documentation and recording. The director of CRN, has referred to this and claims there is "*a black hole in cancer treatment*" [3] as late as 30th of April 2018 in the media. The director of CRN hypothesize the lack of documentation to a lack of "tradition" among physicians to report such treatment variables. The article raises the discussion on the lack of reporting and unwanted regional difference in treatment, which is not possible to investigate without documentation on this. Due to the analysis preformed in this thesis, and the discovery of significant regional difference in cancer survival, such variables are in great need in order to examine if variations in specialized care are the potential reason for the regional survival variations. Furthermore, with the close geographical proximity of Akershus and Oslo, this substantial difference in survival time appears troublesome since these two counties also share the majority of patients between them and corresponding specialized health service affiliation.

This thesis estimated the coherency between early stage diagnosis and potential financial dividend for the Norwegian healthcare sectors. These findings create a compelling case to encourage early stage diagnosis. The director of health at Hdir further commented on 24th of

[3] Retrieved and translated 1st of May 2018 from: https://www.dagensmedisin.no/artikler/2018/04/30/-det-er-et-sort-hull-i-kreftbehandlingen/

April 2018 that; "*These (…regional) variations apply to both the extent of service, treatment practice, waiting times and resource utilization*"[4], which underscores the importance of resource utilization within the Norwegian healthcare system. Although the per patient savings appears small, colorectal and lung cancer represents a large proportion of cancer cases in Norway, yielding a large financial dividend for earlier diagnosis for these patient groups. With any strained national budget, monetary spending averted yields the possibility of resource allocation to other areas, possibly remaining within the healthcare system.

A reduction in morality are implicitly a goal in the 2013 national strategy against cancer; "Together against cancer – National cancer strategy 2013-2017". With the demonstrated cost-effectiveness of a colorectal cancer screening program, a national program will begin in 2019, which will most likely reduce the mortality from colorectal cancer. However, it would be interesting to examine any potential regional difference in participation and if this program could help to limit or dissolve the regional differences in survival time for patients with colorectal cancer. There is yet to be proposed a national screening program for lung cancer, due to uncertain and/or not proven cost-effectiveness, and the decision is to await until more and better documentation is available.

6.4 Limitations

Like with any research or study, there exists limitations within the study and any result should be understood with a degree of caution. The series of limitations exists around the different aspects of the thesis.

There are limitations regarding the costing parameters in the costing model for lung cancer treatment. Firstly, since the CRN dataset does not differentiate between NSCLC and SCLC, only costs related to the treatment of NSCLC was considered in the model. With approximately 95 % of lung cancer cases classified as NSCLC, this assumption was applied. In addition, the overall cost estimates produced by the model most likely does not reflect the true cost of lung cancer treatment, due to the known heterogeneity of cancer and cancer treatment. Since model was built on the basis of the guidelines provided by Hdir, and not actually observed treatment through the dataset, it is difficult to know if the patients received the recommended treatments,

[4] Translated 25th of April 2018 from: https://helsedirektoratet.no/nyheter/helsetjenestene-i-norge-bedre-kvalitet-men-fortsatt-for-store-forskjeller

and hence the adjacent costs. There are also most likely an underestimation of costs, which is common in *incidence costing*. Incidence cancer costs are computed from the time of diagnosis and represent the costs of cancer from an individual perspective which may be aggregated over individuals to provide estimates of the costs of newly diagnosed disease. As before mentioned, this yields a limitation since there is no information as to which patients actually received what main treatment, the actual number of consultations, hospital days, and pharmaceutical prescription. These costs and probabilities of treatment was instead collected through previous studies on NSCLC. The DRG-costs also does probably not cover the full extent on costs within the specialized health services, and the cost of the pharmaceutical regimes were collected through publications from the UK (National Institute for Health and Clinical Excellence), which are probably different to a certain degree.

Furthermore, there exists limitations imbedded in the modelling exercise. The memoryless feature of the state-transition model assumes that transition probabilities to the next state depend only on the current state, disregarding any prior state. However, since the state-transition model applied in this thesis only emphases costs, and not utility, the memoryless property is less invading towards the extent of limitations. There are however certain sources that potentially could deteriorate the model accuracy. Especially the assumption related to "structural uncertainty" – where a series of assumptions made through the model, potentially being an inappropriate one. There could be other model structures deemed more appropriate than the one chosen for this thesis.

In addition, the model duration was limited to 5 years. This duration was chosen due to the fast progression and short life-time of patients diagnosed with NSCLC, yet a longer model duration could alter the costing results. Since the costs were estimated over a 5-year duration, cost of follow-up after this is excluded. In addition, patients could suffer a recurrence at any point and within this model, entailing that any recurrence occurring more than 5 years since primary diagnosis are not included. Since the cost of recurrence are substantial, this could potentially increase the cost estimates for the earlier stage lung cancer patients and lower the cost saving estimates. However, previous literature on lung cancer recurrence indicates that 90 % of all recurrences occur within the first year after primary treatment, which justified the decision of a five-year time horizon.

There are also limitations regarding the dataset received from the CRN. When using registry data, there are always concern about the input reliability and accuracy, as well as the overall

data quality. There are always potential errors in raw register data. Even with the measures taken to remedy errors and duplicate entries, there is still potential for others to occur. The process of transforming the data to make it appropriate for survival analysis could potentially create errors, yet these have been adjusted for to the extent which was possible. The greatest potential error in the survival analysis is the death variable, which was crudely inputted in the dataset, and the assumptions proposed by the analysis could be flawed. However, with lung cancer specifically, the fast progression of the disease and investigation of literature on cancer mortality suggest that death during cancer treatment, and sometime after reaching remission is often caused by cancer related comorbidities.

Other limitations include the weighting of the costs and incremental health gains to match the CRN dataset stage distribution. To fully estimate potential financial gains with earlier diagnosis, the regional variations and exploring the different scenarios certain assumptions was made, which are a likely source of limitation. These assumptions were made on the basis of previous literature and explored through the range of CI in cancer cost literature. There is an underlying premise that there needs to be an increase in cost with the increase in severity in order to achieve financial gains by earlier diagnosis. In addition, the scenarios themselves and the adjacent calculations could be a source of limitations. Other methods than the calculations of mean stage and top/bottom 30 % approach used, for counties above and below the national average could deemed more appropriate than the method chosen for this thesis. In addition, more precise stratification by other factors could have predicted a different result. With the substantial number of patients recorded with unknown stage and the survival suggesting poor could indicate that they may have been diagnosed late and should incur higher costs (e.g. costs coherent with metastatic disease) than what was estimated in this thesis. This also relates to the estimated health gains. Only the incremental gains between the registered diagnostic stage were calculated, hence the survival time for patients with unknown stage could alter the estimated health gains by earlier diagnosis.

Despite some limitations, this thesis makes an important contribution for the identification of regional difference regarding survival time. The analysis of health and cost gains by earlier diagnosis could contribute to place more emphasis on the significance of achieving an earlier diagnosis and resource utilization. The parametric estimation and identification of best model fit by the Weibull distribution could aid future analysis of register data for survival analysis.

7 Conclusion

The main objective of this thesis was to investigate stage-specific and regional variations in survival time for patients diagnosed with colorectal and lung cancer in Norway. In addition, stage-specific costs for the two cancers were estimated, both trough a self-created model exploring the pathway for lung cancer treatment, and a previously published decision-analytic model for colorectal cancer, both from a Norwegian perspective.

There are considerable differences in both stage distribution and survival between the Norwegian counties. The main non-parametric estimation reveled a regional variation (by county of residence) in median survival time of between 6.19-9.23 months for lung cancer patients, and 48.24-66.59 months for patients diagnosed with colorectal cancer. Through model adjudication, the Weibull PH as the best model for the data. This aids future registry survival analysis studies to be performed with the Weibull distribution. The fully fitted Weibull PH model predicted an increase in survival time of 44.78 months for patients diagnosed with lung cancer and localized disease compared to patients with metastatic disease. The corresponding result for colorectal cancer patients was a gain of 168.8 months for patients diagnosed with localized at time of diagnosis. There are also substantial monetary savings to be achieved by earlier diagnosis, by applying a scenario of best regional practice, set by the county with the most “favorable” stage distribution. This yields a potential per patient and yearly cost saving for lung and colorectal cancer of NOK 12 269 and NOK 11 202, and approximately NOK 37 mill and NOK 50 million, respectfully, using the best practice scenario.

With the aim of the Norwegian healthcare sector to provide equal access and services to the population, regional difference in the prognostic staging distribution and survival probability could indicate a disruption within this principle. This thesis can function to serve as a starting point for future research and provides valuable insight to regional differences which may guide decision maker’s actions and future study within the field of cancer survival and resource utilization. Additional research should focus on gathering more detailed documentation and reporting on cancer treatment and socioeconomic variables. Furthermore, more detailed probabilistic decision-analytic modeling could investigate the uncertainty of cancer related costs, and also explore other financial scenarios for different sub-groups.

8 References

1. Norway, C.R.o., *Cancer in Norway 2016: Cancer incidence, mortality, survival and prevalence in Norway*. 2017, Cancer Registry of Norway: Oslo.

2. Norway, C.R.o., *Cancer in Norway 2016 - Cancer incidence, mortality, survival and prevalence in Norway*. 2017, Cancer Registry of Norway.

3. Ringard, Å.S., A. Saunes, I. Lindahl, A, *Norway - Health system review*, in *Health Systems in Transitions*. 2013, Kunnskapssenteret.

4. Joranger, P., et al., *Modeling and validating the cost and clinical pathway of colorectal cancer.* Med Decis Making, 2015. **35**(2): p. 255-65.

5. Gridelli, C., et al., *Non-small-cell lung cancer.* Nat Rev Dis Primers, 2015. **1**: p. 15009.

6. Haggar, F.A. and R.P. Boushey, *Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors.* Clin Colon Rectal Surg, 2009. **22**(4): p. 191-7.

7. InsisiveHealth, *Saving lives, averting costs: An analysis of the financial implications of achieving earlier diagnosis of colorectal, lung and overian cancer*. 2014: Cancer Research UK.

8. Xu, Y., et al., *Comparisons between the National Comprehensive Cancer Network (NCCN) non-small-cell lung cancer (NSCLC) Clinical Practice Guidelines (Chinese version), the NCCN original edition, and the European Society for Medical Oncology NSCLC Guidelines in 2009.* Thorac Cancer, 2010. **1**(2): p. 83-86.

9. Islami, F., L.A. Torre, and A. Jemal, *Global trends of lung cancer mortality and smoking prevalence.* Transl Lung Cancer Res, 2015. **4**(4): p. 327-38.

10. Zisman, A.L., et al., *Associations between the age at diagnosis and location of colorectal cancer and the use of alcohol and tobacco: implications for screening.* Arch Intern Med, 2006. **166**(6): p. 629-34.

11. Croce, C.M., *Oncogenes and cancer.* N Engl J Med, 2008. **358**(5): p. 502-11.

12. Michor, F., Y. Iwasa, and M.A. Nowak, *Dynamics of cancer progression.* Nat Rev Cancer, 2004. **4**(3): p. 197-205.

13. Crowell, R.E., et al., *Regional disparities in treatment and survival of early stage non-small cell lung cancer.* Ethn Dis, 2007. **17**(2): p. 358-64.

14. Masoomi, H., et al., *Population-based evaluation of adenosquamous carcinoma of the colon and rectum.* Dis Colon Rectum, 2012. **55**(5): p. 509-14.

15. Markowitz, S.D. and M.M. Bertagnolli, *Molecular origins of cancer: Molecular basis of colorectal cancer.* N Engl J Med, 2009. **361**(25): p. 2449-60.

16. Wittekind, C., *[Lymph nodes, tumour deposits, and TNM: are we getting better? 7th edition of UICC 2010 TNM classification of malignant tumors]*. Strahlenther Onkol, 2012. **188**(2): p. 191-2.

17. Fisher, E.R., et al., *Dukes' classification revisited. Findings from the National Surgical Adjuvant Breast and Bowel Projects (Protocol R-01).* Cancer, 1989. **64**(11): p. 2354-60.

18. Amin, M.B., et al., *The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging.* CA Cancer J Clin, 2017. **67**(2): p. 93-99.

19. Norway, C.R.o., *Årsrapport 2016 med resultater og forbedringstiltak fra Nasjonalt kvalitetsregister for lungekreft*. 2017: Cancer Registry of Norway.

20. Seeff, L.C., et al., *Comprehensive cancer control programs and coalitions: partnering to launch successful colorectal cancer screening initiatives.* Cancer Causes Control, 2010. **21**(12): p. 2023-31.

21. Bray, F., et al., *[Epidemiology of colorectal cancer in Norway]*. Tidsskr Nor Laegeforen, 2007. **127**(20): p. 2682-7.

22. Norway, C.R.o., *Årsrapport 2015*, N.k.f.t.-o. endetarmskreft, Editor. 2016, Kreftregisteret.

23. Sagerup, C.M.T., et al., *Sex-specific trends in lung cancer incidence and survival: a population study of 40 118 cases.* Thorax, 2011. **66**(4): p. 301-307.

24. Health, N.D.o., *Nasjonalt handlingsprogram med retningslinjer for diagnostikk, behandling og oppfølging av kreft i tykktarm og endetarm.* , N.D.o. Health, Editor. 2017, Norwegian Directorate of Health: Oslo.

25. Health, N.D.o., *Nasjonalt handlingsprogram med retningslinjer for diagnostikk, behandling og oppfølging av lungekreft, mesoteliom og thyom*, N.M.o. Health, Editor. 2017, Norwegian Directorate of Health: Oslo. p. 238.

26. Health, N.D.o., *Nasjonalt handlingsprogram med retningslinjer for diagnostikk, behandling og oppfølging av lungekreft, mesoteliom og thyom.* 2017.

27. Health, N.D.o., *Nasjonalt handlingsprogram med retningslinjer for diagnostikk, behandling og oppfølging av kreft i tykktarm og endetarm.* 2017.

28. Brustugun, O.T. *Om kreft i lunge*. 2014; Available from: http://oncolex.no/Lunge.

29. Nesbakken, A. *Om kreft i tykktarm/endetarm*. 2016; Available from: http://oncolex.no/Tykktarm.

30. Morgensztern, D., et al., *Trends in Stage Distribution for Patients with Non-small Cell Lung Cancer A National Cancer Database Survey.* Journal of Thoracic Oncology, 2010. **5**(1): p. 29-33.

31. Butler, J., et al., *The International Cancer Benchmarking Partnership: an international collaboration to inform cancer policy in Australia, Canada, Denmark, Norway, Sweden and the United Kingdom.* Health Policy, 2013. **112**(1-2): p. 148-55.

32. Economics, O., *Kreft i Norge - Kostnader for pasientene, helsetjenesten og samfunnet*. 2016, Oslo Economics.

33. Chouaid, C., et al., *Economics of the clinical management of lung cancer in France: an analysis using a Markov model.* Br J Cancer, 2004. **90**(2): p. 397-402.

34. Corral, J., et al., *Estimation of lung cancer diagnosis and treatment costs based on a patient-level analysis in Catalonia (Spain).* Bmc Health Services Research, 2015. **15**.

35. Haug, U., et al., *Estimating colorectal cancer treatment costs: a pragmatic approach exemplified by health insurance data from Germany.* PLoS One, 2014. **9**(2): p. e88407.

36. Mar, J., et al., *The cost of colorectal cancer according to the TNM stage.* Cir Esp, 2017. **95**(2): p. 89-96.

37. Zappa, C. and S.A. Mousa, *Non-small cell lung cancer: current treatment and future advances.* Transl Lung Cancer Res, 2016. **5**(3): p. 288-300.

38. Services, N.M.o.H.a.C., *Principles for priority settning in health care.* 2017.

39. Helsedirektoratet, *Leger i kommune- og spesialisthelsetjenesten*. 2017, Helsedirektoratet - Avdeling spesialisering og etterutdanning.

40. HOD, *Strategi: Sammen - mot kreft, nasjonal kreftstrategi 2013-2017*, H.-o. Omsorgsdepartementet, Editor. 2013, Helse- og Omsorgsdepartementet: Oslo.

41. Avila, H.D., T. Eide, M. Vaage, L, *Pakkeforløp for kreft - Evaluering av pakkeforløp med vekt på arbeidsprosesser og organisering*. 2016: Oslo.

42. Norway, C.R.o.; Available from: https://www.kreftregisteret.no/contentassets/6b389374314344d8a37726aa4d79c732/dokumentasjon_av_variabler-v4.1.pdf.

43. Briggs, A.H., K. Claxton, and M.J. Sculpher, *Decision modelling for health economic evaluation*. Oxford handbooks in health economic evaluation. 2006, Oxford: Oxford University Press. x, 237 p.

44. Kapadia, N.S., et al., *Patterns of Treatment and Outcomes for Definitive Therapy of Early Stage Non-Small Cell Lung Cancer.* Ann Thorac Surg, 2017. **104**(6): p. 1881-1888.

45. Rowell, N.P. and C.J. Williams, *Radical radiotherapy for stage I/II non-small cell lung cancer in patients not sufficiently fit for or declining surgery (medically inoperable).* Cochrane Database Syst Rev, 2001(2): p. CD002935.

46. Sugimura, H., et al., *Survival after recurrent nonsmall-cell lung cancer after complete pulmonary resection.* Ann Thorac Surg, 2007. **83**(2): p. 409-17; discussioin 417-8.

47. Lou, F., et al., *Patterns of recurrence and second primary lung cancer in early-stage lung cancer survivors followed with routine computed tomography surveillance.* J Thorac Cardiovasc Surg, 2013. **145**(1): p. 75-81; discussion 81-2.

48. Lou, F., et al., *Differences in patterns of recurrence in early-stage versus locally advanced non-small cell lung cancer.* Ann Thorac Surg, 2014. **98**(5): p. 1755-60; discussion 1760-1.

49. Berthelot, J.M., et al., *Decision framework for chemotherapeutic interventions for metastatic non-small-cell lung cancer.* J Natl Cancer Inst, 2000. **92**(16): p. 1321-9.

50. Sundstrom, S., et al., *Hypofractionated palliative radiotherapy (17 Gy per two fractions) in advanced non-small-cell lung carcinoma is comparable to standard fractionation for symptom control and survival: a national phase III trial.* J Clin Oncol, 2004. **22**(5): p. 801-10.

51. Harrison, L.D., et al., *Comparing effectiveness with efficacy: outcomes of palliative chemotherapy for non-small-cell lung cancer in routine practice.* Curr Oncol, 2015. **22**(3): p. 184-91.

52. Johnson, D.H., et al., *Randomized phase II trial comparing bevacizumab plus carboplatin and paclitaxel with carboplatin and paclitaxel alone in previously untreated locally advanced or metastatic non-small-cell lung cancer.* J Clin Oncol, 2004. **22**(11): p. 2184-91.

53. Norum, J., et al., *Pembrolizumab as second-line therapy in non-small cell lung cancer in northern Norway: budget impact and expected gain-a model-based analysis.* ESMO Open, 2017. **2**(3): p. e000222.

54. Strom, H.H., et al., *Concurrent palliative chemoradiation leads to survival and quality of life benefits in poor prognosis stage III non-small-cell lung cancer: a randomised trial by the Norwegian Lung Cancer Study Group.* Br J Cancer, 2013. **109**(6): p. 1467-75.

55. Fairchild, A., et al., *Palliative thoracic radiotherapy for lung cancer: a systematic review.* J Clin Oncol, 2008. **26**(24): p. 4001-11.

56. Evans, W.K., et al., *Diagnostic and therapeutic approaches to lung cancer in Canada and their costs.* Br J Cancer, 1995. **72**(5): p. 1270-7.

57. Batson, S., et al., *Tyrosine kinase inhibitor combination therapy in first-line treatment of non-small-cell lung cancer: systematic review and network meta-analysis.* Onco Targets Ther, 2017. **10**: p. 2473-2482.

58. Bendaly, E., et al., *Treatment Patterns and Early Outcomes of ALK-Positive Non-Small Cell Lung Cancer Patients Receiving Ceritinib: A Chart Review Study.* Adv Ther, 2017. **34**(5): p. 1145-1156.

59. Zimmermann, F., et al., *Stereotactic body radiation therapy for early non-small cell lung cancer.* Front Radiat Ther Oncol, 2010. **42**: p. 94-114.

60. Hashizume, T., *[Complete remission of locally advanced squamous cell lung cancer induced by chemotherapy with cisplatin and docetaxel]*. Gan To Kagaku Ryoho, 2004. **31**(1): p. 83-5.

61. Hassett, M.J., et al., *Survival after recurrence of stage I-III breast, colorectal, or lung cancer.* Cancer Epidemiol, 2017. **49**: p. 186-194.

62. Drummond, M., *Methods for the economic evaluation of health care programmes*. Fourth edition. ed. Oxford medical publications. 2015, Oxford, United Kingdom ; New York, NY, USA: Oxford University Press. xiii, 445 pages.

63. Clark, T.G., et al., *Survival analysis part I: Basic concepts and first analyses.* British Journal of Cancer, 2003. **89**(2): p. 232-238.

64. Klein, J.P., *Handbook of survival analysis*. Chapman & Hall/CRC handbooks of modern statistical methods. 2014, Boca Raton: CRC Press, Taylor & Franciss Group. xvi, 640 pages.

65. Bradburn, M.J., et al., *Survival Analysis Part II: Multivariate data analysis - an introduction to concepts and methods.* British Journal of Cancer, 2003. **89**(3): p. 431-436.

66. Sorensen, S.V., et al., *Incidence-based cost-of-illness model for metastatic breast cancer in the United States.* Int J Technol Assess Health Care, 2012. **28**(1): p. 12-21.

67. Kakushadze, Z., R. Raghubanshi, and W.L. Yu, *Estimating Cost Savings from Early Cancer Diagnosis.* Data, 2017. **2**(3).

68. Chibaudel, B., et al., *Targeted therapies as adjuvant treatment for early-stage colorectal cancer: first impressions and clinical questions.* Clin Colorectal Cancer, 2010. **9**(5): p. 269-73.

69. Saisho, S., et al., *Post-recurrence survival of patients with non-small-cell lung cancer after curative resection with or without induction/adjuvant chemotherapy.* Interact Cardiovasc Thorac Surg, 2013. **16**(2): p. 166-72.

70. Socinski, M.A., et al., *Treatment of non-small cell lung cancer, stage IV: ACCP evidence-based clinical practice guidelines (2nd edition).* Chest, 2007. **132**(3 Suppl): p. 277S-289S.

71. Rostad, H., et al., *Lung cancer surgery: the first 60 days. A population-based study.* Eur J Cardiothorac Surg, 2006. **29**(5): p. 824-8.

72. (NICE), N.I.f.H.a.C.E., *The diagnosis and treatment of lung cancer - Full guideline.* 2011.

Appendices

9.1 Data approval

Data approval from the Cancer Registry of Norway

1.1 Dato	01-23-2018
1.2 Prosjektleders navn	Eline Aas
1.3 Stilling/akademisk grad	PhD
1.4 Firma/institusjon	Universitetet i Oslo
1.5 Adresse	PB 1089, Blindern 0317 Oslo
1.6 Telefon	41141924
1.7 E-post	eline.aas@medisin.uio.no
1.8 Forskningsansvarlig institusjon	Nina Vøllestad
1.9 Fakturaadresse	1.09) Fakturaadresse: 4350@invoicecenter.net <mailto:4350@invoicecenter.net> 1.10) Merknad til faktura: Stine Bergliot Olsen og «Masteroppgave UiO, Ida Laurendz Kommandantvold»
1.10 Merknad til faktura	Faktura skal sendes til Kreftforeningen. Dette er et prosjekt ved UiO, men datainnsamling skal betales av Kreftforenignen
2.1 Prosjektets tittel	
2.2 Prosjektstart og prosjektslutt	Start 01.01.18 Slutt 31.12.19
2.3 Tidspunkt for når de omsøkte data skal slettes/anonymiseres?	01-31-2025

4. Dersom det svares "ja" på noen av nedenstående spørsmål, skal det legges ved dokumentasjon som bekrefter godkjenning. Både søknaden og selve godkjenningen skal legges ved.

	Ja	Nei
4.1 Er det nødvendig med godkjenning av andre databehandlingsansvarlige institusjoner? *	-	
4.2 Er prosjektet fremleggelsespliktig for Regional komité for medisinsk og helsefaglig forskningsetikk (REK)? **	-	
4.3 Krever prosjektet konsesjon fra Datatilsynet?	-	
4.4 Krever prosjektet samtykke fra deltakerne? ***	-	
4.5 Dersom nei i punkt 4.3, er deltakerne informert på annen måte? ****	-	
4.6 Er det nødvendig å søke dispensasjon fra taushetsplikten for å få tilgang på data?	-	

5.1 Antall vedlegg til søknaden:	2
5.2 Forskningsprotokoll	Forskningsprotokoll
5.3 Andre vedlegg	CV prosjektleder
5.4 Eventuell annen relevant informasjon	Meldeskjema sendes til Datatilsynet

Data approval from the Norwegian Data Protection Authority (DPA).

Datatilsynet mottok 2018-01-23 13:50:20 en melding med referanse til denne e-postadressen.
Meldingsnummeret for denne meldingen er: 82993
Meldingsnøkkelen for denne meldingen er: 7ccf43f15b0f0a6b68cd412a13bd983d
Behandlingsansvarlig virksomhet er oppgitt til:
Universitetet i Oslo, Avdeling for helseledelse og helseøkonomi

9.2 Background

Table A 1: Recommended treatment guidelines by stage of diagnosis for lung and colorectal cancer.

Lung cancer

Patients diagnosed in stage I and II typically have surgery to remove either a lobe or section of the lung. This is done if the tumor is found to be resectable, often in combination with the patients' overall health and suspected tolerance of surgery. In some cases however, the tumor is non-resectable, which means that the location hinders good surgical vision/access, or other health issues may prevent the patient from being eligible for surgical interventions. If the tumor is non-resectable or the patient is unfit for surgery, the patient could receive stereotactic body radiation therapy (SBRT), which delivers a single high dose of radiation therapy in a highly precise location. After surgery, the removed tissue of the lung is examined to identify if there are cancerous cells on the edges of the removed section, call *positive margins*. If positive margins are identified, a second surgery could be performed in order to ensure that all the cancer has been removed. Another option to remove the positive margins is use radiation therapy after surgery. For tumors which shows indication of possible recurrence after surgery, based on size, location and other factors, adjuvant chemotherapy after surgery is given to lower the risk of recurrent cancer. However, while clinical studies of provision of adjuvant chemotherapy have found it to hinder or delay recurrence, not all patients receive adjuvant therapy due to inconclusive results. For stage IIB tumors, adjuvant chemotherapy, regardless of whether or not positive margins are identified. Neo-adjuvant chemotherapy is also used for stage IIB patients to shrink the tumor before surgery [28, 37].

While stage I and II often receives similar treatment strategies, stage III is more complex and requires inputs and consultations from a medical oncologist, radiation oncologist and a thoracic surgeon. In stage IIIA, the treatment depends on both size and location of the tumor, to which lymph nodes the cancer has spread, overall health and predicted tolerability among the treatment strategies. The treatment options includes some combination of radiation therapy, chemotherapy, and/or surgery. For patients who indicates tolerability, treatment often starts with chemotherapy in combination with radiation therapy. Surgery might be an option following initial treatment if the thoracic surgeon finds the tumor resectable and whether any remaining cancer can be removed. If surgery is performed, another chemotherapy regime is also given after surgery [28, 69].

In stage IIIB the cancer has spread to lymph nodes that are in proximity to the other lung of in the neck, and might also have grown into important structures in the chest. This cancer cannot be removed by surgery. As with previously, treatment depends on the patients overall health. Intensive chemotherapy regimes in combination with radiation therapy in otherwise healthy patient are given with curative intent. Stage IIIB lung cancer offers poor prognosis for patients with poor overall health, and are often offered participation in clinical trials with life-extending intent [28].

Stage IV lung cancer has metastasized when diagnosed and are most often not given treatment with curative intent. Lung cancer most commonly spread the brain. If the patient is otherwise good health, surgery, chemotherapy, targeted therapy, immunotherapy and radiation therapy is given to prolong life-expectancy and to relive symptoms. If the cancer has only metastasized to one other site, and the patient is in otherwise good health, curative treatments might be considered. This is however uncommon, and treatment for the lung cancer itself is then based on the T and N stages and may include surgery, chemotherapy, radiation therapy, or some combination of these [70].

For stage IV lung cancer that has spread widely throughout the body, the tumor will be tested for gene mutations (such as the EGFR, ALK, ROS 1 or BRAF genes). If one of these genes are mutated in the cancer cells the first-line treatment can include targeted therapy [28].

Stage I colon cancer includes cancers that were part of a polyp, and if the polyp is removed completely during a colonoscopy and there are no positive margins of the removed piece, there might be no need for further treatment. If the cancer in the polyp is high grade or there are positive margins, further surgical intervention is required. Surgical treatment is also performed if the polyp could not be completely removed. For cancers that did not originate for a polyp, surgery to remove the section of the colon that has cancerous cells, a partial colectomy, is the standard treatment. Further treatment is usually not necessary. In stage I rectal cancer, the same procedure is performed if the cancer originated from a polyp. If not, surgical inventions depends on location in the rectum. This includes resecting the cancerous area through the anus without abdominal incision is the tumor is small, using transanal endoscopic microsurgery (TEM). For larger tumors within stage I rectal cancer, a low anterior resection (LAR), proctectomy with colo-anal anastomosis, or an abdominoperineal resection (APR) may be performed, again depending on location of tumor in the colon [27, 29, 68].

In stage II colon cancer, the tumor has grown through the wall of the colon and possibly into nearby tissue. Surgery to remove the section of the colon invaded by the tumor, a partial colectomy, in addition to the nearby lymph nodes is the primary treatment strategy. If the cancer shows indication of possible recurrence, adjuvant chemotherapy is given. In colon cancer, certain factors indicates a higher risk of recurrence. This includes whether the cancerous cells appear very abnormal (high grade), the cancer has grown into nearby blood or lymph vessels, surgical removal of less than 12 lymph nodes, positive margins of the removed section, the tumor had obstructed the colon of caused a perforation of the wall of the colon. For patients with stage II rectal cancer, the treatment strategy involves chemotherapy, radiation therapy and surgery, in different orders depending on the patients overall health. The most common treatment order is a combination of chemotherapy and radiation therapy, followed by one of the surgical methods mentioned for stage I rectal cancer. The treatment pathway usually ends with a 6 month chemotherapy regime [27, 29, 68].

In stage III colon cancer, the cancer has spread to nearby lymph nodes. A partial colectomy and surgery to remove nearby lymph nodes, followed by adjuvant chemotherapy is the standard treatment for this stage of progression. A combination of radiation therapy and chemotherapy is the main option for patients who are in poor overall health. The treatment of stage III rectal cancer is similar to the treatment received in stage II. However, in stage III, chemotherapy is most often given along with radiation therapy, called chemoradiation. The intent is to shrink the cancer, resulting in less invasive surgery. Chemoradiation also hinders recurrence in the pelvis and has less side-effects that radiation after surgery [27, 29, 68].

Stage IV colon cancer involves metastasis to distant organs and tissue. Colon cancer most commonly spread to liver, but also to the lung, brain, peritoneum (lining of the abdominal cavity) or to distant lymph nodes. Surgery is unlikely to be given with curative intent, but is considered with life-prolonging intent if only a few small areas of metastasis. If major metastasis, chemotherapy is the main treatment. Surgery might be given if the cancer is blocking the colon to relive symptoms. The main intent of treatment of stage IV colon cancer is to control the cancer, and is provided through target therapy regimes. The choice of regimens depends on several factors, including any previous treatment received.

For advanced colon cancer, radiation therapy can be used to prevent or relive symptoms. In patients with gene changes in their cancer cells, immunotherapy drugs after initial chemotherapy is an option. The treatment option for stage IV rectal cancer is considered more invasive and most often involves surgery, chemotherapy and radiation therapy. If the cancer is more widespread and cannot be removed by surgery, treatment with chemotherapy and/or target therapy drugs is likely. As with colon cancer, immunotherapy drugs after initial chemotherapy is an option for patients with specific cancer gene changes [27, 29, 68].

Table A 2: Total healthcare spending for primary and secondary health care services from 2011-2014 for diagnostic and treatment of all cancers. All values given in mil NOK.

	2011	2012	2013	2014
Primary health services	**362**	**397**	**431**	**472**
Secondary health services	**10 552**	**11 093**	**11 873**	**12 983**
Private specialists	86	87	95	105
Outpatient care	147	143	146	174
Inpatient care	6 337	6 368	6 979	7 132
Outpatient consultation	2 848	2 402	2 595	2 717
Outpatient diagnostic imaginary[a]	248	288	302	324
Outpatient laboratory tests[a]	366	425	446	478
Other costs in the secondary health services[b]	520	1 380	1 311	2 053
Pharmaceuticals (extradited in pharmacies)[c]	**743**	**833**	**921**	**1 054**
Share of cancer specific pharmaceuticals extradited in pharmacies[c]	501	577	679	810
Total health care costs	**11 657**	**12 323**	**13 225**	**14 509**

[a)] 2015 prices, adjusted for growth in lab-/x-ray expenses. [b)] Including ambulatory services, pension, patient transportation, overhead costs, and administration. [c)] Pharmacy purchase price tax deducted. Source: [32]

Table A 3: General Practitioners (practices) according to counties. Per 31.12.16. All numbers are given in percentages (%)

County	Lists without regular GP*	GP lists with available capacity	Individuals without GP**	Individuals opting out***
Østfold	0.0	43.8	0.7	0.1
Akershus	0.4	50.0	0.7	0.2
Oslo	0.4	59.4	1.5	0.2
Hedmark	0.6	46.1	0.9	0.4
Oppland	1.6	48.2	0.6	0.2
Buskerud	0.0	55.7	1.0	0.1
Vestfold	0.0	49.2	0.7	0.1
Telemark	1.3	50.3	0.7	0.1
Aust-Agder	1.8	54.5	0.5	0.1
Vest-Agder	0.0	47.4	0.6	0.2
Rogaland	0.5	44.6	0.7	0.1
Hordaland	2.2	50.3	0.8	0.2
Sogn og Fjordane	4.0	56.3	1.5	0.2
Møre og Romsdal	3.1	47.1	0.9	0.2
Sør-Trøndelag	1.4	48.2	0.8	0.1
Nord-Trøndelag	3.1	42.3	0.5	0.1
Nordland	7.0	41.6	0.8	0.3
Troms	3.8	42.1	1.0	0.1
Finnmark	8.4	56.8	1.4	0.3
Unknown county[1]	-	-	40.0	21.2
Total	**1.6**	**49.4**	**1.3**	**0.4**

*Registered GP practice, but without regular GP doctor **Participating in the GP scheme, but not registered on any list *** Citizens opting out of the GP scheme voluntarily [1]Refugees and their families which are not registered citizens at a specific municipality, and individuals which are registered citizens but without registered address. Source. *[39]*

Table A 4: Number of specialist doctors according to regional health authority in 2017.

Health region	Total number of doctors*	Per 10 000 inhabitants
Helse Nord RHA	1 484	31
Helse Midt-Norge RHA	1 858	25
Helse Vest RHA	2 263	21
Helse Sør-Øst RHA	7 047	25

* Including both specialist doctors and employed doctors in specialized training (LIS). Source: [39]

9.3 Data and materials

Table A 5: Patient characteristics stratified by successive 4-year (2016 alone) and 17-year (Total 2000-2016) diagnostic period. Percentage (%) in parenthesis. Patients diagnosed with lung cancer by county, age-group and stage.

	2000-2003	2004-2007	2008-2011	2012-2015	2016	Total
No. (%)	**8 780**	**9 875**	**11 140**	**12 277**	**3 206**	**45 278**
County						
Østfold	558 (6.4)	577 (5.8)	710 (6.4)	869 (7.1)	198 (6.2)	**2 912 (6.4)**
Akershus	825 (9.4)	937 (9.5)	1 096 (9.8)	1 159 (9.4)	319 (10)	**4 336 (9.6)**
Oslo	936 (10.7)	1 014 (10.3)	1 043 (9.4)	1 011 (8.2)	268 (8.4)	**4 272 (9.4)**
Hedmark	392 (4.5)	473 (4.8)	502 (4.5)	575 (4.7)	132 (4.1)	**2 074 (4.6)**
Oppland	320 (3.6)	354 (3.6)	485 (4.4)	504 (4.1)	129 (4)	**1 792 (4)**
Buskerud	450 (5.1)	546 (5.5)	610 (5.5)	616 (5)	162 (5.1)	**2 383 (5.3)**
Vestfold	524 (6)	541 (5.5)	629 (5.7)	726 (5.9)	163 (5.1)	**2 583 (5.7)**
Telemark	362 (4.1)	379 (3.8)	418 (3.8)	449 (3.7)	118 (3.7)	**1 726 (3.8)**
Aust-Agder	242 (2.8)	267 (2.7)	323 (2.9)	349 (2.8)	82 (2.6)	**1 263 (2.8)**
Vest-Agder	373 (4.3)	432 (4.4)	476 (4.3)	485 (4)	145 (4.5)	**1 911 (4.2)**
Rogaland	684 (7.8)	699 (7.1)	842 (7.6)	1 034 (8.4)	272 (8.5)	**3 531 (7.8)**
Hordaland	768 (8.8)	851 (8.6)	978 (8.9)	1 173 (9.6)	313 (9.8)	**4 083 (9)**
Sogn og Fjordane	169 (1.9)	232 (2.4)	236 (2.1)	270 (2.2)	71 (2.2)	**978 (2.2)**
Møre og Romsdal	394 (4.5)	550 (5.6)	600 (5.4)	635 (5.2)	168 (5.2)	**2 347 (5.2)**
Sør-Trøndelag	509 (5.8)	589 (6)	645 (5.8)	651 (5.3)	171 (5.3)	**2 567 (5.7)**
Nord-Trøndelag	231 (2.6)	289 (2.9)	291 (2.6)	361 (2.9)	74 (2.3)	**1 246 (2.8)**
Nordland	534 (6.1)	570 (5.8)	602 (5.4)	646 (5.3)	184 (5.7)	**2 536 (5.6)**
Troms	297 (3.4)	342 (3.5)	373 (3.4)	387 (3.2)	81 (2.5)	**1 480 (3.3)**
Finnmark	170 (1.9)	191 (1.9)	219 (2)	229 (1.9)	61 (1.9)	**870 (1.9)**
Unknown	42 (0.5)	42 (0.4)	62 (0.6)	148 (1.2)	95 (3)	**389 (0.9)**
Age						
0-49	381 (4.3)	335 (3.4)	351 (3.2)	295 (2.4)	74 (2.3)	**1 436 (3.2)**
50-59	1 426 (16.2)	1 413 (14.3)	1 335 (12)	1 251 (10.2)	275 (8.6)	**5 700 (12.6)**
60-69	2 228 (25.4)	2 831 (28.7)	3 421 (30.7)	3 849 (31.5)	943 (29.4)	**13 272 (29.3)**
70-79	3 255 (37.1)	3 309 (33.5)	3 631 (32.6)	4 168 (34)	1 245 (38.8)	**15 608 (34.5)**
80-89	1 396 (15.9)	1 841 (18.6)	2 177 (19.5)	2 405 (19.6)	577 (18)	**8 396 (18.5)**
90 and over	94 (1.1)	146 (1.5)	225 (2)	309 (2.5)	92 (2.9)	**866 (1.9)**
Stage						
Localised	1 195 (13.6)	1 371 (13.9)	1 951 (17.5)	2 337 (19)	572 (17.8)	**7 426 (16.4)**
Regional	2 229 (25.4)	2 848 (28.8)	2 988 (26.8)	3 393 (27.6)	852 (26.6)	**12 310 (27.2)**
Metastatic	4 206 (47.9)	4 591 (46.5)	5 174 (46.5)	5 114 (41.7)	1 208 (37.7)	**20 293 (44.8)**
Unknown*	1 150 (13.1)	1 065 (10.8)	1 022 (9.2)	1 433 (11.7)	574 (17.9)	**5 244 (11.6)**

* Unknown stage at diagnosis.

Table A 6: Patient characteristics stratified by successive 4-year (2016 alone) and 17-year (Total 2000-2016) diagnostic period. Percentage (%) in parenthesis. Patients diagnosed with colorectal cancer by county, age-group and stage.

	2000-2003	2004-2007	2008-2011	2012-2015	2016	Total
No. (%)	**13 636**	**14 568**	**15 740**	**17 465**	**4 634**	**66 043**
County						
Østfold	827 (6.1)	921 (6.3)	1 057 (6.7)	1 120 (6.4)	285 (6.2)	**4 210 (6.4)**
Akershus	1 173 (8.6)	1 344 (9.2)	1 529 (9.7)	1 874 (10.7)	501 (10.8)	**6 421 (9.7)**
Oslo	1 417 (10.4)	1 334 (9.2)	1 484 (9.4)	1 522 (8.7)	392 (8.5)	**6 149 (9.3)**
Hedmark	641 (4.7)	671 (4.6)	679 (4.3)	718 (4.1)	181 (3.9)	**2 890 (4.4)**
Oppland	569 (4.2)	629 (4.3)	630 (4)	703 (4)	190 (4.1)	**2 721 (4.1)**
Buskerud	742 (5.4)	785 (5.4)	898 (5.7)	1 000 (5.7)	254 (5.5)	**3 679 (5.6)**
Vestfold	698 (5.1)	738 (5.1)	832 (5.3)	888 (5.1)	262 (5.7)	**3 418 (5.2)**
Telemark	473 (3.5)	503 (3.5)	508 (3.2)	573 (3.3)	159 (3.4)	**2 216 (3.4)**
Aust-Agder	237 (2)	307 (2.1)	333 (2.1)	373 (2.1)	93 (2)	**1 379 (2.1)**
Vest-Agder	435 (3.2)	463 (3.2)	536 (3.4)	573 (3.3)	144 (3.1)	**2 151 (3.3)**
Rogaland	1 164 (8.5)	1 144 (7.9)	1 329 (8.4)	1 407 (8.1)	390 (8.4)	**5 434 (8.2)**
Hordaland	1 452 (10.6)	1 618 (11.1)	1 670 (10.6)	1 806 (10.3)	478 (10.3)	**7 024 (10.6)**
Sogn og Fjordane	392 (2.9)	423 (2.9)	416 (2.6)	481 (2.8)	115 (2.5)	**1 827 (2.8)**
Møre og Romsdal	846 (6.2)	913 (6.3)	1 006 (6.4)	1 061 (6.1)	313 (6.8)	**4 139 (6.3)**
Sør-Trøndelag	830 (6.1)	849 (5.8)	857 (5.4)	1 000 (5.7)	261 (5.6)	**3 797 (5.8)**
Nord-Trøndelag	389 (2.9)	423 (2.9)	458 (2.9)	552 (3.2)	121 (2.6)	**1 943 (2.9)**
Nordland	746 (5.5)	846 (5.8)	814 (5.2)	982 (5.6)	268 (5.8)	**3 656 (5.5)**
Troms	369 (2.7)	444 (3.1)	494 (3.1)	565 (3.2)	157 (3.4)	**2 029 (3.1)**
Finnmark	160 (1.2)	161 (1.1)	165 (1.1)	205 (1.2)	48 (1)	**739 (1.1)**
Unknown	40 (0.3)	52 (0.4)	45 (0.3)	62 (0.4)	22 (0.5)	**221 (0.3)**
Age						
0-49	659 (4.8)	702 (4.8)	771 (4.9)	956 (5.5)	265 (5.7)	**3 362 (5.1)**
50-59	1 559 (11.4)	1 606 (11)	1 687 (10.7)	1 775 (10.2)	487 (10.5)	**7 114 (10.8)**
60-69	2 731 (20)	3 197 (22)	3 760 (23.9)	4 231 (24.2)	1 090 (23.5)	**15 009 (22.7)**
70-79	4 520 (33.2)	4 416 (30.3)	4 652 (29.6)	5 261 (30.1)	1 489 (32.1)	**20 338 (30.8)**
80-89	3 648 (26.8)	4 054 (27.8)	4 139 (26.3)	4 341(24.9)	1 064 (23)	**17 246 (26.1)**
90 and over	519 (3.8)	593 (4.1)	731 (4.6)	892 (5.1)	239 (5.2)	**2 974 (4.5)**
Stage						
Localised	2 747 (20.2)	2 990 (20.5)	2 655 (16.9)	3 590 (20.6)	1 018 (22)	**13 000 (19.7)**
Regional	6 230 (45.7)	7 167 (49.2)	8 336 (53)	8 795 (50.4)	2 253 (48.6)	**32 781 (49.6)**
Metastatic	3 115 (22.8)	3 207 (22)	3 757 (23.9)	3 795 (21.7)	952 (20.4)	**14 826 (22.5)**
Unknown	1 544 (11.3)	1 204 (8.3)	991 (6.3)	1 285 (7.4)	411 (8.9)	**5 435 (8.2)**

*

Table A 7: Non time dependent transition probabilities related to movement through the decision tree in the NSCLC model.

	Probability	Source
Stage I and surgery	0.80	[71]
Stage I and SBRT*	0.15	[59]
Stage I and curative RT**	0.05	[45]
Stage II and surgery	0.75	[71]
Stage II and curative RT	0.24	[45]
Stage III and surgery	0.28	[71]
Stage III and Chemotherapy + RT	0.58	[52]
Stage III and palliative treatment	0.14	[52]
Stage IV and curative intent	0.08	[52]
Stage IV and palliative treatment	0.92	[52]
Surgery stage I and response	0.89	[71]
Surgery stage I and poor response	0.11	[71]
SBRT stage I and response	0.82	[59]
SBRT stage I and poor response	0.18	[59]
Curative RT stage I and response	0.79	[45]
Curative RT stage I and poor response	0.21	[45]
Surgery stage II and response	0.81	[71]
Surgery stage II and poor response	0.19	[71]
Curative RT stage II and response	0.78	[45]
Curative RT stage II and poor response	0.12	[45]
Surgery stage III and response	0.18	[71]
Surgery stage III and poor response	0.82	[71]
Stage III and Chemotherapy + RT response	0.29	[52]
Stage III and Chemotherapy + RT poor response	0.71	[52]
Curative intent stage IV and response	0.01	[52]
Curative intent stage IV and poor response	0.99	[52]

*SBRT; stereotactic body radiation ** RT=radiation therapy

Table A 8: Transition probabilities related to movement through the time-dependent/state-transition NSCLC model

Transition	Probability	Source
A1		
Stage I		
Monthly probability of distant recurrence after disease free state	0.00297	[44]
Monthly probability of local recurrence after disease free state	0.00198	[44]
Monthly probability of death given disease free state	0.00262	[44]
Monthly probability of staying in disease free	0.99243	[44]
Probability of remission given local recurrence	0.77637	[61]
Probability of palliative care given local recurrence	0.17313	[61]
Probability of death given local recurrence	0.05050	[61]
Probability of remission given distant recurrence	0.46258	[61]
Probability of palliative care given distant recurrence	0.39457	[61]
Probability of death given distant recurrence	0.14285	[61]
Monthly probability of staying in remission given recurrence (both local and distant)	0.98649	[48]
Monthly probability of palliative care given remission	0.01089	[48]
Monthly probability of death given remission	0.00262	[48]
Stage II		
Monthly probability of distant recurrence after disease free state	0.00445	[44]
Monthly probability of local recurrence after disease free state	0.00148	[44]
Monthly probability of death given disease free state	0.00262	[44]
Monthly probability of staying in disease free	0.99145	[44]
Probability of remission given local recurrence	0.66542	[61]
Probability of palliative care given local recurrence	0.18598	[61]
Probability of death given local recurrence	0.14860	[61]
Probability of remission given distant recurrence	0.46258	[61]
Probability of palliative care given distant recurrence	0.39457	[61]
Probability of death given distant recurrence	0.14285	[61]
Monthly probability of staying in remission given local recurrence (both local and distant)	0.98169	[48]
Monthly probability of palliative care given remission	0.01307	[48]
Monthly probability of death given remission	0.00524	[48]

Continuing Table A 9: Transition probabilities related to movement in the time-dependent/state-transition NSCLC model

Stage III		
Monthly probability of distant recurrence after disease free state	0.03344	[48]
Monthly probability of local recurrence after disease free state	0.00991	[48]
Monthly probability of death given disease free state	0.00580	[48]
Monthly probability of staying in disease free	0.95085	[48]
Probability of remission given local recurrence	0.66542	[61]
Probability of palliative care given local recurrence	0.29743	[61]
Probability of death given local recurrence	0.03715	[61]
Probability of remission given distant recurrence	0.46542	[61]
Probability of palliative care given distant recurrence	0.29743	[61]
Probability of death given distant recurrence	0.03715	[61]
Monthly probability of staying in remission given local recurrence (both local and distant)	0.98311	[60]
Monthly probability of palliative care given remission	0.01361	[60]
Monthly probability of death given remission	0.00327	[60]
A2		
Stage IV		
Monthly probability of staying in remission	0.70009	[52]
Monthly probability of recurrence after remission	0.10056	[52]
Monthly probability of death given remission	0.19935	[52]
Probability of palliative care given recurrence	0.87379	[51]
Probability of death given recurrence	0.12621	[51]
B		
Monthly probability of death while in palliative care	0.30064	[51]

Table A 9: Identification of costs associated with diagnostics and treatment for all stages of Non-Small cell lung cancer (NSCLC). Separated by diagnostic, planning, treatment and other relevant direct hospital costs. All costs given in NOK as unit cost per utilization.

NSCLC		
Cost parameters	**Unit cost**	**Source**
Diagnostics		
Outpatient consultation	2 258[1]	DRG*
Spirometry	2 345	DRG*
PET scan	341	[7, 72]
Bronchoscopy	5511	DRG*
EBUS	15 015	[7, 72]
Chest contrast CT	3 949	[7, 72]
CT biopsy	43 208	[7, 72]
Pre-treatment (planning)		
Outpatient consultation – simulation radiation therapy	11 465[4]	DRG*
Outpatient consultation – ordinary radiation therapy	13 767[4]	DRG*
Outpatient consultation – complex radiation therapy	14 288[4]	DRG*
Outpatient consultation – Surgical preparation	2 736[4]	DRG*
Outpatient consultation – chemotherapy preparation	7 385[4]	[7, 72]
Admission before treatment	7 860	DRG*
Surgery		
Lobectomy – Scheduled	56 243	[7, 72]
Wedge resection – Scheduled	57 365	[7, 72]
Lobectomy – Emergency	61 655	[7, 72]
Wedge resection – Emergency	61 986	[7, 72]
Complicated thoracic procedure	172 583	DRG*
Endobronchial stent	14 102	[7, 72]
Endobronchial debulking	13 982	[7, 72]
Intracranial surgery	166 850	DRG*
Radiation therapy[2]		
SBRT	44 861	DRG*
Curative radiation therapy (+ adjuvant and neo-adjuvant)	25 124	[7, 72]
SBRT (brain metastasis)	61 956	DRG*
Palliative radiation therapy	21 340	[7, 72]
Chemotherapy drugs[2]		[7, 72]
In-hospital chemotherapy	6 340[1]	DRG*
Cisplatin + Vinorelbine	49 247[3]	[7, 72]
Docetaxel monotherapy (Taxotere)	73 854[3]	[7, 72]
Pemetrexed (Altima)	132 836[3]	[7, 72]
Vinorelbine (Navalbine)	35 673[3]	[7, 72]
Erlotinib (Tarceva)	134 200[3]	[7, 72]
Gefitinib (Iressa)	112 487[3]	[7, 72]
Other		[7, 72]
Surgical lung complication	53 577	DRG*
Outpatient contact in specialized palliative department	4 299[1]	DRG*

[1]Per consulation/meeting. [2]Per round/cycle. [3]Assumes full compliance with regime. EBUS= Endobronchial ultrasound with biopsy. SBRT = stereotactic body radiation. *Source: *Innsatsstyrt finansiering 2018*

9.4 Analysis and methods

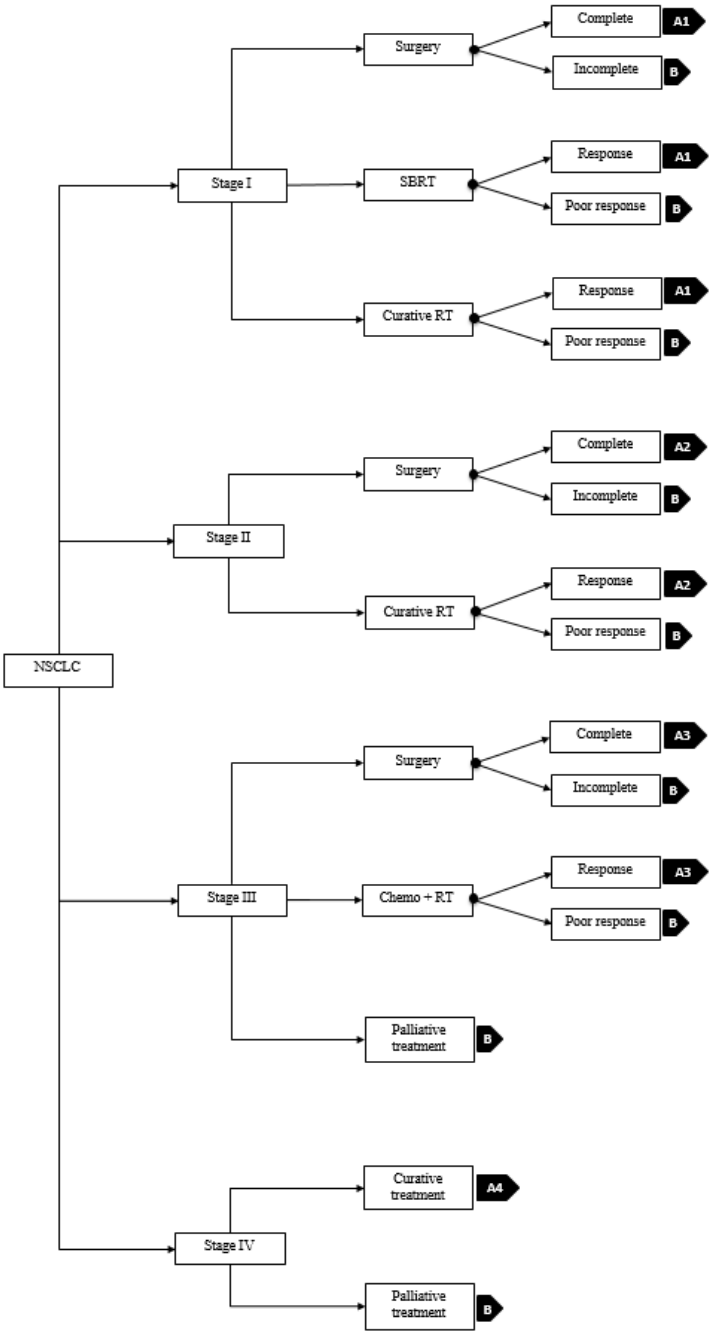


Figure A 1: The decision tree structure representing the "stage-specific" options according to stage of diagnosis for the NSCLC model. Separated by stage, main treatment and treatment response. Arrows represents which Markov model the patient cohort is direct to after 1 month (duration of the decision tree)

9.5 Results

Table A 10: Distribution of diagnostic stage of lung cancer according to county of residence in the diagnostic period of 2000-2016. Percentage (%) in parenthesis.

County	Local disease	Regional disease	Distant disease	Unknown	Total
Østfold	512 (17.6)	813 (27.9)	1 302 (44.7)	285 (9.8)	2 912
Akershus	789 (18.2)	1 143 (26.4)	2 031 (46.9)	373 (8.6)	4 336
Oslo	742 (17.4)	1 086 (25.4)	1 887 (44.2)	557 (13)	4272
Hedmark	267 (12.9)	508 (24.5)	916 (44.2)	381 (18.4)	2 072
Oppland	254 (14.2)	461 (25.7)	799 (44.6)	278 (15.5)	1 792
Buskerud	375 (15.7)	634 (26.6)	1 151 (48.3)	223 (9.4)	2 383
Vestfold	393 (15.2)	751 (29.1)	1 173 (45.4)	266 (10.3)	2 583
Telemark	261 (15.1)	475 (27.5)	693 (40.2)	297 (17.2)	1 726
Aust-Agder	197 (15.6)	337 (26.7)	581 (46)	148 (11.7)	1 263
Vest-Agder	320 (16.8)	559 (29.3)	885 (46.3)	147 (7.8)	1 911
Rogaland	616 (17.5)	882 (25)	1 673 (47.4)	359 (10.2)	3 530
Hordaland	645 (15.8)	1 170 (28.7)	1 819 (44.6)	449 (11)	4 083
Sogn og Fjordane	113 (11.5)	301 (30.8)	473 (48.4)	91 (9.3)	978
Møre og Romsdal	424 (18.1)	634 (27)	1 026 (43.7)	263 (11.2)	2347
Sør-Trøndelag	430 (16.8)	772 (30.1)	1 150 (44.9)	212 (8.3)	2 567
Nord-Trøndelag	184 (14.8)	387 (31.1)	587 (47.1)	88 (7.1)	1 246
Nordland	705 (17.4)	705 (27.8)	1 050 (41.4)	341 (13.5)	2 536
Troms	286 (19.3)	398 (26.9)	621 (42)	175 (11.8)	1 480
Finnmark	153 (17.6)	239 (27.5)	466 (42.1)	112 (12.9)	870
Unknown*	25 (6.4)	55 (14.1)	110 (28.3)	199 (51.2)	389
Norway	**7 426 (16.4)**	**12 310 (27.2)**	**20 293 (44.8)**	**5 244 (11.6)**	**45 273**

* Refugees and their families which are not registered citizens at a specific municipality, and individuals which are registered citizens but without registered address

Table A 11: Distribution of diagnostic stage of colorectal cancer according to county of residence. Percentage (%) in parenthesis.

County	Local disease	Regional disease	Distant disease	Unknown	Total
Østfold	847 (20.8)	2 084 (49.5)	926 (22)	326 (7.7)	4 210
Akershus	1 198 (18.7)	3 339 (52)	1 471 (22.9)	413 (6.4)	6 421
Oslo	1 150 (18.7)	2 986 (48.6)	1 387 (22.6)	626 (10.2)	6 149
Hedmark	531 (18.4)	1 470 (50.9)	655 (22.7)	234 (8.1)	2 890
Oppland	510 (18.7)	1 358 (49.9)	680 (25)	173 (6.4)	2 721
Buskerud	659 (17.9)	1 908 (51.9)	847 (23)	265 (7.2)	3 679
Vestfold	622 (18.2)	1 759 (51.5)	774 (22.6)	263 (7.7)	3 418
Telemark	574 (25.9)	967 (43.6)	494 (22.3)	181 (8.2)	2 216
Aust-Agder	229 (16.6)	711 (51.6)	308 (22.3)	131 (9.5)	1 379
Vest-Agder	453 (21.1)	1 062 (49.4)	491 (22.8)	145 (6.7)	2 151
Rogaland	1 087 (20)	2 588 (47.6)	1 240 (22.8)	519 (9.6)	5 434
Hordaland	1 399 (19.9)	3 498 (49.8)	1 454 (20.7)	673 (9.6)	7 024
Sogn og Fjordane	376 (20.6)	930 (50.9)	380 (20.8)	141 (7.7)	1 827
Møre og Romsdal	889 (21.5)	1 994 (48.2)	906 (21.9)	349 (8.4)	4 138
Sør-Trøndelag	679 (17.9)	1 950 (51.4)	873 (23)	295 (7.8)	3 797
Nord-Trøndelag	349 (18)	997 (51.3)	444 (22.9)	153 (7.9)	1 943
Nordland	793 (21.7)	1 755 (48)	803 (22)	305 (8.3)	3 656
Troms	427 (21)	992 (48.9)	449 (22.1)	161 (7.9)	2 029
Finnmark	173 (23.4)	349 (47.2)	168 (22.7)	49 (6.6)	739
Unknown*	28 (12.7)	84 (38)	76 (34.4)	33 (14.9)	221
Total	**13 000 (19.7)**	**32 781 (49.6)**	**14 826 (22.5)**	**5 435 (8.2)**	**66 042**

* Refugees and their families which are not registered citizens at a specific municipality, and individuals which are registered citizens but without registered address

Table A 12: One-, 3- and 5-year absolute survival for patients diagnosed with lung cancer in the diagnostic period 2000-2016 by county CI= confidence interval.

County	1-year absolute survival (95% CI)	3-year absolute survival (95% CI)	5-year absolute survival (95% CI)
Østfold	34.13 (32.39-35.88)	16.42 (15.03-17.86)	10.95 (9.73-12.26)
Akershus	39.50 (38.02-40.97)	19.54 (18.32-20.78)	14.32 (13.21-15.48)
Oslo	37.80 (36.32-39.27)	17.77 (16.58-18.99)	13.36 (12.27-14.49)
Hedmark	38.12 (36.00-40.23)	15.89 (14.26-17.59)	11.06 (9.61-12.63)
Oppland	*34.12 (31.9-36.35)*	14.81 (13.12-16.6)	10.32 (8.83.11.93)
Buskerud	36.79 (34.83-38.75)	16.71 (15.17-18.32)	12.04 (10.65-13.53)
Vestfold	37.37 (35.48-39.25)	16.88 (15.39-18.42)	12.29 (10.93-13.73)
Telemark	37.92 (35.60-40.24)	16.77 (14.96-18.68)	11.71 (10.10-13.44)
Aust-Agder	39.32 (36.59-42.04)	18.50 (16.31-20.81)	12.84 (10.88-14.97)
Vest-Agder	**43.59 (41.32-45.84)**	**20.33 (18.45-22.27)**	**15.39 (13.66-17.22)**
Rogaland	38.51 (36.87-40.14)	16.68 (15.39-18.01)	11.89 (10.73-13.12)
Hordaland	41.02 (39.48-42.55)	18.10 (16.86-19.37)	13.49 (12.35-14.67)
Sogn og Fjordane	38.89 (35.78-41.98)	*14.77 (12.50-17.21)*	*9.67 (7.74-11.85)*
Møre og Romsdal	42.29 (40.25-44.31)	19.76 (18.09-21.49)	14.67 (13.14-16.28)
Sør-Trøndelag	39.37 (37.46-41.28)	19.10 (17.54-20.71)	14.67 (13.23-16.19)
Nord-Trøndelag	35.59 (32.90-38.29)	16.96 (14.84-19.20)	12.70 (10.75-14.81)
Nordland	39.42 (37.49-41.35)	16.63 (15.12-18.19)	12.30 (10.94-13.74)
Troms	38.93 (36.41-41.44)	17.75 (15.55-19.63)	12.48 (10.70-14.40)
Finnmark	40.35 (37.03-43.64)	16.65 (14.15-19.34)	11.21 (9.03-13.63)
Unknown*	45.58 (40.25-50.75)	20.13 (15.46-25.26)	10.27 (6.49-15.04)

* Refugees and their families which are not registered citizens at a specific municipality, and individuals which are registered citizens but without registered address.

Table A 13: One-, 3- and 5-year absolute survival for patients diagnosed with colorectal cancer in the diagnostic period 2000-2016 by county CI= confidence interval.

County	1-year absolute survival (95% CI)	3-year absolute survival (95% CI)	5-year absolute survival (95% CI)
Østfold	76.29 (74.96-77.56)	58.33 (56.75-59.87)	47.42 (45.76-49.06)
Akershus	77.78 (76.72-78.79)	60.94 (59.67-62.19)	51.74 (50.38-53.08)
Oslo	74.51 (73.39-75.59)	*55.90 (54.60-57.19)*	46.47 (45.12-47.81)
Hedmark	*74.41 (72.76-75.98)*	56.86 (54.95-58.72)	47.44 (45.47-49.39)
Oppland	74.75 (73.05-76.36)	57.08 (55.12-59.00)	48.06 (46.02-50.07)
Buskerud	75.66 (74.22-77.03)	57.87 (56.18-59.52)	47.70 (45.93-49.5)
Vestfold	74.73 (73.22-76.17)	57.04 (55.28-58.76)	48.35 (46.53-50.16)
Telemark	75.68 (73.81-77.43)	57.40 (55.21-59.53)	*46.08 (43.79-48.34)*
Aust-Agder	76.43 (74.06-78.60)	60.83 (58.09-63.45)	51.87 (48.97-54.68)
Vest-Agder	76.67 (74.79-78.42)	59.15 (56.94-61.30)	50.25 (47.92-52.54)
Rogaland	77.01 (75.85-78.12)	58.96 (57.57-60.31)	47.85 (46.40-49.30)
Hordaland	76.86 (75.85-77.85)	59.93 (58.72-61.12)	50.44 (49.16-51.70)
Sogn og Fjordane	78.45 (76.46-80.29)	60.29 (57.90-62.59)	50.01 (47.49-52.48)
Møre og Romsdal	78.51 (77.20-79.75)	58.94 (57.34-60.50)	49.30 (47.63-50.96)
Sør-Trøndelag	77.22 (75.84-78.54)	59.23 (57.58-60.85)	50.52 (48.78-52.23)
Nord-Trøndelag	77.73 (75.79-79.54)	58.82 (56.49-61.08)	49.51 (47.08-51.90)
Nordland	77.25 (75.83-78.59)	59.96 (58.27-61.60)	50.73 (48.96-52.48)
Troms	77.11 (75.19-78.90)	58.96 (56.68-61.16)	48.62 (46.23-50.97)
Finnmark	**79.75 (76.62-82.50)**	**62.61 (58.84-66.14)**	**52.28 (48.28-56.11)**
Unknown*	72.39 (65.78-77.93)	49.25 (41.95-56.14)	36.85 (29.78-43.92)

* Refugees and their families which are not registered citizens at a specific municipality, and individuals which are registered citizens but without registered address

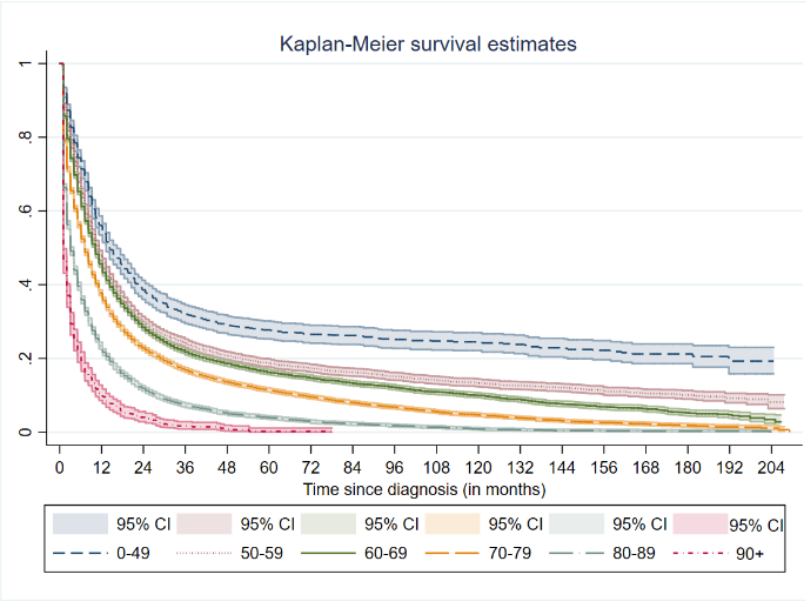


Figure A 2: Kaplan-Meier survival curve for lung cancer since time of diagnosis by age-groups. CI=confidence interval.

χ^2 = 33551, 5 d.f, p<0.001

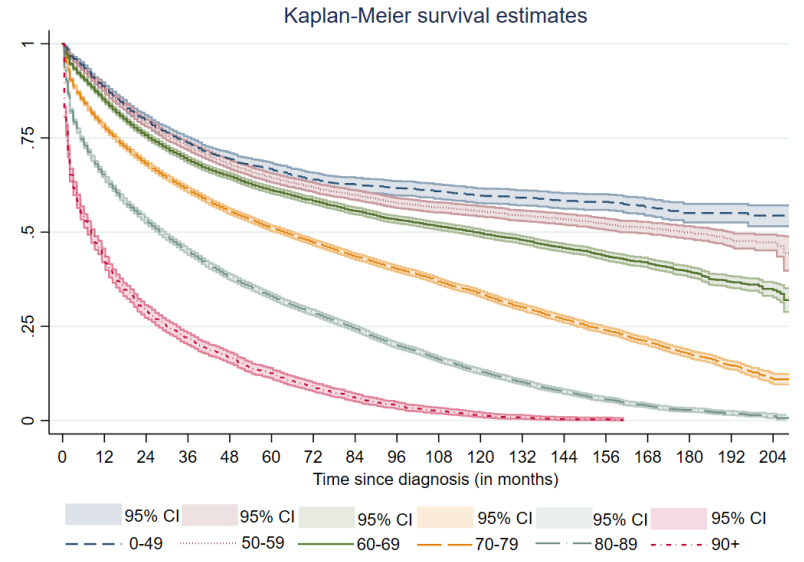


Figure A 3: Kaplan-Meier survival curve for lung cancer since time of diagnosis by age-group. CI=confidence interval.

χ^2 = 33551, 5 d.f, p<0.001

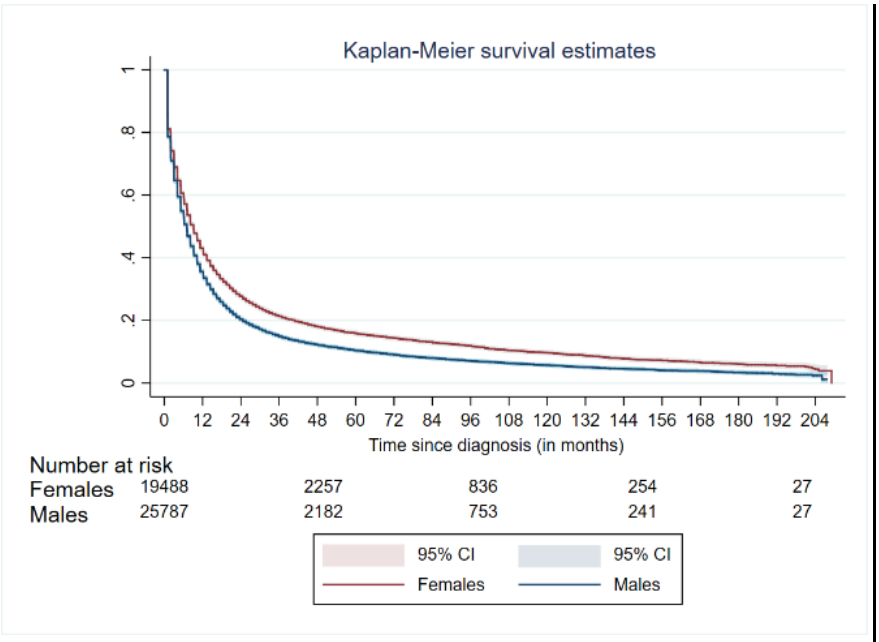


Figure A 4: Kaplan-Meier survival curve for lung cancer since time of diagnosis by gender. CI=confidence interval. χ^2 = 370, 5 d.f, p<0.001

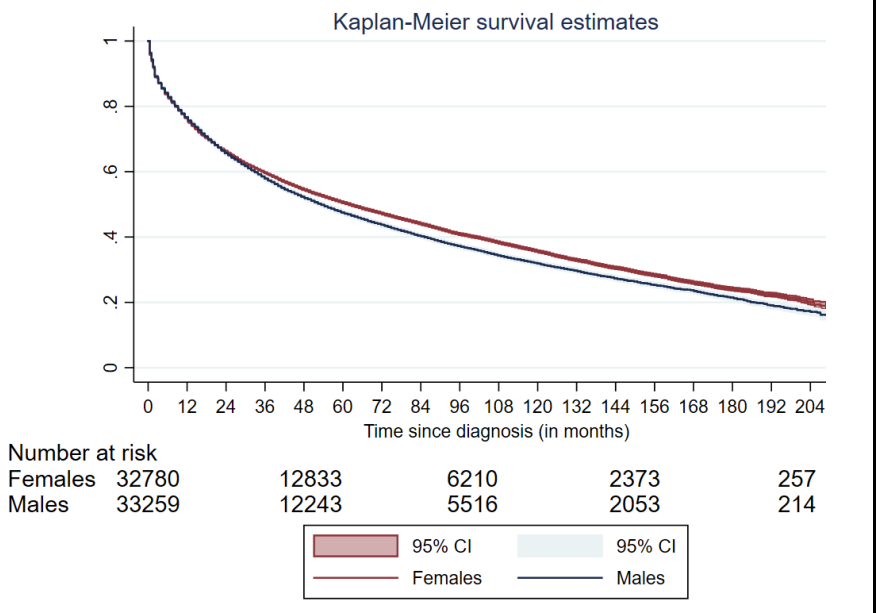


Figure A 5: Kaplan-Meier survival curve for colorectal cancer since time of diagnosis by gender. CI=confidence interval. χ^2 = 44.4, 1 d.f, p<0.001

Table A 14: Difference in survival for patients diagnosed with lung cancer from 2000-2016, by county. CI= confidence interval. Kaplan-Meier estimation and log-rank test for equality of the survivor function.

County	Number of deaths (O_i)	Median survival time (months) (95% CI)	Expected number of deaths (E_i)
Østfold	2 524	6.19 (2.05-20.45)	2 329.3
Akershus	3 657	8.15 (2.05-24.62)	3 781.9
Oslo	3 669	7.17 (2.05-22.53)	3 605.1
Hedmark	1 797	7.14 (1.99-21.49)	1 699.1
Oppland	*1 584*	*6.19 (2.05-19.47)*	*1 429.1*
Buskerud	2 071	7.14 (2.05-19.47)	1 966.5
Vestfold	2 237	7.17 (2.05-20.48)	2 163.9
Telemark	1 500	7.21 (2.05-20.5)	1 443.2
Aust-Agder	1 074	8.22 (2.05-23.6)	1 102.6
Vest-Agder	**1 573**	**9.23 (3.07-25.67)**	**1 773.9**
Rogaland	3 010	8.19 (2.05-19.44)	2 965.8
Hordaland	3 442	8.25 (2.08-23.58)	3 608.4
Sogn og Fjordane	859	8.19 (3.07-25.63)	820.2
Møre og Romsdal	1 973	9.19 (3.07-25.63)	2 140.5
Sør-Trøndelag	2 161	7.24 (2.05-23.58)	2 263.6
Nord-Trøndelag	1 076	7.17 (2.05-20.51)	1 045.3
Nordland	2 183	8.19 (2.99-21.49)	2 200.2
Troms	1 277	8.19 (2.99-22.60)	1 290.9
Finnmark	757	8.22 (3.03-21.59)	766.4
Unknown*	280	10.24 (2.09-28.69)	308.3

* Refugees and their families which are not registered citizens at a specific municipality, and individuals which are registered citizens but without registered address. $\chi^2 = 114.6$, 19 d.f, $p<0.001$

Table A 15: Difference in survival for patients diagnosed with colorectal cancer from 2000-2016, by county. CI= confidence interval. Kaplan-Meier estimation and log-rank test for equality of the survivor function.

County	Number of deaths (O_i)	Median survival time (months) (95% CI)	Expected number of deaths (E_i)
Østfold	2 438	55.31 (13.31-154.81)	2 346.6
Akershus	**3 363**	**66.59 (15.33-189.64)**	**3 678.6**
Oslo	*3 660*	*48.17 (11.28-162.96)*	*3 390.5*
Hedmark	1 751	52.31 (11.28-144.57)	1 614.9
Oppland	1 603	54.29 (11.32-159.93)	1 538.9
Buskerud	2 122	52.31 (12.29-160.98)	2 046.8
Vestfold	1 943	54.36 (11.28-168.12)	1 908.6
Telemark	1 304	52.24 (12.21-154.78)	1 233.7
Aust-Agder	757	65.65 (13.34-162.99)	796.8
Vest-Agder	1 215	60.49 (14.35-155.89)	1 226.2
Rogaland	3 075	54.29 (13.34-168.12)	3 069.1
Hordaland	3 906	61.51 (14.28-174.32)	4 052.7
Sogn og Fjordane	1 027	60.46 (15.29-177.38)	1 067.6
Møre og Romsdal	2 324	58.41 (15.36-174.32)	2 375.3
Sør-Trøndelag	2 148	60.56 (13.34-163.07)	2 179.2
Nord-Trøndelag	1 110	58.41 (14.35-159.93)	1 111.8
Nordland	2 022	62.55 (14.28-178.39)	2 121.7
Troms	1 122	57.36 (13.37-175.36)	1 135.6
Finnmark	418	62.59 (17.45-167.07)	439.9
Unknown*	138	35.84 (10.24-139.45)	111.4

* Refugees and their families which are not registered citizens at a specific municipality, and individuals which are registered citizens but without registered address. $\chi^2 = 97.13$, 19 d.f, $p<0.001$

Table A 16: Regression output from the Cox proportional hazard model for the lung cancer dataset. Presented with coefficients, hazard ratios, 96 % confidence intervals and p-value. The univariate regression on the left-hand side, and the multivariate regression on the right-hand side.

	Univariate analysis				Multivariate analysis			
Covariate	**Coef. (b_i)**	**HR [exp(b_i)]**	**95%CI**	**p-value**	**Coef. (b_i)**	**HR [exp(b_i)]**	**95%CI**	**p-value**
Gender								
Female*								
Male	0.198	1.219	1.19-1.25	<0.001	0.171	1.187	1.16-1.21	<0.001
Stage								
Localized disease*								
Regional disease	0.982	2.669	2.56-2.79	<0.001	0.815	2.258	2.17-2.34	<0.001
Metastatic disease	1.844	6.321	6.07-6.58	<0.001	1.672	5.323	5.14-5.51	<0.001
Unknown	1.119	3.063	2.91-3.23	<0.001	1.159	3.186	3.05-3.32	<0.001
Age (years)								
0-49*								
50-59	0.253	1.286	1.19-1.38	<0.001	0.248	1.281	1.19-1.37	<0.001
60-69	0.295	1.343	1.25-1.44	<0.001	0.402	1.496	1.40-1.59	<0.001
70-79	0.454	1.575	1.47-1.69	<0.001	0.667	1.949	1.83-2.08	<0.001
80-89	0.792	2.208	2.06-2.37	<0.001	1.093	2.985	2.79-3.19	<0.001
90 and over	1.167	3.212	2.87-3.59	<0.001	1.531	4.624	4.22-5.07	<0.001
County								
Østfold*								
Akershus	-0.114	0.892	0.84-0.94	<0.001	-0.140	0.869	0.83-0.91	<0.001
Oslo	-0.063	0.939	0.89-0.99	0.015	-0.097	0.907	0.86-0.95	<0.001
Hedmark	-0.024	0.976	0.92-1.04	0.433	-0.099	0.904	0.85-0.96	0.001
Oppland	0.023	1.023	0.96-1.09	0.480	-0.052	0.949	0.89-1.01	0.103
Buskerud	-0.029	0.972	0.92-1.03	0.336	-0.092	0.911	0.86-0.97	0.002
Vestfold	-0.047	0.954	0.90-1.01	0.105	-0.096	0.908	0.86-0.96	0.001
Telemark	-0.042	0.959	0.90-1.02	0.200	-0.086	0.917	0.86-0.98	0.008
Aust-Agder	-0.107	0.899	0.84-0.97	0.003	-0.145	0.864	0.80-0.93	<0.001
Vest-Agder	-0.201	0.819	0.77-0.87	<0.001	-0.245	0.783	0.73-0.83	<0.001
Rogaland	-0.066	0.937	0.89-0.99	0.015	-0.124	0.884	0.84-0.93	<0.001
Hordaland	-0.128	0.88	0.84-0.93	<0.001	-0.214	0.807	0.77-0.85	<0.001
Sogn og Fjordane	-0.034	0.967	0.89-1.04	0.390	-0.153	0.858	0.79-0.93	<0.001
Møre og Romsdal	-0.162	0.851	0.80-0.90	<0.001	-0.232	0.793	0.75-0.84	<0.001
Sør-Trøndelag	-0.129	0.881	0.83-0.93	<0.001	-0.162	0.850	0.80-0.90	<0.001
Nord-Trøndelag	-0.051	0.949	0.88-1.02	0.158	-0.121	0.885	0.82-0.95	0.001
Nordland	-0.088	0.916	0.86-0.97	0.003	-0.119	0.887	0.84-0.94	<0.001
Troms	-0.091	0.913	0.85-0.97	0.008	-0.101	0.904	0.84-0.97	0.003
Finnmark	-0.093	0.912	0.84-0.99	0.025	-0.134	0.875	0.81-0.95	0.001
Unknown	-0.176	0.838	0.74-0.94	0.005	-0.308	0.735	0.65-0.83	<0.001

*Reference group

Table A 17: Regression output from the Weibull proportional hazard model for the lung cancer dataset. Presented with coefficients, hazard ratios, 96 % confidence intervals and p-value. The univariate regression on the left-hand side, and the multivariate regression on the right-hand side.

	Univariate analysis				Multivariate analysis			
Covariate	**Coef. (b_i)**	**HR [exp(b_i)]**	**95%CI**	**p-value**	**Coef. (b_i)**	**HR [exp(b_i)]**	**95%CI**	**p-value**
Gender								
Female*								
Male	0.222	1.249	1.22-1.27	<0.001	0.192	1.211	1.18-1.24	<0.001
Stage								
Localized disease*								
Regional disease	0.877	2.404	2.32-2.49	<0.001	0.919	2.507	2.42-2.59	<0.001
Metastatic disease	1.834	6.259	6.05-6.48	<0.001	1.895	6.650	6.43-6.88	<0.001
Unknown	1.321	3.747	3.59-3.91	<0.001	1.258	3.519	3.37-3.67	<0.001
Age (years)								
0-49*								
50-59	0.338	1.402	1.31-1.49	<0.001	0.294	1.336	1.25-1.43	<0.001
60-69	0.476	1.609	1.51-1.72	<0.001	0.479	1.616	1.52-1.72	<0.001
70-79	0.723	2.060	1.93-2.19	<0.001	0.778	2.176	2.04-2.32	<0.001
80-89	1.189	3.286	3.08-3.51	<0.001	1.259	3.523	3.29-3.76	<0.001
90 and over	1.719	5.578	5.09-6.11	<0.001	1.731	5.644	5.15-6.19	<0.001
County								
Østfold*								
Akershus	-0.141	0.867	0.82-0.91	<0.001	-0.163	0.849	0.81-0.89	<0.001
Oslo	-0.079	0.923	0.88-0.97	0.002	-0.114	0.893	0.85-0.94	<0.001
Hedmark	-0.021	0.979	0.92-1.04	0.499	-0.102	0.903	0.85-0.96	0.001
Oppland	0.025	1.025	0.96-1.09	0.435	-0.056	0.945	0.89-1.01	0.080
Buskerud	-0.036	965	0.91-1.02	0.230	-0.107	0.898	0.85-0.95	<0.001
Vestfold	-0.049	0.952	0.89-1.01	0.091	-0.115	0.891	0.85-0.94	<0.001
Telemark	-0.048	0.954	0.89-1.01	0.145	-0.096	0.909	0.85-0.97	0.003
Aust-Agder	-0.111	0.895	0.83-0.96	0.002	-0.156	0.856	0.79-0.92	<0.001
Vest-Agder	-0.224	0.799	0.75-0.85	<0.001	-0.268	0.765	0.72-0.81	<0.001
Rogaland	-0.056	0.946	0.89-0.99	0.039	-0.126	0.882	0.84-0.93	<0.001
Hordaland	-0.140	0.869	0.83-0.92	<0.001	-0.237	0.789	0.75-0.83	<0.001
Sogn og Fjordane	-0.018	0.982	0.91-1.06	0.652	-0.159	0.853	0.79-0.92	<0.001
Møre og Romsdal	-0.179	0.835	0.79-0.89	<0.001	-0.254	0.776	0.73-0.82	<0.001
Sør-Trøndelag	-0.157	0.854	0.81-0.91	<0.001	-0.187	0.829	0.78-0.88	<0.001
Nord-Trøndelag	-0.059	0.942	0.88-1.01	0.100	-0.138	0.871	0.81-0.94	<0.001
Nordland	-0.102	0.903	0.85-0.96	0.001	-0.132	0.876	0.83-0.93	<0.001
Troms	-0.096	0.909	0.85-0.97	0.005	-0.106	0.899	0.84-0.96	0.002
Finnmark	-0.095	0.909	0.84-0.99	0.022	-0.147	0.863	0.79-0.94	<0.001
Unknown	-0.144	0.866	0.76-0.98	0.022	-0.271	0.762	0.67-0.86	<0.001

*Reference group

Table A 18: Regression output from the Exponential proportional hazard model for the lung cancer dataset. Presented with coefficients, hazard ratios, 96 % confidence intervals and p-value. The univariate regression on the left-hand side, and the multivariate regression on the right-hand side.

	Univariate analysis			Multivariate analysis		
Covariate	**HR**	**95%CI**	**p-value**	**HR**	**95%CI**	**p-value**
Gender						
Female*						
Male	1.335	1.31-1.36	<0.001	1.252	1.23-1.28	<0.001
Stage						
Localized disease*						
Regional disease	2.703	2.61-2.80	<0.001	2.768	2.76-2.87	<0.001
Metastatic disease	8.867	8.58-9,16	<0.001	8.738	8.45-9.03	<0.001
Unknown	4.328	4.15-4.51	<0.001	3.944	3.78-4.11	<0.001
Age (years)						
0-49*						
50-59	1.522	1.42-1.63	<0.001	1.396	1.30-1.49	<0.001
60-69	1.879	1.76-2.00	<0.001	1.754	1.65-1.87	<0.001
70-79	2.581	2.42-2.75	<0.001	2.466	2.31-2.63	<0.001
80-89	4.850	4.54-5.18	<0.001	4.361	4.09-4.56	<0.001
90 and over	10.599	9.68-11.61	<0.001	7.733	7.09-8.47	<0.001
County						
Østfold*						
Akershus	0.816	0.78-0.86	<0.001	0.822	0.78-0.87	<0.001
Oslo	0.885	0.84-0.93	<0.001	0.869	0.83-0.19	<0.001
Hedmark	0.981	0.92-1.04	0.536	0.895	0.84-0.95	<0.001
Oppland	1.031	0.97-1.09	0.339	0.939	0.88-1.00	0.048
Buskerud	0.944	0.89-1.00	0.053	0.876	0.83-0.93	<0.001
Vestfold	0.938	0.89-0.99	0.029	0.864	0.82-0.91	<0.001
Telemark	0.934	0.88-0.99	0.038	0.893	0.84-0.95	0.001
Aust-Agder	0.871	0.81-0.93	<0.001	0.838	0.78-0.89	<0.001
Vest-Agder	0.748	0.70-0.79	<0.001	0.737	0.69-0.79	<0.001
Rogaland	0.954	0.90-1.00	0.078	0.871	0.83-0.92	<0.001
Hordaland	0.831	0.78-0.88	<0.001	0.758	0.72-0.79	<0.001
Sogn og Fjordane	1.001	0.93-1.01	0.987	0.839	0.78-0.91	<0.001
Møre og Romsdal	0.792	0.75-0.84	<0.001	0.747	0.70-0.79	<0.001
Sør-Trøndelag	0.794	0.75-0.84	<0.001	0.803	0.76-0.85	<0.001
Nord-Trøndelag	0.920	0.86-0.99	0.022	0.855	0.79-0.92	<0.001
Nordland	0.866	0.82-0.92	<0.001	0.859	0.87-0.91	<0.001
Troms	0.887	0.83-0.95	<0.001	0.889	0.93-0.95	0.001
Finnmark	0.888	0.82-0.96	0.004	0.838	0.77-0.91	<0.001
Unknown	0.915	0.81-1.03	0.158	0.776	0.69-0.88	<0.001

*Reference group

Table A 19: Regression output from the Cox proportional hazard model for the colorectal cancer dataset. Presented with coefficients, hazard ratios, 96 % confidence intervals and p-value. The univariate regression on the left-hand side, and the multivariate regression on the right-hand side.

	Univariate analysis				Multivariate analysis			
Covariate	**Coef. (b_i)**	**HR [exp(b_i)]**	**95%CI**	**p-value**	**Coef. (b_i)**	**HR [exp(b_i)]**	**95%CI**	**p-value**
Gender								
Female*								
Male	0.068	1.071	1.04-1.09	<0.001	0.159	1.172	1.15-1.19	<0.001
Stage								
Localized disease*								
Regional disease	0.343	1.409	1.36-1.45	<0.001	0.349	1.419	1.37-1.47	<0.001
Metastatic disease	1.797	6.031	5.83-6.24	<0.001	1.976	7.21	6.97-7.46	<0.001
Unknown	1.217	3.377	3.24-3.52	<0.001	1.149	3.157	3.03-3.29	<0.001
Age (years)								
0-49*								
50-59	0.112	1.119	1.04-1.19	0.002	0.146	1.157	1.07-1.24	<0.001
60-69	0.313	1.367	1.28-1.46	<0.001	0.396	1.486	1.39-1.58	<0.001
70-79	0.736	2.087	1.96-2.22	<0.001	0.933	2.512	2.39-2.70	<0.001
80-89	1.282	3.602	3.39-3.83	<0.001	1.538	4.654	4.38-4.95	<0.001
90 and over	1.936	6.928	6.46-7.43	<0.001	2.186	8.903	8.29-9.56	<0.001
County								
Østfold*								
Akershus	-0.128	0.879	0.84-0.93	<0.001	-0.110	0.897	0.85-0.94	<0.001
Oslo	0.038	1.039	0.99-1.09	0.143	-0.033	0.968	0.92-1.02	0.213
Hedmark	0.043	1.043	0.98-1.11	0.173	-0.001	0.999	0.94-1.06	0.986
Oppland	0.003	1.003	0.94-1.07	0.936	-0.060	0.942	0.88-1.00	0.061
Buskerud	-0.002	0.998	0.94-1.06	0.943	-0.035	0.966	0.91-1.02	0.244
Vestfold	-0.020	0.979	0.92-1.04	0.503	-0.059	0.943	0.89-1.00	0.054
Telemark	0.017	1.017	0.95-1.09	0.617	-0.029	0.972	0.91-1.04	0.403
Aust-Agder	-0.089	0.914	0.84-0.99	0.032	-0.135	0.874	0.81-0.95	0.001
Vest-Agder	-0.047	0.954	0.89-1.02	0.177	-0.093	0.912	0.85-0.98	0.008
Rogaland	-0.036	0.964	0.91-1.02	0.181	-0.088	0.916	0.87-0.97	0.001
Hordaland	-0.075	0.928	0.88-0.98	0.004	-0.101	0.904	0.86-0.95	<0.001
Sogn og Fjordane	-0.077	0.926	0.86-0.99	0.038	-0.127	0.881	0.82-0.95	0.001
Møre og Romsdal	-0.060	0.942	0.88-0.99	0.038	-0.142	0.868	0.82-0.92	<0.001
Sør-Trøndelag	-0.052	0.949	0.89-1.01	0.075	-0.094	0.911	0.86-0.97	0.002
Nord-Trøndelag	-0.039	0.961	0.89-1.03	0.271	-0.102	0.903	0.84-0.97	0.005
Nordland	-0.086	0.917	0.86-0.97	0.004	-0.082	0.921	0.87.0.98	0.006
Troms	-0.050	0.951	0.89-1.02	0.164	-0.059	0.943	0.88-1.01	0.103
Finnmark	-0.089	0.915	0.82-1.01	0.091	-0.042	0.959	0.86-1.06	0.429
Unknown	0.176	1.192	1.00-1.42	0.045	0.111	1.117	0.94-1.33	0.206

*Reference group

Table A 20: Regression output from the Weibull proportional hazard model for the colorectal cancer dataset. Presented with coefficients, hazard ratios, 96 % confidence intervals and p-value. The univariate regression on the left-hand side, and the multivariate regression on the right-hand side.

	Univariate analysis				Multivariate analysis			
Covariate	**Coef. (b_i)**	**HR [exp(b_i)]**	**95%CI**	**p-value**	**Coef. (b_i)**	**HR [exp(b_i)]**	**95%CI**	**p-value**
Gender								
Female*								
Male	0.070	1.073	1.05-1.09	<0.001	0.156	1.169	1.15-1.19	<0.001
Stage								
Localized disease*								
Regional disease	0.341	1.406	1.36-1.45	<0.001	0.342	1.407	1.36-1.45	<0.001
Metastatic disease	1.826	6.210	6.01-6.42	<0.001	1.998	7.371	7.13-7.63	<0.001
Unknown	1.233	3.431	3.23-3.58	<0.001	1.171	3.225	3.09-3.36	<0.001
Age (years)								
0-49*								
50-59	0.109	1.116	1.04-1.19	0.002	0.142	1.152	1.07-1.24	<0.001
60-69	0.306	1.358	1.27-1.45	<0.001	0.388	1.474	1.38-1.57	<0.001
70-79	0.726	2.067	1.94-2.19	<0.001	0.926	2.526	2.38-2.69	<0.001
80-89	1.271	3.566	3.35-3.79	<0.001	1.529	4.614	4.34-4.91	<0.001
90 and over	1.950	7.031	6.55-7.54	<0.001	2.198	9.009	8.39-9.67	<0.001
County								
Østfold*								
Akershus	-0.128	0.879	0.83-0.93	<0.001	-0.108	0.898	0.85-0.95	<0.001
Oslo	0.039	1.040	0.99-1.09	0.129	-0.031	0.969	0.92-1.02	0.238
Hedmark	0.042	1.043	0.98-1.11	0.182	-0.000	0.999	0.94-1.06	0.997
Oppland	0.002	1.002	0.94-1.07	0.956	-0.059	0.942	0.88-1.00	0.065
Buskerud	-0.001	0.999	0.94-1.06	0.984	-0.034	0.966	0.91-1.02	0.248
Vestfold	-0.021	0.979	0.92-1.04	0.496	-0.055	0.947	0.89-1.00	0.071
Telemark	0.019	1.019	0.95-1.09	0.570	-0.029	0.971	0.91-1.04	0.384
Aust-Agder	-0.091	0.913	0.84-0.99	0.028	-0.134	0.875	0.81-0.95	0.001
Vest-Agder	-0.049	0.952	0.89-1.02	0.160	-0.097	0.908	0.85-0.97	0.006
Rogaland	-0.036	0.964	0.91-1.02	0.180	-0.087	0.917	0.87-0.97	0.001
Hordaland	-0.076	0.927	0.88-0.97	0.003	-0.099	0.905	0.86-0.95	<0.001
Sogn og Fjordane	-0.075	0.928	0.86-0.99	0.043	-0.124	0.883	0.82-0.95	0.001
Møre og Romsdal	-0.061	0.941	0.89-0.99	0.035	-0.145	0.865	0.82-0.92	<0.001
Sør-Trøndelag	-0.053	0.948	0.89-1.00	0.071	-0.094	0.911	0.86-0.97	0.002
Nord-Trøndelag	-0.041	0.960	0.89-1.03	0.262	-0.104	0.901	0.84-0.97	0.004
Nordland	-0.087	0.916	0.86-0.97	0.004	-0.079	0.942	0.87-0.98	0.009
Troms	-0.051	0.951	0.89-1.02	0.160	-0.064	0.938	0.87-1.01	0.078
Finnmark	-0.092	0.912	0.82-1.01	0.082	-0.049	0.952	0.86-1.05	0.078
Unknown	0.18	1.199	1.01-1.42	0.032	0.113	1.120	0.94-1.33	0.349

*Reference group

Table A 21: Regression output from the Exponential proportional hazard model for the colorectal cancer dataset. Presented with coefficients, hazard ratios, 96 % confidence intervals and p-value. The univariate regression on the left-hand side, and the multivariate regression on the right-hand side.

	Univariate analysis			Multivariate analysis		
Covariate	**HR**	**95%CI**	**p-value**	**HR**	**95%CI**	**p-value**
Gender						
Female*						
Male	1.096	1.08-1.12	<0.001	1.195	1.17-1.23	<0.001
Stage						
Localized disease*						
Regional disease	1.458	1.41-1.51	<0.001	1.447	1.40-1.49	<0.001
Metastatic disease	8.195	7.93-8.47	<0.001	9.291	8.99-9.60	<0.001
Unknown	3.553	3.41-3.70	<0.001	3.389	3.25-3.53	<0.001
Age (years)						
0-49*						
50-59	1.114	1.04-1.19	0.002	1.166	1.08-1.25	<0.001
60-69	1.393	1.31-1.48	<0.001	1.526	1.43-1.63	<0.001
70-79	2.194	2.06-2.33	<0.001	2.714	2.55-2.89	<0.001
80-89	4.132	3.89-4.39	<0.001	5.296	4.98-5.63	<0.001
90 and over	9.679	9.03-10.38	<0.001	11.660	10.87-12.51	<0.001
County						
Østfold*						
Akershus	0.868	0.82-0.91	<0.001	0.889	0.84-0.94	<0.001
Oslo	1.035	0.98-1.09	0.185	0.959	0.91-1.01	0.112
Hedmark	1.038	0.98-1.10	0.238	0.994	0.93-1.06	0.845
Oppland	0.989	0.93-1.05	0.753	0.931	0.87-0.99	0.026
Buskerud	0.998	0.94-1.06	0.943	0.965	0.91-1.02	0.227
Vestfold	0.968	0.91-1.03	0.290	0.941	0.89-0.99	0.046
Telemark	1.017	0.95-1.09	0.621	0.970	0.91-1.04	0.377
Aust-Agder	0.897	0.83-0.97	0.009	0.861	0.79-0.93	<0.001
Vest-Agder	0.945	0.88-1.01	0.107	0.898	0.84-0.96	0.002
Rogaland	0.959	0.91-1.01	0.129	0.912	0.87-0.96	0.001
Hordaland	0.911	0.87-0.96	<0.001	0.895	0.85-94	<0.001
Sogn og Fjordane	0.915	0.85-0.98	0.016	0.873	0.81-0.94	<0.001
Møre og Romsdal	0.932	0.88-0.99	0.016	0.857	0.81-0.91	<0.001
Sør-Trøndelag	0.935	0.88-0.99	0.024	0.906	0.85-0.96	0.001
Nord-Trøndelag	0.955	0.89-1.03	0.207	0.899	0.84-0.97	0.003
Nordland	0.894	0.84-0.95	<0.001	0.913	0.86-0.97	0.003
Troms	0.952	0.89-1.02	0.175	0.928	0.86-99	0.038
Finnmark	0.894	0.81-0.99	0.034	0.931	0.84-1.03	0.176
Unknown	1.214	1.02-1.44	0.027	1.093	0.92-1.29	0.311

*Reference group

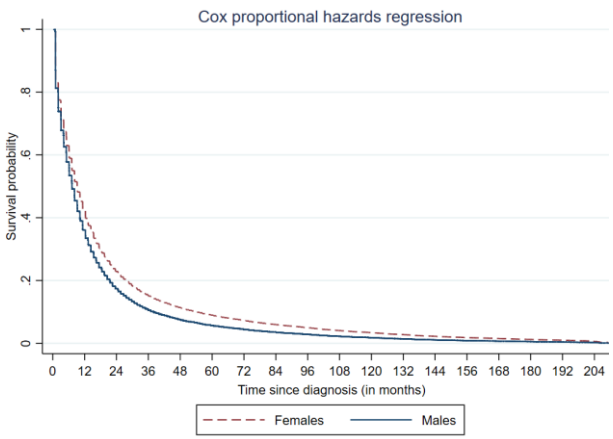
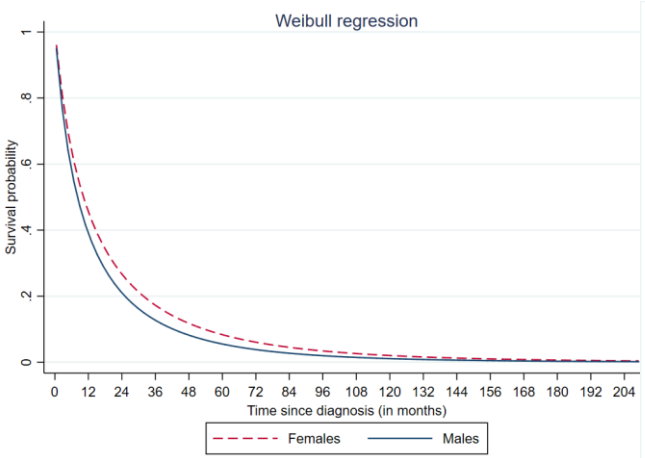
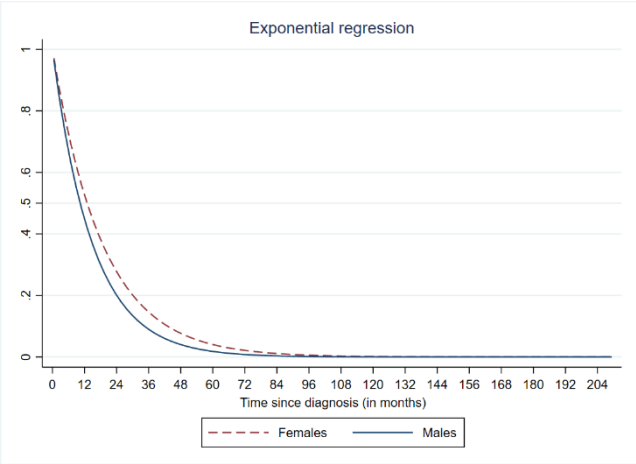


Figure A 7 -1: Semi-Parametric and parametric survival probability – Lung cancer (gender). regression for lung cancer, survival probability according to gender. p<0.001.



A 7 -2: Weibull PH regression for the lung cancer dataset, survival probability according to gender. *p*<0.001.



A 7 -3: Exponential PH regression for the lung cancer dataset, survival probability according to gender. *p*<0.001.

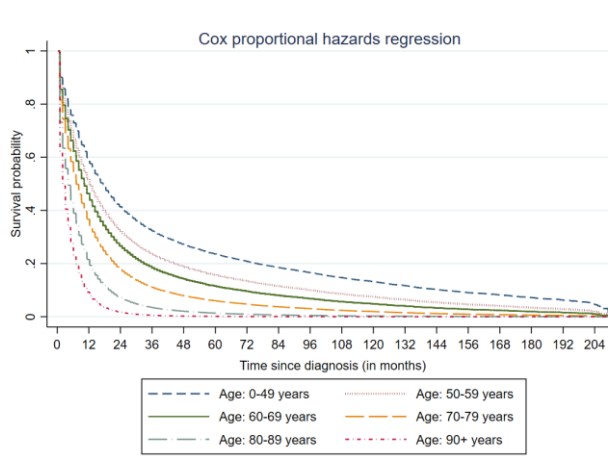
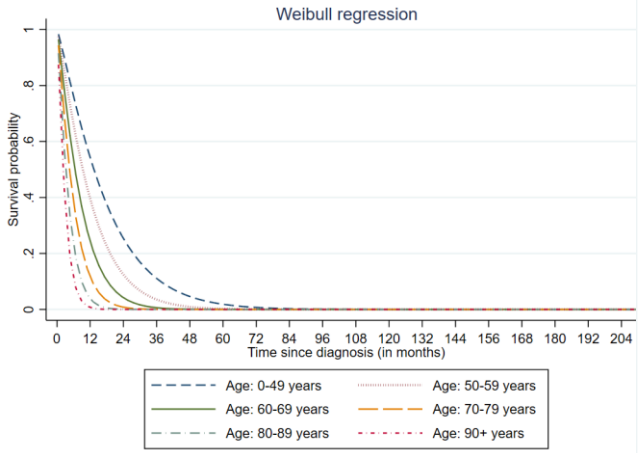
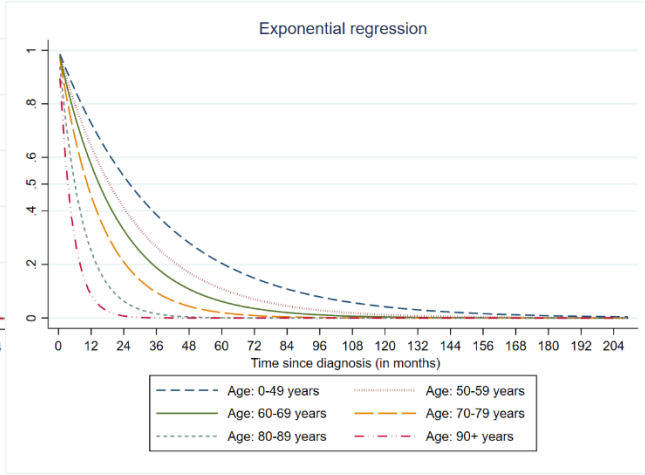


Figure A 6 - 1: Semi-Parametric and parametric survival probability – Lung cancer (age-groups) regression for lung cancer, survival probability according to age groups (10-year age-intervals). Age at time of diagnosis. p<0.001.



A 6 - 2: Weibull PH regression for the lung cancer dataset, survival probability according to age groups (10-year age-intervals). Age at time of diagnosis. *p*<0.001.



A 6 - 3: Exponential PH regression for the lung cancer dataset, survival probability according to age groups (10-year age-intervals). Age at time of diagnosis. *p*<0.001.

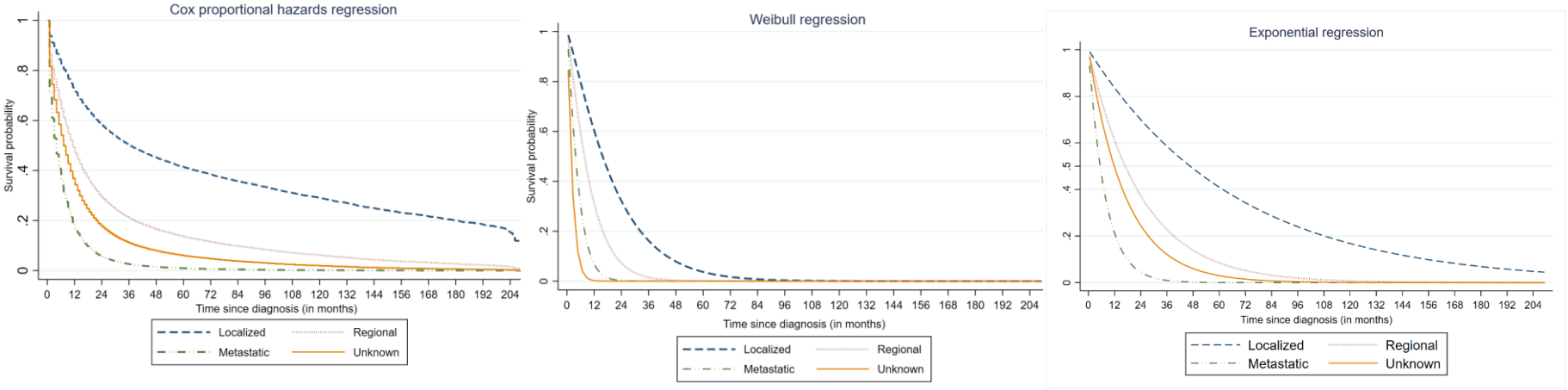


Figure A 9 -1: Semi-Parametric and parametric survival probability – Lung cancer (stage): Cox PH regression for the lung cancer dataset, survival probability according to diagnostic stage at time of diagnosis. p<0.001.

A 9 - 2: Weibull PH regression for the lung cancer dataset, survival probability according to diagnostic stage at time of diagnosis. *p*<0.001.

A 9 - 3: Exponential PH regression for the lung cancer dataset, survival probability according to diagnostic stage at time of diagnosis. *p*<0.001.

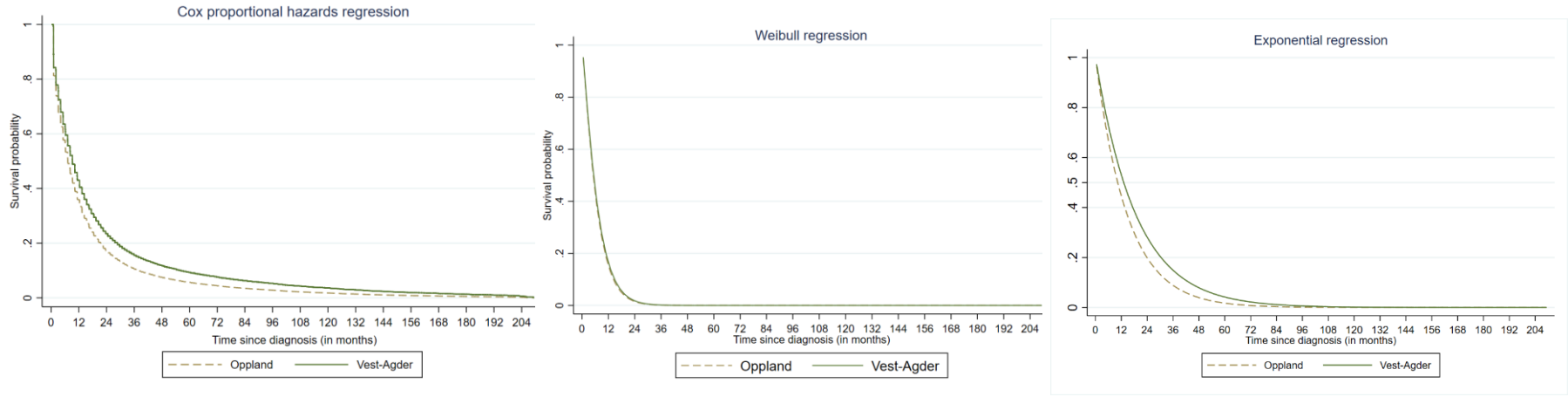
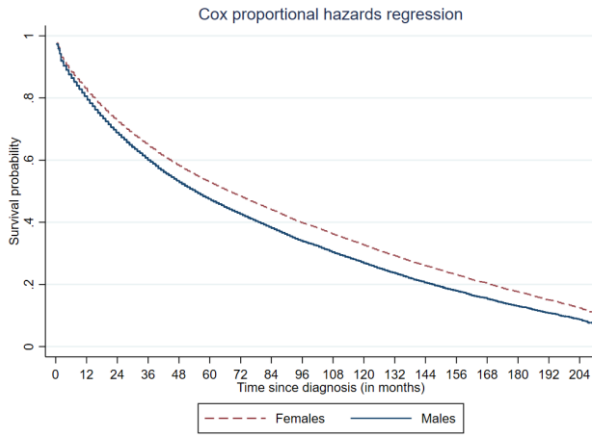
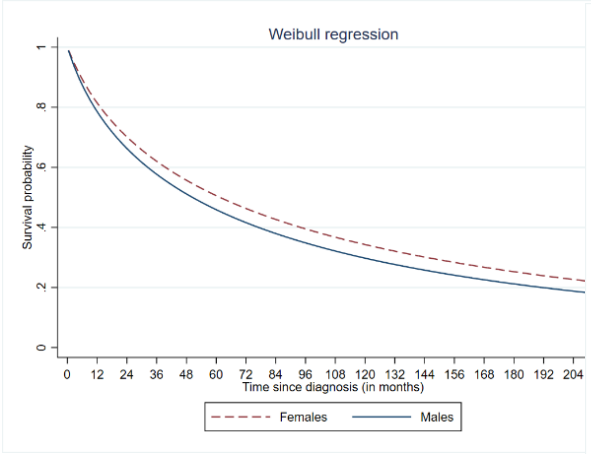


Figure A 8 -1: Semi-Parametric and parametric survival probability – Lung cancer (county) regression for the lung cancer dataset, survival probability according to residence at time of diagnosis. Only depicted the county with the longest and shortest median survival time, as estimated by the Kaplan-Meier estimator. p<0.001.

A 8 - 2: Weibull PH regression for the lung cancer dataset, survival probability according to residence at time of diagnosis. Only depicted the county with the longest and shortest median survival time, as estimated by the Kaplan-Meier estimator. *p*<0.001.

A 9 - 3: Exponential PH regression for the lung cancer dataset, survival probability according to residence at time of diagnosis. Only depicted the county with the longest and shortest median survival time, as estimated by the Kaplan-Meier estimator. *p*<0.001.





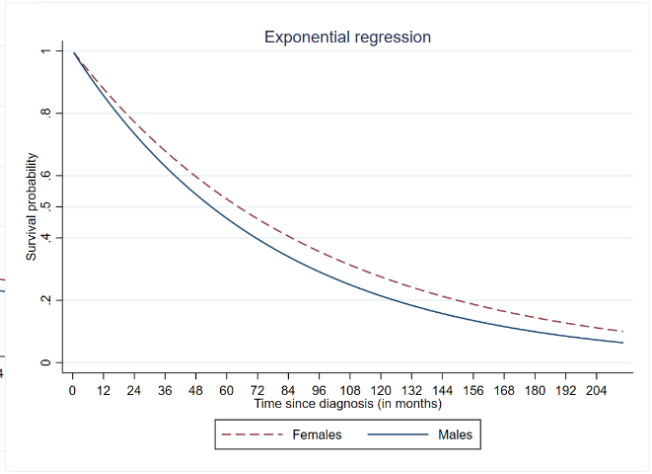
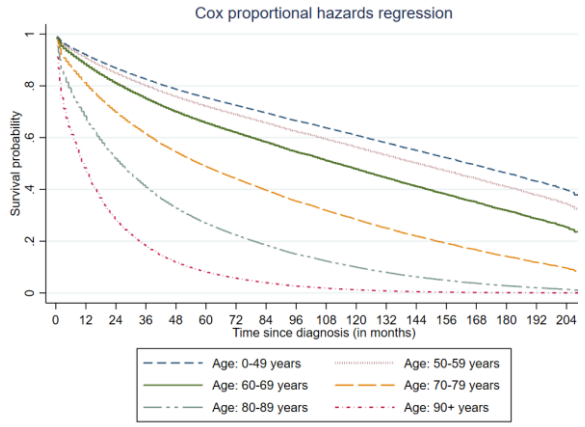
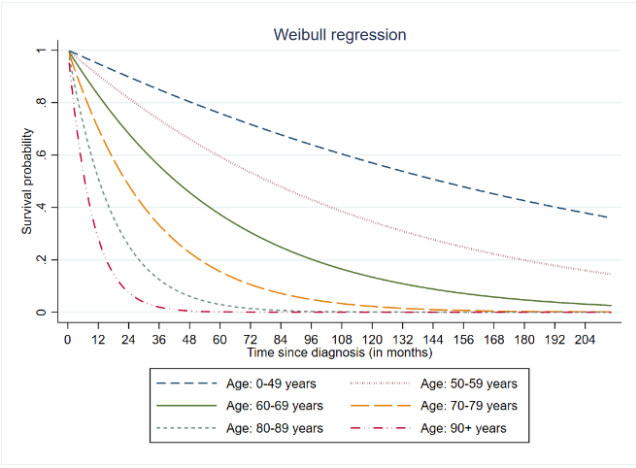


Figure A 11 -1: Semi-Parametric and parametric survival probability – colorectal cancer cancer (gender) regression for the colorectal cancer dataset, survival probability according to gender. p<0.001.

A 11 - 2: Weibull PH regression for the colorectal cancer dataset, survival probability according to gender. *p*<0.001.

A 11 - 3: Exponential PH regression for the colorectal cancer dataset, survival probability according to gender. *p*<0.001.





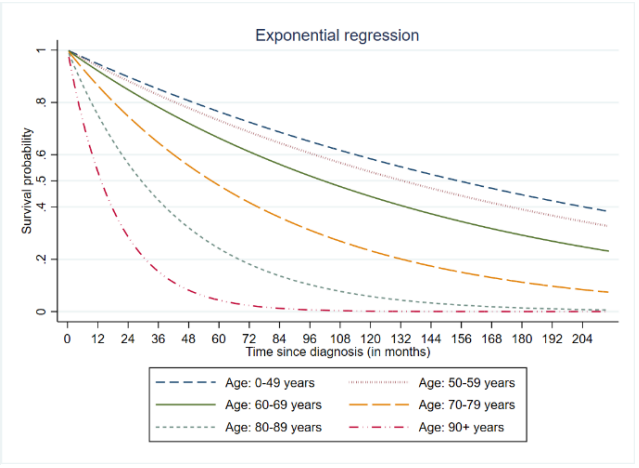


Figure A 10 -1: Semi-Parametric and parametric survival probability – Colorectal cancer (age-groups): regression for the colorectal cancer dataset, survival probability according to age groups (10-year age-intervals). Age at time of diagnosis. p<0.001.

A 10 - 2: Weibull PH regression for the colorectal cancer dataset, survival probability according to age groups (10-year age-intervals). Age at time of diagnosis. *p*<0.001.

A 10 - 3: Exponential PH regression for the colorectal cancer dataset, survival probability according to age groups (10-year age-intervals). Age at time of diagnosis. *p*<0.001.

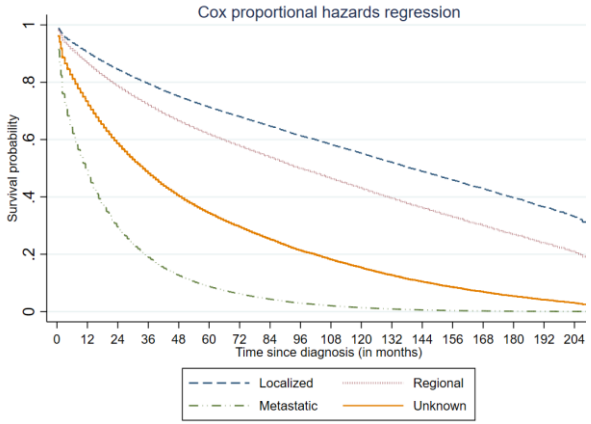
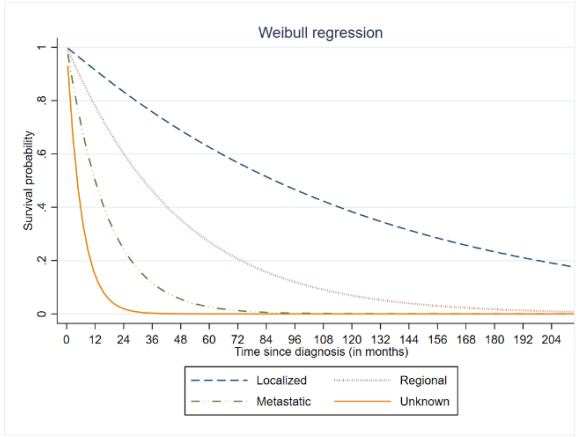
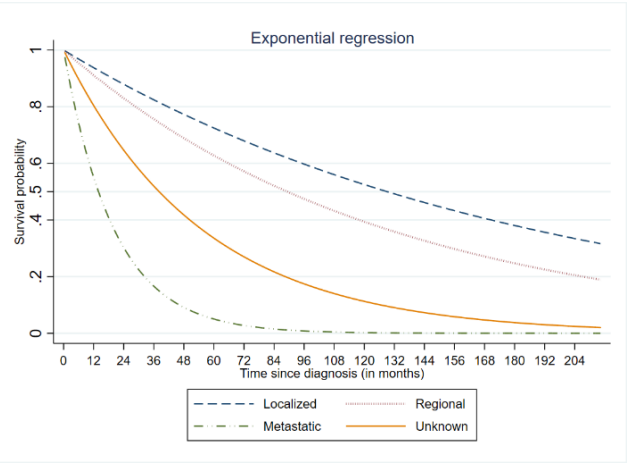


Figure A 13 -1: Semi-Parametric and parametric survival probability – colorectal cancer (stage) regression for the colorectal cancer dataset, survival probability according to diagnostic stage at time of diagnosis. p<0.001.



A 13- 2: Weibull PH regression for the colorectal cancer dataset, survival probability according to diagnostic stage at time of diagnosis. *p*<0.001.



A 13 - 3: Exponential PH regression for the colorectal cancer dataset, survival probability according to diagnostic stage at time of diagnosis. *p*<0.001.

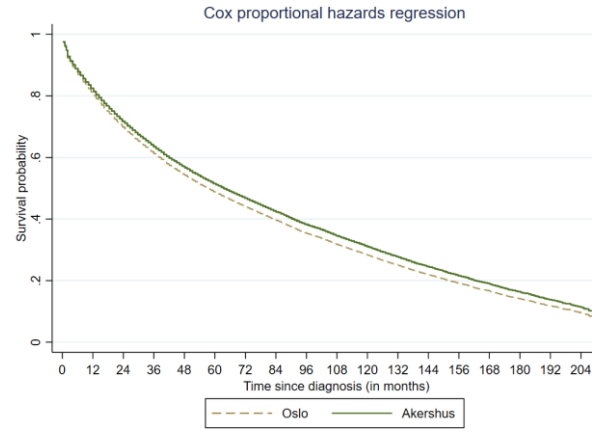
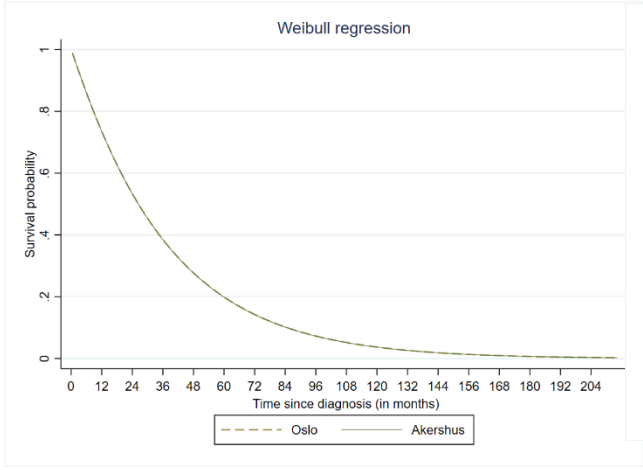
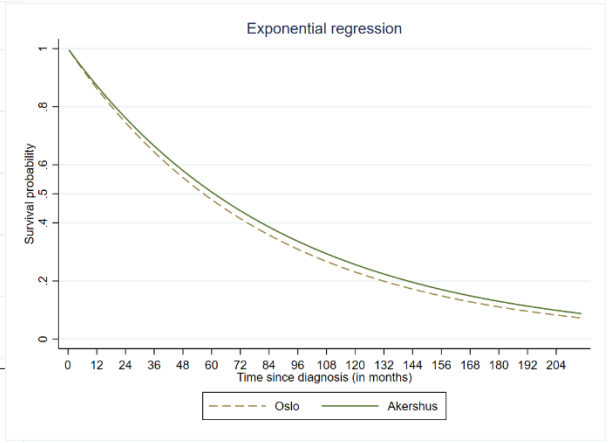


Figure A 12 -1: Semi-Parametric and parametric survival probability – colorectal cancer cancer (county) regression for the colorectal cancer dataset, survival probability according to residence at time of diagnosis. Only depicted the county with the longest and shortest median survival time, as estimated by the Kaplan-Meier estimator. p<0.001.



A 12 - 2: Weibull PH regression for the colorectal cancer dataset, survival probability according to residence at time of diagnosis. Only depicted the county with the longest and shortest median survival time, as estimated by the Kaplan-Meier estimator. *p*<0.001.



A 12 - 3: Exponential PH regression for the colorectal cancer dataset, survival probability according to residence at time of diagnosis. Only depicted the county with the longest and shortest median survival time, as estimated by the Kaplan-Meier estimator. *p*<0.001.

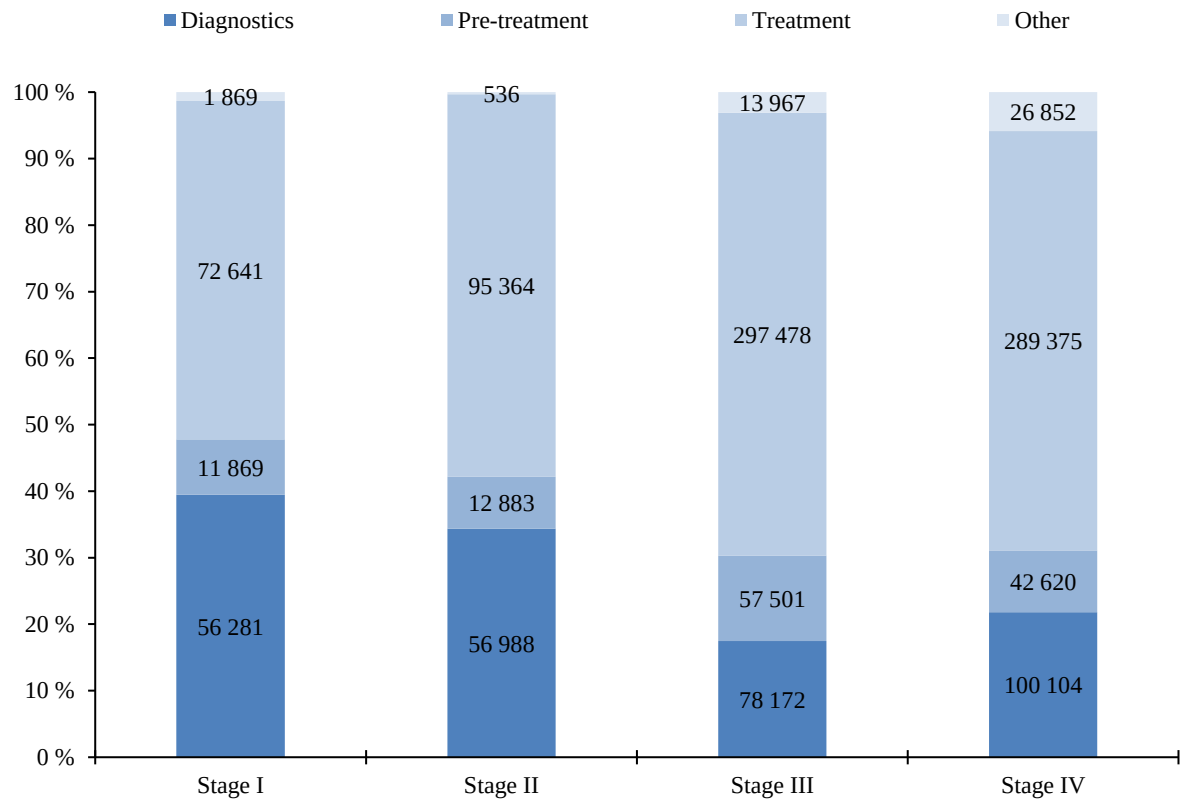


Figure A 14: Distribution of the main cost components estimated by the pathway model for lung cancer, over a five-year time horizon separated by stage I-IV. The cost component other consists of surgical complications as a result of the main treatment and costs related to at-home palliative care.