A population-based study of lung cancer in Norway –
the importance of resection rate and factors associated
with treatment and survival

Yngvar Nilssen, MSc

PhD thesis

Faculty of Medicine
University of Oslo
Contents

Acknowledgements .................................................................................................................. 4
List of papers .......................................................................................................................... 6
Abbreviations ......................................................................................................................... 7
Introduction ............................................................................................................................. 9
  Background ........................................................................................................................... 9
  Incidence, prevalence and mortality ................................................................................... 9
Aetiology ................................................................................................................................ 11
Anatomy and histology .......................................................................................................... 13
Treatment ............................................................................................................................... 14
Prognostic factors .................................................................................................................. 16
Survival .................................................................................................................................. 23
Aims of the study ................................................................................................................... 25
Material and methods ........................................................................................................... 26
  Data sources ......................................................................................................................... 26
  Data linkage .......................................................................................................................... 28
  Classification of variables .................................................................................................... 29
  Statistical methods ............................................................................................................... 33
Main results ............................................................................................................................ 40
  Lung cancer survival in Norway (Paper I) ........................................................................ 40
  Lung cancer treatment (Paper II) ....................................................................................... 41
  Resection in relation to survival (Paper III) ...................................................................... 41
Discussion ................................................................................................................................ 42
  Overview of results ............................................................................................................. 42
  Methodological considerations ........................................................................................... 42
  Discussion of the results ....................................................................................................... 50
Conclusion ................................................................................................................................ 58
Future perspectives ............................................................................................................... 59
Errors in published papers .................................................................................................... 61
References ............................................................................................................................... 62
Papers I-III .............................................................................................................................. 76
Acknowledgements

This project was conducted at the Cancer Registry of Norway, Institute of Population-based Cancer Research, in Oslo from 2013 to 2016. I am grateful to everyone who has contributed in any way to my PhD thesis, but some deserve extra acknowledgement.

First, I would like to thank the Cancer Registry of Norway, led by Director Giske Ursin, for giving me the opportunity to write this thesis. The working facilities have been excellent. I would also like to thank the South-Eastern Norway Regional Health Authority for understanding the importance of this work through their generous financial grant.

I would like to thank my main supervisor and initiator of this project, Bjørn Møller, for his outstanding supervision over the last three years. Through his understanding, patience and ability to challenge me, as well as, his willingness to share his superb expertise in the field of statistics and cancer epidemiology, Bjørn has been able to guide me through this period in a way I have found to be nothing less than perfect. Further, my questions and suggestions have always been received with a positive attitude, something I have appreciated. I could not have asked for a better supervisor.

Thank you to my project group that consisted of Bjørn Møller, Trond-Eirik Strand, Lars Fjellbirkeland, Kristian Bartnes, Odd-Terje Brustugun (co-author on Paper II), Xue Qin Yu and Dianne L. O’Connell. This well-organised, multi-disciplinary group consisted of senior epidemiologists, clinicians and statisticians who all contributed to the content of this thesis by providing constructive feedback, all within reasonable timeframes.
I would also like to thank Cancer Council New South Wales, Australia, for hosting me from October 2014 to September 2015, and particularly Professor Dianne L. O’Connell for granting my wish to work abroad as a part of my project. I am grateful that I could discuss project challenges and learn from the expertise and knowledge she and Qin possess. An extra thanks to everyone in the Health Services Research group. They all welcomed and included me in a way that was overwhelming, warm, and beyond every expectation I had for my stay.

I want to thank all my fellow PhD students at the CRN for all their contributions throughout the different phases of my project, for providing a good working environment in our team-office, and for the relaxing chats and breaks.

Thanks to the house statistician Tor Åge Myklebust and the Stata-expert Bjarte Aagnes for outstanding assistance to any and all statistical- and Stata-related challenges. Their help has been invaluable.

Finally, I would like to thank my dearest family and friends for always believing, motivating and supporting me, even at times when I struggled to see the light at the end of the tunnel.
List of papers

Paper I:


Paper II:


Paper III:

Abbreviations

ALK – Anaplastic lymphoma kinase
CCI – Charlson comorbidity index
CI – Confidence interval
CRN – Cancer Registry of Norway
cTNM – Clinical tumour, node, metastasis
EGFR – Epidermal growth factor receptor
EOD – Extent of disease
Gy – Gray
HR – Hazard ratio
ICD-O-3 – International Classification of Disease for Oncology, 3rd Edition
ICD-10 – International Classification of Disease, 10th Edition
MAR – Missing at random
MCAR – Missing completely at random
MNAR – Missing not at random
NPR – Norwegian Patient Register
NSCLC – Non-small cell lung cancer
PET – Positron emission tomography
pTNM – Pathological tumour, node, metastasis
SBRT – Stereotactic body radiation therapy
SCC – Squamous cell carcinoma
SCLC – Small-cell lung cancer
SES – Socioeconomic status
SSB – Statistics Norway
TNM – Tumour, node, metastasis
Introduction

Background

Lung cancer is the most common cancer in the world with 1.8 million new cases in 2013, which accounts for 13% of all new cancer diagnoses (1). There has been a lot of stigmatisation related to lung cancer, and a nihilistic attitude characterised the field for years (2-5). However, an indication of an upward trend in survival has been observed in Norway (6). The relationships between prognostic factors and survival, as well as, predictors and treatment, have previously been reported by studies that have used selected groups of lung cancer patients, or hospital materials (7-10). Therefore, the present project was initiated by the Cancer Registry of Norway (CRN), with funding from the South-Eastern Norway Regional Health Authority, to use a national population-based material to (i) study prognostic factors and the improvement in lung cancer survival in Norway, (ii) describe predictors for treatment for lung cancer and (iii) explore the relationship between resection rates and survival, while studying geographical variation.

Incidence, prevalence and mortality

In Norway, 3 019 new lung cancer cases (9.5% of all new cancers) were diagnosed in 2014, making lung cancer the third most common cancer type, after prostate and breast (6). Lung cancer is the most common cause of death from cancer worldwide, responsible for approximately 1.6 million deaths in 2013 (1). In Norway, lung cancer was responsible for 2 158 deaths in 2014, which is more than prostate cancer and breast cancer deaths combined, and it was also responsible for almost as many years of life lost as colon, breast and prostate cancers combined (6, 11).
Figure 1: Age-standardised lung cancer incidence (C33-34) in the Nordic countries (excl. Iceland) among men and women.

Footnote: Swedish rates are not directly comparable to those from the other Nordic countries, since the Cancer Registry of Sweden does not include information about cancer patients diagnosed based on death certificate only. Source: NordCan (12, 13).

Historically, the age-standardised incidence rates among men with lung cancer in Norway reached a plateau in the 1990s, which was 10 years and 20 years after the observed peaks in Denmark and Finland, respectively (Figure 1). For women, the overall incidence is still increasing and the trend in Norway is similar to that in Denmark, but steeper than those observed in Sweden and Finland. When comparing Nordic lung cancer incidence trends stratified by age-groups, a decreasing trend was observed in women under 65, and men under 80, while the incidence continued to increase among older female patients (12, 13). A recent study also showed that large differences in age-specific lung
cancer incidence does exist between counties in Norway (14). Since lung cancer is a lethal disease, the mortality rate follows the incidence rate closely, with an overall estimated mortality-to-incidence ratio of 0.91 and 0.85 among Norwegian men and women, respectively (15). At the end of 2014, there were 6 619 people alive after a lung cancer diagnosis in Norway, and out of these, less than 20% (1 197) were diagnosed more than five years ago (6).

Aetiology

Lung cancer is one of the few cancer types where the aetiology is known for the majority of the cases. The study by Doll and Hill in 1950 established the association between lung cancer and smoking (16). It is estimated that approximately 90% of all lung cancer cases are related to smoking (17). Other well-known causes for lung cancer are radon, asbestos and occupational exposures, as well as, both indoor (e.g. solid fuel combustion, environmental tobacco smoking) and outdoor air pollution (18).

The risk for lung cancer increases with the number of cigarettes smoked, number of years smoking and if a person started smoking at an early age (19). The close relation between smoking and lung cancer can be observed in how the historical incidence of lung cancer follows a similar shaped curve as the smoking prevalence, with a time lag. While the prevalence of daily smokers among men aged 16–74 in Norway has steadily decreased from 42% in 1973 to 16% in 2012, the decrease among women did not start until the end of the 1990s (20). The same report showed that the decrease in smoking prevalence has been largest among people aged between 16–24, regardless of gender (20).

Smoking habits have been shown to vary by region in Norway. The national average of daily smokers from 2008 to 2012 in Norway was 19%, with the prevalence varying from 14% in the region of Oslo to
28% in Finnmark (21). It is also well known that smoking habits are strongly related to socioeconomic status (SES) in the population. Smoking prevalence among people with an elementary school education (34%) is four times higher than those with a university degree or similar (8%) in Norway (Figure 2) (20).

**Figure 2**: The proportion of Norwegian population aged 16–74 that are current smokers, stratified by level of education, 1976–2015.
Figure 3: Anatomy of the respiratory system, showing the trachea and both lungs with their lobes and airways. Lymph nodes are also illustrated.

Footnote: For the National Cancer Institute © 2006 Terese Winslow, U.S. Govt. has certain rights

Anatomy and histology

The lung is anatomically divided into lobes, three on the right and two on the left side (Figure 3). From the trachea, there is a main bronchus going into each of the lungs. Lung cancer is characterised by uncontrolled growth of abnormal cells, which do not develop new healthy lung tissue. According to histological type, lung cancer is broadly divided into two main groups: non-small cell lung cancer (NSCLC) and small-cell lung cancer (SCLC). Both international and Norwegian data report that NSCLC, with the most common types being adenocarcinomas, squamous cell carcinomas (SCC) and large cell
carcinomas, represents 80–85% of all lung cancer diagnoses (22, 23). Large differences in histological type between genders have also been observed, as 28% and 42% of all lung cancers are adenocarcinomas in men and women, respectively and 44% of men and 25% of women with lung cancer have SCC (24, 25). These differences are likely to be caused by an earlier historical introduction to smoking among men compared with women (26). SCLC is named after the size of the tumour cells. These tumours are often located at the centre of the lung, and tend to grow and spread quickly.

**Treatment**

Possible treatment modalities for lung cancer patients are surgical resection, radiotherapy, and chemotherapy. Chemotherapy has become more personalised in recent years (27). Combinations of these modalities are also possible. The treatment decision is based on the histopathological diagnosis, which may be supplemented by an immunohistochemical and cancer genome (mutational) examination to obtain a more specific subgroup of histology. In addition, the localisation and spread of the tumour (using the International Classification of Disease for Oncology [ICD-O-3] and clinical stage I–IV\(^a\)), the patient’s performance status, presence of comorbidities, as well as, the patient’s own preferences are considered important when deciding the treatment (27).

Among patients diagnosed between 2010 and 2014 in Norway, 44% were diagnosed with distant spread (6). An advanced stage of disease reduces the possibility for curative treatment and the probability of achieving cure. Patients diagnosed with NSCLC in stage I or II are candidates for surgery with curative intent. It is known that approximately 20% of lung cancer patients in Norway are resected every year (23, 29). Adjuvant platinum-based chemotherapy is offered for patients in stage

\(^a\) Classified according to *Staging Manual in Thoracic Oncology* into stage I: \((T1, N0, M0)\), stage II: \((T2, N0, M0)\), stage III: \((T1, T2, N1, M0)\) or \((T1, T2, N2, M0)\) or \((T3, N0, N1, N2, M0)\), and stage IV: \((T4, any \; N, M0)\) or \((any \; T, N3, M0)\) or \((any \; T, any \; N, M1)\) (28).
II. The majority of resections performed are (bi-) lobectomies, where one or two of the lobes are surgically removed. Other resection alternatives include pneumonectomy and sub-lobar resection. Neoadjuvant or adjuvant radiotherapy, i.e. radiotherapy given before or after surgery, respectively, can also be used, if required. If the patient is not considered a surgical candidate due to technical or medical inoperability, either traditional radiotherapy or stereotactic body radiation therapy (SBRT), can be offered. During SBRT, high-dose radiation is directly aimed at the tumour from multiple angles. Compared to traditional radiotherapy, the benefits of SBRT include better preservation of the normal tissue surrounding the tumour and improved survival. However, this technique is only possible in N0-situations, i.e. no lymph node metastases. Due to the heterogeneity of patients diagnosed in stage III, the differentiation between treatment decisions is based on T- and N-stage.

While the debate continues as to whether or not patients with stage III disease and limited N2-metastases should undergo surgical resection, Norwegian national guidelines recommend that this group of patients receive radiotherapy in concomitant combination with chemotherapy.

For stage IV patients, treatment should be given with life-prolonging or palliative intent, i.e. chemotherapy, palliative radiotherapy, or a combination of the two. Cytostatic treatment with cisplatin or carboplatin, in combination with vinorelbine, gemcitabine, docetaxel, paclitaxel or pemetrexed, is considered standard first-line therapy in patients with no genetic aberrations amenable for targeted therapy. Approximately 10% of patients harbour mutations in the epidermal growth factor receptor (EGFR) gene and should be treated with first-line erlotinib, gefitinib or afatanib. Crizotinib is the treatment of choice in the approximately 5% of patients with changes in the anaplastic lymphoma kinase (ALK)-gene. These targeted therapies are all oral drugs, containing a kinase inhibitor, and are taken daily as long as they provide benefit to the patient. All patients will inevitably relapse, and approved second line therapies include docetaxel or pemetrexed for non-mutated patients. Novel kinase inhibitors are preferable for patients with known mutations.
Mainly SCLC patients with stage I (<10%) are considered candidates for surgery with adjuvant chemotherapy, given as with four courses of cisplatin in combination with etoposide. All resected SCLC patients should also undergo prophylactic brain irradiation. For all other SCLC patients (>90%), surgery is considered futile. For non-resectable patients with limited spread of disease, four courses of cisplatin and etoposide should be given, with 3 weeks of radiotherapy intercalated between courses two and three. For SCLC patients with extended disease and good general health, four courses of platinum and etoposide are recommended, while for patients with reduced general health, four courses of doxorubicin, cyclophosphamide and vincristine are recommended.

Prognostic factors

Prognostic factors (or predictors) are defined as variables that can account for some of the heterogeneity in the course of the disease and the ultimate outcome for the patients (30, 31). Understanding these factors may help answer a wide range of important questions, such as predicting the outcome or prognosis for individual patients, and providing information about possible differences in the quality and health care services provided between sub-groups of patients (32). For the purpose of this thesis, prognostic factors are divided into three different groups: tumour-, patient- and treatment-related factors. Tumour-related factors are those directly related to the biological aspects of the tumour, such as extent of disease (EOD) and histology. Patient-related factors are not directly related to the cancer, but more specific to the individual, such as gender, SES, comorbidity, smoking status and area of residence. Treatment-related factors include the treatment modality, procedure and the expertise of the clinician (33). While it is possible to quantify tumour- and patient-related factors and whether or not patients receive surgery, other factors that are associated with quality of treatment and personal experiences, cannot as easily be measured.
Tumour-related

Extent of disease

Tumour stage is an important prognostic factor, as the 5-year survival rates range from 50% (stage IA) to 2% (stage IV), and from 73% (stage IA) to 13% (stage IV), according to clinical tumour, node, metastasis (cTNM) and pathological tumour, node, metastasis (pTNM), respectively (34). A study from 2012 including lung cancer patients from Australia, Canada, Denmark, Norway, Sweden and the UK, showed significant differences in stage distribution between the countries, as well as, significant survival differences between the different stages (35).

Histology

A number of studies have examined the association between histological group and survival. A study from 2012 showed that histology may be an independent prognostic factor (36). While a Norwegian study considering all lung cancer patients found no difference between SCC and adenocarcinoma, they found a 12% increased relative risk of death when comparing large-cell carcinoma with adenocarcinoma (37). Another Norwegian study, only including resected patients, found a 44% and 33% increased mortality when comparing adenocarcinoma and large-cell carcinoma, respectively, with SCC (38). Results from the International Association for the Study of Lung Cancer (IASLC) identified histology to be an independent prognostic factor among resected patients (39). While large-cell carcinoma was associated with a 19% increased mortality compared to SCC, a difference only observed among men, the results slightly favoured SCC compared to adenocarcinoma (hazard ratio [HR] = 0.86, p<0.0001). A study using population-based data from seven regions in Spain, reported varying 5-year relative survival estimates with histological type, favouring SCC and adenocarcinoma, while SCLC patients had the worst prognosis (40). Therefore, histological subgroup should be considered a potential prognostic factor when studying survival.
Patient-related

Socioeconomic status

One of the patient-related prognostic factors, SES, is intended to measure a person’s social position (41). SES is difficult to measure and therefore, income, education, marital status and/or area of residence are commonly used as proxies. In Norway, individual-level information regarding both income and education is available from Statistics Norway, however, there may be situations where agreement between these measures and a person’s SES is poor. For example, patients who are not working would be registered with a low income and hence categorised with a low SES. However, these patients may have partners who are working and who can financially support both. Hence, the patient’s personal income as a proxy for SES would cause a misclassification, while including information about the total household income would be more appropriate.

A recent systematic review and meta-analysis showed that lung cancer patients with high SES are more likely to receive both surgery and chemotherapy, while the influence of SES on whether the patient receives radiotherapy remains inconclusive (42). These results were found both in SCLC and NSCLC patients (42-52). When reviewing literature on the relationship between SES and survival, it was reported that these survival estimates were ambiguous (53). However, studies from Sweden, Denmark and England have reported that the survival of lung cancer patients is affected by the patients’ SES (43-45, 49, 54, 55). Differences in lifestyle, culture and behaviour (e.g. smoking habits), may also be related to SES, and these may further influence the patients’ health. Similar to countries like Sweden, Denmark and England, Norway has a universal healthcare system where healthcare is equally available to everyone independent of social factors and area of residence. This contrasts with the insurance-based system in the US, and therefore, care should be taken when comparing the SES estimates of Norway, with those from non-universal healthcare systems, where the differences between SES groups are expected to be larger.
Smoking status

In addition to being the dominating aetiological factor, smoking has been studied in relation to cancer recurrence and survival. A review article from 2013 showed that patients who continue to smoke after a lung cancer diagnosis are 1.9 times more likely to get a recurrent tumour, 2.3 times more likely to get a second tumour, and have 2.9 times higher overall mortality than patients who quit smoking at the date of diagnosis (56). This review also found that for patients receiving palliative treatment, smoking cessation at time of diagnosis was associated with improved pulmonary function, weight gain and better overall quality of life. While there are some inconsistencies in reported results, a systematic review and meta-analysis showed that among early stage NSCLC patients, continuing smoking compared with smoking cessation was associated with almost a 3-fold increase in all-cause mortality, while the comparable increase for limited SCLC was almost 2-fold (57-60). Other studies have shown that there is a significant positive effect on survival of being a never smoker compared to an ever smoker, with estimates varying from a 5–50% reduction in risk of death (61, 62). However, among patients with stage II and III, the results are inconclusive (63, 64).

Comorbidity

As a large proportion of the lung cancer patients are current or former smokers, they are more likely to have reduced lung function, in addition to a number of other conditions (65). These other conditions are called comorbidities, and are often summarised into a score or index used in epidemiological studies. The most commonly used comorbidity index is called the Charlson comorbidity index (CCI), which gives a score to 17 different chronic diseases based on their severity (66). This score includes chronic obstructive pulmonary disease, the most common comorbidity for lung cancer, as well as, myocardial infarction, congestive heart failure, peripheral vascular disease, cerebrovascular disease, dementia, rheumatic disease, peptic ulcer disease, mild liver disease,
diabetes with and without chronic complications, hemiplegia or paraplegia, renal disease, cancer, moderate or severe liver disease, metastatic cancer and AIDS/HIV (66, 67). From national and international guidelines, it is known that information about the patient’s general health, lung function and comorbidities are important when making a treatment decision. While it would be optimal to have information on all three factors, comorbidity information which often serves as a proxy of general health, is most accessible in large national population-based materials. One study showed that compared with patients who have a low level of comorbidity, patients with a high level had a 35% increased 1-year mortality and a 26% increased 5-year mortality (68). A literature review on the association between comorbidity and lung cancer survival found that having more comorbidities was associated with a 10-50% increased mortality (69). However, as the prognosis for lung cancer is considered poor, it has been shown that having comorbidities is of relatively low prognostic importance. (70, 71).

Symptoms

The most common symptoms that lung cancer patients display are progressive shortness of breath, coughing (blood), chest pain/oppression, hoarseness or loss of voice and pneumonia (72). Unfortunately, these symptoms are most likely not to be present until the tumour has metastasised beyond the primary site. Vague symptoms can lead to later patient contact with a doctor, and hence later diagnosis. There have been studies examining the presence and duration of symptoms in relation to survival. A Korean study showed that among NSCLC patients there was a significant reduction in the risk of death (odds ratio [OR]: 0.23 95% confidence interval [CI] 0.22–0.52) for asymptomatic compared to symptomatic patients at the time of diagnosis, while no effect was observed among SCLC patients (73). In a local centre study from England where they considered 5-year survival among all resected NSCLC patients between 2000 and 2009, there was no significant difference between asymptomatic and symptomatic patients (74). In India they found that patients who had symptoms for less than one month before diagnosis had a greater than 50% reduction in
the risk of death (HR = 0.44 95% CI 0.26–0.74) during the next 30 months (75). A systematic review from 2009 and a review article from 2014 identified the presence of symptoms to be a significant negative prognostic factor for lung cancer outcome (76, 77). However, others have shown that persons living with symptoms for a short period of time had worse prognosis compared to those living longer with symptoms. Hence, it is not only the presence of symptoms that is important to consider, but also its duration.

Gender

Gender differences in regard to survival have been studied extensively, and the results consistently show that women have a better prognosis than men (78). Previous studies from Norway showed that women had a 14% and 41% improved survival compared to men, when analysing all lung cancer patients and resected patients, respectively (38, 79). The latter result is comparable to a study from the United States (US) that reported a 50% lower 30-day post-operative mortality among women compared with men (80). Other studies examined the gender differences among NSCLC and SCLC patients, separately, and found a 15–20% higher survival among women in both groups (81-83). A Polish study found that men had a 15% increased risk of death compared to women (84). Another study of NSCLC patients found better overall survival as well as stage-specific survival among women (85). Unless these results are (residually) confounded by factors that could not be adjusted (sufficiently) for, it is important to adjust for gender as a prognostic factor. An example of a possible confounder is smoking, as it is well known that smoking habits differ between men and women. Therefore, it is important to interpret the previous results with care, and keep in mind that there may be other associations disguised as a gender difference.
Area of residence

Another patient-related prognostic factor is area of residence. The results of a number of European studies have shown that the probability of getting treatment differs within countries (29, 43, 86-88). The proportion of patients being resected varied between 15–31% in Norway, 3–18% in England, 13–24% in Denmark and 8–16% in Ireland. Marked survival differences between regions in Norway, Sweden and England have been reported (43, 88-90). From England, it is known that the regional differences in survival decreased between the periods 1991 to 1995 and 2001 to 2006, both among men and women (91). Even though the geographical differences in overall survival got smaller, two studies showed that among NSCLC patients there is still significant variation in survival (43, 92). Hence, area of residence seems to be an important prognostic factor for survival.

Treatment-related

The 5-year overall survival among all lung cancer patients is approximately 15%, however, resected patients experience a 5-year survival ranging from 20% to 80% depending on their stage (93). A Norwegian population-based study, examining prognostic factors among resected patients, showed that having a more extensive procedure than lobectomy was associated with a worse prognosis (38). Even though studies showed significant association between resection rates and survival, the optimal proportion of patients who should be resected has not been found. In addition, the varying definitions of a resection rate between countries make it difficult to compare results (94-96). For patients receiving radical radiotherapy in stages I and II, the 5-year survival was 17% according to a published Cochrane report (97). If left untreated, barely any of these lung cancer patients would be alive after three years (98).

Further, a strong association between overall survival and regional variation in resection rates has been reported (43, 92). Results from the Lung Cancer Register of Central Sweden, observed that the
risk of death was 20–40% higher in other county centres than the reference centre (88). However, after adjusting for treatment (surgery, radiotherapy, chemotherapy), county of residence was no longer considered a prognostic factor.

Survival

Worldwide, while survival after a lung cancer diagnosis has been considered poor for decades, there have been recent indications of a promising positive trend in survival (99-101). In addition, variation in survival estimates between countries has been observed (15, 99, 100, 102-107). In 2013 in Norway, the median survival time was 8.2 and 12.3 months for men and women, respectively (27). Excluding Iceland due to its small population which would affect the relative survival estimates by random variation from year to year, there have historically been differences between the Nordic countries in survival estimates among men and women. The NordCan database shows that for the period from 2009 to 2013, the 5-year relative survival estimates varied from 10 to 15% and from 16 to 19% among men and women, respectively (Figure 4) (12, 13). The survival estimates were 15% for men and 19% for women in Norway during this period. Norway and Sweden had the highest survival estimates, while Denmark and Finland had the lowest, according to the NordCan database (12, 13).

Varying 5-year relative survival has been observed in regions with similar universal healthcare systems, ranging from 8.7% in England to 20.1% in Manitoba, Canada, for the period from 1995 to 2007 (103). It was argued that the low survival in England can be attributed to the fact that cancer patients seek contact with their doctors at a later stage of disease (104). The EUROCARE-5 study estimated the 5-year European mean relative survival of lung cancer to be 13.0%, which was the poorest of the ten index cancer sites. The CONCORD-2 study, which incorporates worldwide information from 279 population-based registries in 67 different countries including over 25 million cancer patients, reported a 5-year relative lung cancer survival of <20% all over Europe, 15–19% in North America and 7–9% in some parts of Asia, in the period from 1995 to 2009 (100).
**Figure 4:** Showing the improvement in 5-year relative survival from 1999–2003 to 2004–2008, as well as, from 2004–2008 to 2009–2013 in the Nordic countries (excl. Iceland) among men and women.

*Source: NordCan (12, 13).*
Aims of the study

The aims of the present study were to:

- examine changes in survival, and patient-, tumour- and treatment-related factors affecting survival, among resected and non-resected lung cancer patients (Paper I).
- identify subgroups of age, gender, SES, histology and treatment in relation to improvement in survival over the last decade (Paper I).
- examine and quantify the association between possible predictors and surgical treatment, radical radiotherapy or palliative radiotherapy for lung cancer patients (Paper II).
- examine if regional variation in survival exists among lung cancer patients in Norway and, if it does, can the variation be explained by varying resection rates (Paper III).
- explore the relationship between survival and resection rate, and investigate whether an optimal resection rate can be identified (Paper III).
Material and methods

Data sources

This national, population-based study involves three sources of information: Cancer Registry of Norway (CRN), Statistics Norway (SSB), and Norwegian Patient Register (NPR). A unique personal identification number has been assigned to every Norwegian citizen since 1964. This personal identification number allows linkage of information on all Norwegian citizens across institutions and national health registries.

Cancer Registry of Norway

It is mandatory for all hospitals, pathology laboratories and general practitioners in Norway to report all newly diagnosed malignant neoplasms to the CRN. The CRN has data on cancers in Norway dating back to 1953. The CRN also receives death certificates for all patients with a cancer diagnosis from the Cause of Death Registry, which is operated by the Norwegian Institute of Public Health. Using the personal identification number, the CRN is linked monthly to the National Population Register to update vital status (death or emigration), and three times per year with the NPR to ensure completeness of cancer cases (6). The quality (i.e. comparability, completeness, validity and timeliness) of the data in the CRN has been evaluated to be high (108).

Statistics Norway

Statistics Norway was established in 1876 and is a governmental entity that falls under the Ministry of Finance (109). It is considered to be an independent scientific institution as it decides when and what to publish. Its main objective is to publish statistics about Norwegian society regarding many different areas, such as population, health, finance and education. Statistics Norway does not directly collect data from the population regarding education and income, but receives that information from other relevant administrative registers. The errors in these data are considered to be negligible (110).
Every Norwegian citizen has to declare his income and wealth annually to the tax authorities, who also collect these data from employers, banks etc. These data are then transferred to Statistics Norway. In addition to tax files, Statistics Norway also collect various tax-free transfers from other administrative registers, primarily from The Labour and Welfare Administration (110, 111). When Statistics Norway publishes information about education, it uses information from “Nasjonal utdanningsdatabase”, “Nasjonal vitnemålsdatabase”, “Helsepersonellregisteret” and “Utlendingsdatabasen”. For immigrants with unknown education, a small proportion of persons are directly contacted to collect this information (112). Then SSB classifies the education level based on the Norwegian Standard Classification of Education (113).

**Norwegian Patient Register**

The Norwegian Patient Register (NPR) falls under the Norwegian Directorate of Health and is a national health registry covering all sectors of specialised health care services. Reporting to NPR is mandatory, and the register includes data on all patients treated in Norwegian government-funded institutions. Personal identification numbers have only been reported to the NPR from 2008 onwards. This enables researchers and health planners to follow the disease trajectory of patients between sectors and hospitals. In addition, alignment of data and validation with other national health registries are made feasible. The NPR data consist of three main sources for statistics: visits for medical treatment for in- and outpatients at publicly financed hospitals, private hospitals and private specialist practices. It is important to have data from all these sources, as the government purchases medical treatment from private hospitals and private specialist practices as a supplement to services at the public hospitals. The NPR does not include data on privately financed hospital treatments, however, in 2008 only around 0.5% of all health care services were provided by these hospitals (114). The basic data unit in the NPR is hospital visits. However, when a patient is transferred between wards at the same hospital, the individual data records are aggregated. Each episode of national
hospital data contains one or more diagnoses, coded according to the International Classification of Diseases, 10th Edition (ICD-10) classification.

**Data linkage**

In this study, the national, population-based data registered in the CRN were used to identify all patients diagnosed with malignant neoplasm of bronchus and lung (ICD-10 code C34) between 1 January 1997 and 31 December 2011 in Norway (n = 34 157). In order to have a homogeneous group of lung cancer patients, those with tracheal cancer (C33) were excluded (n = 49). If a patient was registered with multiple tumours, only the first case within ICD-10 group C34 was included. For example, if a patient was diagnosed with a tumour in one lobe and later diagnosed with another independent tumour in a different lobe, only the first diagnosis would be included in this study. Patients registered as dead before diagnosis, and patients whose diagnosis was solely based on death certificate and autopsy, were excluded from the study (n = 886). Information from Statistics Norway was linked with indicators for SES, i.e. education, personal and household incomes. Education was measured by the highest achieved education at the year of diagnosis, while personal and household incomes were measured the year prior to lung cancer diagnosis. Education and personal income were available for the whole study period, while household income was only available after 2003. Finally, information about co-existing diseases (i.e. comorbidities) during one-year prior to diagnosis was obtained from the NPR, but this was only available for patients diagnosed after 1 January 2009.

In Paper I, data on all lung cancer patients registered in the CRN from 1997 to 2011 were used. The rationale behind starting the study period in 1997 was that this was the first year radiotherapy data on a national level were available. In Paper II, all lung cancer patients identified in the CRN from 2002 to 2011 were included, as health trusts were defined and came into effect in Norway from January 1
2002, and since regional variation was an aim of the study (115). For Paper III, only NSCLC patients who were diagnosed from 2002 to 2011 were included. Patients with SCLC rarely undergo surgery, so they were excluded in this study (27). Figure 5 shows the linkage of the different data sources and shows which patients were included in the different papers.

**Figure 5:** Flowchart showing the data linkage and study populations for papers I–III.

![Flowchart](image)

**Classification of variables**

The following information was available for the lung cancer patients in the study: date of diagnosis, date of death or last observation, age, gender, EOD, education, personal income, household income (after 2004), histology, topography, date and type of treatment (surgery and radical or palliative radiotherapy), laterality, comorbidity (after 2008), symptom duration and smoking status (2004–2010). In addition, due to the extensive quality control work done by H. Rostad and T.E. Strand internally at the CRN, surgical procedure, pTNM, tumour size and resection margin information were also available for resected patients.
Radiotherapy

The CRN receives electronic records from all radiotherapy centres annually, with information about ICD-10 group, date of treatment, treatment intention, total radiation dose and the number of fractions. The possible treatment intentions are curative, local control, prophylactic or palliative, and for the purpose of this thesis the first three intentions were grouped together as radical (116). For 95% of the lung cancer cohort, the intention was known, while for the last 5% of patients, total radiation dose given was used to categorise them as either radical or palliative. For NSCLC patients who did not undergo surgery, a total radiation dose over 60 Gray (Gy) was considered as radical, and the comparable limit for resected patients was 50 Gy. For SCLC patients, radiation doses of 42 Gy or higher were classified as radical. Any radiation dose lower than those described above, was classified as palliative.

Histology

For Papers I and III, information about histology was obtained using all information registered in the CRN and by choosing the most informative and specific subgroup. For Paper II, information about the most specific histological subgroup registered in the CRN before the time of resection was used for the resected patients. However, if the patients were not resected, the approach for classifying histology was the same in Paper II as it was in Papers I and III.

Histology was classified based on the 2004 version of the World Health Organization classification as SCC, adenocarcinoma, small-cell carcinoma, large-cell carcinoma, other specified carcinomas, carcinoma not specified and unknown histology (72). Table 1 shows which ICD-O-3 codes that were included in the different histology groups. Since sarcomas are biologically different from the other tumours, and because the number of sarcoma patients is very low (0.2%), these tumours were not included in this study.
**Table 1**: Shows which morphology codes (ICD-O-3) are included in the different histological groups.

<table>
<thead>
<tr>
<th>Histological group</th>
<th>ICD-O-3 Codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Squamous cell carcinoma</td>
<td>8050–8076</td>
</tr>
<tr>
<td>Adenocarcinoma</td>
<td>8140, 8211, 8230–8231, 8250–8260, 8333, 8341, 8480–8490, 8550, 8560, 8570, 8574</td>
</tr>
<tr>
<td>Small cell carcinoma</td>
<td>8040–8045</td>
</tr>
<tr>
<td>Large cell carcinoma</td>
<td>8012–8031, 8310</td>
</tr>
<tr>
<td>Other specified carcinoma</td>
<td>8046, 8082, 8123, 8200, 8240, 8244, 8246, 8249, 8430</td>
</tr>
<tr>
<td>Carcinoma, not specified</td>
<td>8010, 8032, 8033</td>
</tr>
<tr>
<td>Unknown</td>
<td>6900, 6999, 8000, 8001</td>
</tr>
</tbody>
</table>

**Extent of disease**

It is important to include information about EOD both for analysing prognosis and likelihood of treatment for lung cancer patients. In the CRN, the variable describing EOD at the time of diagnosis is grouped into localised, regional, metastatic or unknown according to the condensed tumour, node, metastasis (TNM) status (117). Localised tumours are defined as tumours with no direct growth into neighbouring tissue, lymph nodes or organs, however, there can be micro-invasive growth or carcinoma with the beginning of a microscopically infiltrating tumour.Regional tumours are defined as tumours with metastasis to regional lymph nodes or microscopic/macroscopic growth into neighbouring tissue. Metastatic disease involves metastasis to distant lymph nodes or metastasis to organs in the same/different part of the body as the primary tumour. The EOD is coded as unknown if a metastasis is found but the location of the primary tumour is uncertain.

Before 2008, EOD was coded as unknown if the coding was solely based on a pathology report, i.e. no valid clinical notification, and there was no information about metastasis at the time of diagnosis.

After 2008, these cases were coded as localised if they received curative surgery. To obtain consistency in the data and to avoid bias in the analyses, all stage information post-2008 were considered unknown. This approach led to some methodological challenges when it came to
analysing changes in either survival (Papers I, III) and treatment over time (Paper II), which will be addressed later.

**Socioeconomic status**

All the analyses included education and income, both serving as proxies for SES. In Papers I and III, personal income was used as a proxy for SES, while in Paper II, household income was used. While both personal and household incomes were explored as predictors in Paper II, it was concluded that household income was a better predictor of treatment. This was especially noted for women; while for men using personal or household income, did not change the results markedly. Education was categorised based on the number of years of education: low (1–9 years, lower secondary school), intermediate (10–12 years, upper secondary school) and high (12+ years, university or similar). Data on both personal and household incomes were defined based on the percentiles. The cut-points were set at the 33rd (low) and 66th (high) percentiles, and were redefined every year to adjust for the increase in income over time. In addition, when redefining these cut-points for personal income, gender was taken into account.

**Health trust**

In 2011, Norway consisted of 21 health trusts, which are responsible for general healthcare and management of all patients residing in its geographical catchment area. If a health trust does not provide certain services (e.g. lung cancer resection), these patients are referred to another health trust that offer the appropriate treatment. The study variable denoting health service region (health trust) was chosen instead of county, because it represents the actual catchment area of the different treatment institutions. Health trust is an explanatory variable that is based on the patient’s place of residence at the time of diagnosis, independent of where the patient was treated.
Comorbidity

Comorbidity information was measured using a modified version of the CCI, which was constructed by using diagnostic codes (ICD-10) from hospitalisations within one-year prior to, and including, the date of diagnosis for a list of 17 chronic diseases. A score was determined for each of a patient’s recorded comorbid disease based on its severity, and the combination of these scores resulted in a modified CCI. The index was categorised into: “no hospital admissions before lung cancer diagnosis” (CCI = -1), low (CCI = 0), intermediate (CCI = 1, 2) and high (CCI ≥ 3) (66, 118).

Statistical methods

In addition to standard descriptive statistics, a number of different statistical methods were applied in this study.

Cox proportional hazard regression

In Paper I, Cox proportional hazard regression analyses were performed to identify and examine the effects that different prognostic factors have on 1-year survival among the following three groups of lung cancer patients: all patients diagnosed, non-resected patients, and resected patients. Cox regression was also used in Paper I to identify which sub-groups of patients had the largest improvement in survival over time. The underlying assumption (i.e. the proportional hazard assumption) of this model is that the explanatory covariates are multiplicatively related to the baseline hazard, and that the ratio of the hazards comparing groups with different values of the explanatory variables, remains constant over time. The Cox regression model can generally be expressed as \( h(t) = h_0(t) \exp(\beta_1 x_1 + \ldots + \beta_n x_n) \), where \( h(t) \) is the hazard function, \( h_0(t) \) is the baseline hazard function and \( x_1, \ldots, x_n \) are the covariates with their corresponding parameters \( \beta_1, \ldots, \beta_n \) (119).
Logistic regression

In Paper II, three separate multivariable logistic regression models were estimated to investigate how different covariates influence the odds of receiving surgical treatment, as well as, curative and palliative radiotherapy, as a patient’s first treatment within one-year of a lung cancer diagnosis. In each of the three models, the dichotomous outcome was defined as either receiving treatment or not. The logistic regression model can be expressed as \( \ln \left( \frac{p}{1-p} \right) = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n \), where \( p \) is the success probability of the event of interest, in this case, the patient receiving treatment, \( \beta_0 \) is the intercept which is the estimate as all covariates are zero, \( x_1, \ldots, x_n \) are the covariates and \( \beta_1, \ldots, \beta_n \) are the corresponding parameters (120).

Net survival and excess mortality

Net cancer survival is the probability of surviving in a hypothetical world where cancer is the only possible cause of death (121). It provides a measure of the excess mortality associated with being diagnosed with cancer. Cause-specific survival and relative survival are two ways of estimating net survival. The major advantage of using a relative survival framework, is that it does not require any cause of death information, as this information cannot always be trusted, especially for older patients (\( \geq 85 \) years) (122). Net survival can be described as \( NS(t) = \frac{1}{n} \sum_{i=1}^{n} \frac{S_i(t)}{S_i^*(t)} \), where \( S_i(t) \) is the all-cause survival for patient \( i \) weighted using the inverse of the cumulative expected survival for a comparable, lung cancer-free individual (\( S_i^*(t) \)). This comparable individual is found by matching the lung cancer patient cohort with an age, gender and calendar year stratified Norwegian lifetable obtained from Statistics Norway. The lifetable used in this thesis was not adjusted for smoking status.
in the general population, however, it has been shown that additional adjustments may have little
effect on the net survival estimates (123). Five-year net survival was estimated using the method
proposed by Pohar-Perme in 2011, implemented in the Stata command strs, for both resected
and non-resected lung cancer patients separately in Paper I (124, 125)b. In Paper III, the relative
excess risk of death among patients with localised, regional and metastatic disease by health trust
was modelled. The excess risk of death, \( \lambda(t) \) at time \( t \) can be expressed as \( \lambda(t) = h(t) - h(t^*) \), where
\( h(t) \) is the all-cause mortality rate experienced by the patients and \( h(t^*) \) is the corresponding
expected mortality rate. Comparing the excess risk of death for the different health trusts to that of
the entire country resulted in estimates of the relative excess risk of death. In order to account for
case-mix differences between health trusts that might affect survival, a Poisson regression model
was used to adjust for available explanatory variables. The Poisson regression model estimated the
expected number of deaths due to causes other than lung cancer. These were estimated using the
general population mortality rates (125, 126). The general form of a Poisson regression with a
logarithmic link function can be written as \( \log(\mathbb{E}(Y \mid x)) = \alpha + \beta_1 x_1 + ... + \beta_n x_n \), where \( Y \) represents
the number of counts of an event, \( \alpha \) is a scaling variable called the offset term which is used to make
the different groups comparable and \( x_1, ..., x_n \) are covariates with their corresponding parameters
\( \beta_1, ..., \beta_n \). In Paper III, the link function proposed by Dickman et al. in 2004 was used when modelling
relative excess risk of death (125).

Competing risk

While net survival aims to estimate the hypothetical survival probability in a world where cancer is
the only possible cause of death, competing risk can be used to estimate real-world survival
probabilities, i.e. survival probabilities in a situation where other causes of death also exists (127). In

b In Papers I and III the term relative survival was used to denote net survival.
Paper II, the cumulative incidence using the Aalen-Johansen estimator, i.e. the probability of patients experiencing surgery and radical or palliative radiotherapy as their first treatment within one-year of lung cancer diagnosis, was estimated using Stata’s `stcomp` command, under the assumption that any of the other treatment modalities (competing risks) could happen (128, 129).

Joinpoint regression

A joinpoint is a point where two linear lines with different slopes meet. By using the Joinpoint Regression Program available from Surveillance, Epidemiology, and End Results Program (SEER), it is possible to analyse linear trends, identify changes in linear trends and determine the number of significant joinpoints (130). The program fits the simplest model as a straight line, i.e. zero joinpoints, and then determines if additional joinpoints should be added to the model. In Paper III where the relationship between surgery and survival was studied, the relative excess risks of death with standard error, as well as, the resection rates were estimated for each health trust per year. Using these estimates in the joinpoint program to plot the relative excess risk of death against the resection rates, an optimal rate of resected patients, both among patients with localised and regional diseases, was sought. The program calculated the annual percentage change, which in this case was interpreted as the average change in relative excess risk of death as the resection rate category increased by one unit. Hence, an inflection point in the relative excess risk of death where the survival stabilises or declines, for increasing levels of resection, would identify an optimal resection rate.

Multiple imputation

In medical and epidemiological research there will be missing information in the data for a number of reasons. The degree of missing information varies from dataset to dataset and from study to study.
There may be many possible reasons for data to be missing, but three kinds of missing mechanisms are identified in the literature: missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR). For data to be MCAR, the information missing should be completely independent from both observed and unobserved data. Data are considered MAR if the probability of missing data does not depend on unobserved data. And finally, data are MNAR if the probability of missing data does depend on unobserved data.

Historically, there have been different ways to deal with missing information, e.g. complete case, mean imputation, last observation carried forward and treating missing data as a new category (131). If missing data are handled inadequately, the statistical analyses will lead to biased and/or inefficient estimates. Treating the unknown data as a separate category has been shown to be a poor option, even when the data are MCAR, since severe bias can arise in parameter estimations (132, 133). Using the complete case approach, which is deleting all observations with unknown information, will lead to correct and unbiased estimates, but only if the missing data are MCAR. However, as part of the given data will be excluded, the statistical power of the analysis will decrease.

Imputation as an alternative approach to handle missing data needs to be carefully considered. Using a single imputation approach, i.e. replacing the missing observation with a single value, will result in standard error estimates that are too small. In contrast, multiple imputation procedures have become widely accepted as a standard approach to handling missing data. In order to use this methodology and obtain unbiased and efficient estimates, the nature of the missing data is important. To achieve asymptotically unbiased estimates, it is important that the data meet at least the MAR assumption. Multiple imputation can be performed even on MNAR data, however, then the mechanism of missing data needs to be modelled as well.
Multiple imputation is generally performed in three steps: (i) generating multiple ($m$) imputed data sets, (ii) analysing each of the imputed data sets and (iii) pooling the estimates from the different analyses together. Multiple imputation by chained equations (MICE) is an approach to construct $m$ imputed datasets. These are based on a set of imputation models, i.e. one model for each variable with missing values. The initial step involves filling the missing values in each variable with a random replacement from the observed data. Then the missing values of a variable, say $y_1$, will be regressed on the other variables $y_2, \ldots, y_n$, using the individuals where $y_1$ is observed. Initial missing values of $y_1$ is replaced with simulated draws from the obtained posterior predictive distribution of $y_1$. The same procedure follows for $y_2$, which is regressed on $y_1, y_3, \ldots, y_n$ restricted to the individuals with observed $y_2$, using the imputed values on $y_1$. Values for missing $y_2$ is replaced by draws from the posterior predictive distribution of $y_2$. This is repeated for all other variables and the process is often called a cycle. In order to stabilise the results, several cycles are performed resulting in one imputed dataset. This procedure is then repeated $m$ times to give $m$ imputed data sets. The second step is analysing the $m$ datasets separately and is usually easy as standard analysis tools can be used.

Finally, using Rubin’s rule, the estimates and variance-covariance matrix from the $m$ different imputed datasets can be pooled together. The combined estimate of the individual obtained estimates ($\hat{\theta}$) can be found by taking the average over all imputations, i.e. $\hat{\theta} = \frac{1}{m} \sum_{j=1}^{m} \theta_j$. The combined variance-covariance matrix includes the components describing both within-imputation variability (i.e. variation between the different imputed dataset) and between-imputation (i.e. reflection of the uncertainty due to missing information). The total variance obtained can be expressed as $\text{var}(\hat{\theta}) = W + \left(1 + \frac{1}{m}\right)B$, where the within-imputation variance $W = \frac{1}{m} \sum_{j=1}^{m} \text{var}(\theta_j)$ and the between-imputation variance $B = \frac{1}{m-1} \sum_{j=1}^{m} (\theta_j - \hat{\theta})^2$. To find a sufficient number of imputations, $m$ should be chosen to fulfil $m \geq 100^*$ the percentage of incomplete cases in the data (134, 135).
It is possible to use multiple imputation with different kinds of variables, i.e. continuous, binary and categorical variables. Categorical variables can be modelled using a multinomial logistic regression. When choosing which variables to use in the imputation model, it is important that all variables, explanatory and outcome, are included in the same form as they appear in the final analysis model. If the analysis is based on a survival model, the outcome variables to include are a censoring indicator and some function of follow-up time. An alternative to using the follow-up time is to use the Nelson-Aalen cumulative hazard estimate (136, 137).

In all papers, multiple imputation was used to impute missing values on the variables education, income, EOD, histology and smoking status. In addition, laterality, tumour size and resection procedure were imputed in Paper I, while symptom status was imputed in Paper II. All of these variables were categorical variables, and therefore, multinomial logistic regressions were performed using the actual variables from the analyses model as explanatory variables. Unfortunately, there is no formal way to test the underlying assumption of data being MAR, and the best approach is to condition the variables with missing values on the available variables with a known or plausibly important association (138). In the data, there was no reason to suspect that there was any association between the missing structure of the variables and their true values or with any other variables that was not adjusted for in the imputation model.
Main results

Lung cancer survival in Norway (Paper I)

From 1997 to 2011, the 1- and 5-year relative survival for all lung cancer patients increased from 35.4% to 47.7% and 11.6% to 17.5%, respectively. The 1- and 5-year relative survival among resected patients increased from 77.2% to 93.3% and 47.0% to 62.2%, respectively. The corresponding figures for non-resected patients were 28.4% to 37.0% and 3.6% to 6.3%, respectively. The HRs for death among resected and non-resected patients up to two years after resection/diagnosis showed that these were consistently lower in 2004–2011 compared to 1997–2003, varying from 0.50 (95% CI 0.37–0.68) to 0.71 (95% CI 0.62–0.82) for resected patients and 0.75 (95% CI 0.70–0.81) to 0.88 (95% CI 0.83–0.94) for non-resected patients. The largest improvements in survival occurred among resected, as well as, adenocarcinoma patients, while patients aged 80 years or older experienced the smallest increase.

There were several factors affecting the improvement in survival during the study period from 1997 to 2011. The waiting time (median number of days) from diagnosis to resection increased from 26 to 34 days. Diagnostic improvements were also made during this period, which can be seen by the fact that more patients received their diagnosis based not only on a histological examination but also with additional examinations. Hence, the proportion of patients receiving diagnosis based only on histological examination decreased from 68% in 1997 to 30% in 2011. The proportion of molecular genetic examinations (including EGFR testing) increased from being less than 1% up until 2009 to 26% in 2011.
Lung cancer treatment (Paper II)

The resection rate among lung cancer patients in Norway remained fairly constant around 18% while the proportion of radical radiotherapy administered increased from 8.6% to 14.1% in 2002–2011. The proportion of patients not receiving either surgery or radiotherapy decreased from 50.0% to 38.7% in the same period.

Older patients (i.e. patients aged 80+), patients with low household income, and patients from certain health trusts, were less likely to receive any treatment. Compared to patients with a high level of education, patients with a low level were found to have a lower odds of resection. Having a low level of education was identified as a negative predictor for receiving surgery. A smoking history was positively associated with both radical and palliative radiotherapy, while comorbidity and symptoms prior to diagnosis were independently associated with receiving palliative radiotherapy.

Resection in relation to survival (Paper III)

The existence of regional variation in survival among NSCLC patients in Norway was established in the study period (2002–2011) both among patients with localised and regional spread of disease. These differences between regions persisted after adjusting for case-mix which included information about whether or not patients underwent resection.

For patients with localised disease an increasing resection rate was associated with a monotone decrease in the relative excess risk of death, while among patients with regional disease a point was identified where a further increase in resection would not contribute to any survival benefit.
Discussion

Overview of results

The results confirmed that survival after a lung cancer diagnosis for all patients has improved over the last 15 years in Norway, with a larger improvement seen among resected than non-resected patients (Paper I). When investigating the predictors for receiving different lung cancer treatments, SES variables and place of residence were identified as independent predictive factors (Paper II). Differences in survival between geographical regions could not be explained by regional differences in resection rate (Paper III).

Methodological considerations

All studies have analysed risk by either analysing survival in relation to surgery and other prognostic factors, or the chance of receiving treatment.

Study design

Within epidemiological studies, the two main branches of studies are observational and experimental. Observational studies examine the individuals of interest without any intervention, only based on recording, classifying, counting and performing statistical analyses (18). These studies can further be divided into descriptive and analytical, where analytical studies usually measure the effect of different risk factors in relation to a specified outcome. Further, there are two types of analytical studies, i.e. case-control and cohort studies. A cohort is defined as a “group of individuals who are followed or traced over a period of time” (139). Starting from a given time, the relationship between different factors and the outcome of interest are explored for the cohort. There are two types of cohort studies, which differ in when the data are collected, i.e. prospectively and historically (140). All papers contributing to this thesis are historical cohort studies. This kind of study is
important in order to reveal possible differences in health services or in care given, and the results can be used to help decision makers change health policies. For example, the results of Paper II, identified groups of patients who were less likely to receive treatment, and therefore, efforts should be made to address the inequity. And in Paper III, an optimal resection rate was found where performing additional surgery has no further survival benefit, and this result may affect the recommended patterns of surgical care.

Validity

The validity of an epidemiological study is usually divided into internal and external validity (141). Internal validity means that the inference performed in the study is valid for the study subjects themselves. However, (mainly) three different types of biases can impair a study’s internal validity, namely, selection bias, information bias and confounding (discussed later). Since an unselected population-based lung cancer cohort was used, the results from the papers are considered valid and representative for the Norwegian population. External validity refers to the applicability of the results found within the study population to other populations, i.e. the generalisability of the results. External validity assumes internal validity, but also relies on comparability of characteristics between the study and target groups. The results that were obtained from these data regarding lung cancer patients in Norway can be considered externally valid and comparable to other populations that have similar demographics and a universal health care system, where treatment and care are equally available for everyone, independent of social factors.

Selection bias is a systematic error that may be a result of the procedure of selecting participants for a study. This bias will occur if the association between the exposure and the disease differs among patients that are and are not included. The mandatory reporting about all new cancer cases to the CRN prevents the data from suffering from selection bias. As all lung cancer patients registered in the
CRN were included in Papers I and II, these studies are very unlikely to be affected by selection bias. The results from the third paper are only valid for NSCLC patients, however, within this group there have not been any further selections. Another type of bias called information bias is related to measurement errors registered for the patients. That is, the covariates or outcome variables may be of different quality, and thus misclassification or measurement error will vary between the comparison groups. As the quality of the data in the CRN is considered to be high, the risk of information bias is considerably reduced. However, random errors can still occur and are related to typing errors and data processing. For example, if a coder at the CRN is in doubt of what is written on the clinical report regarding stage or histology, he will make a decision regarding the classification of this patient. However, it is unlikely that there are systematic errors. In all the papers, stage information has been adjusted for as a possible confounder. Due to the changes in the coding practice (described earlier) after 2008, our results for the period 2002–2011 would be influenced by information bias. However, all information regarding stage was considered as unknown after 2008 in order to minimise this kind of bias.

The definition of a confounder is a variable that is correlated to the dependent variable and causally linked to the outcome (18, 142). Confounding variables either falsely create an association that do not really exist, or hide an already existing relation between the groups being compared. In the data analysis process there are two ways of dealing with confounders: by stratification or by adjustment in a statistical regression model (139). In the papers both techniques have been used, e.g. in Paper I, the analyses were stratified by resection status and then adjusted for possible confounders, such as age and stage.

In Paper I, the effect of education level on survival for a cohort of lung cancer patients was examined. Survival depends on the stage of the disease, and therefore, performing a multivariable regression...
without stage, could be confounded if those with lower education level have more advanced stage at time of diagnosis. Therefore, in all the papers, the internal coding practice at the CRN which divides stage into localised, regional, and metastatic, was used. A stratified analysis may still be affected by confounding within the three strata, as it will only be able to control for confounding between, and not within the different strata. This within strata confounding is also referred to as residual confounding. Residual confounding may appear in any epidemiological study when a certain confounding variable is not sufficiently adjusted for. A way of testing if an observed effect is affected by residual confounding is to first perform a multivariable regression excluding the confounding variable of interest, and then to examine the change in the estimates when the confounding variable is included. If the estimate for the variable of interest changes towards a smaller effect, it indicates that confounding still exists within the variable that was adjusted for. For example, stage information (condensed TNM) was used in the regression model in Paper I. If instead information about the patients’ TNM, which is a more detailed grouping of stage, was available, the estimates for education could move closer to 1. This would indicate that the observed effect of education, when adjusted for stage categorised in broad groups, could be a result of residual confounding. However, when stage was included in the multivariable model, the estimates for intermediate and high education went from 0.92 (95% CI 0.89–0.95) and 0.87 (95% CI 0.83–0.92) to 0.91 (95% CI 0.88–0.94) and 0.85 (95% CI 0.81–0.89), respectively, indicating that the estimates were not likely to be a result of residual confounding.

Statistical methods

Multiple imputation was used in all three papers to handle the difficulties related to changes in coding practice on tumour stage in the CRN. Tumour stage was considered as missing for 100% of the patients from 2009 to 2011, as there was no way to re-code these patients similar to those coded before 2009. Can one really use multiple imputations when missing a whole year of information? To
validate the method used in the papers, sensitivity analyses based on observed, historical data from
2002 to 2008 were performed comparing tumour stage distribution in two different scenarios: 1) imputation on the patients that were actually missing tumour stage, and 2) imputation on data that came from sequentially deleting all information about tumour stage for entire yearly cohorts. The first scenario took the data as it is registered in the CRN and imputed the unknown stage information based on the other known variables of the patients. The second scenario was performed by re-coding all stage information for patients diagnosed in 2002 as missing. To obtain the imputed stage distribution for this cohort, stage information regarding patients diagnosed from 2003 to 2008 and other available variables for the 2002-cohort were used. The same procedure was then applied sequentially for all cohorts in years 2003 to 2008.
Figure 6: Comparing 1-year survival when dealing with missing stage information in two different ways. First, multiple imputation is used to impute stage data the way it has been registered at the Cancer Registry of Norway (method 1). Second, multiple imputation was used to impute stage when assuming that all information about stage was missing for entire years (method 2).

Since the outcome in Papers I and III were related to survival and prognosis, the survival estimates in these two scenarios were compared. The distribution and survival for the different tumour stages at different years are shown in Figures 6 and 7.
Figure 7: Comparing the stage distribution when dealing with missing stage information in two different ways. First, multiple imputation is used to impute stage data the way it has been registered at the Cancer Registry of Norway (method 1). Second, multiple imputation was used to impute stage when assuming that all information about stage was missing for entire years (method 2).

Figures 6 and 7 show that survival and the distribution for stage for the two scenarios were similar, except for the marginal difference in the stage-distributions in 2002 and 2008. This indicates that the imputed stage information is reliable, and that using multiple imputation techniques to deal with whole years of missing data seems reasonable.
Strengths and limitations of register-based studies

In this thesis the results of three Norwegian population-based studies, of which two analysed all lung cancer patients, and one was restricted to NSCLC patients, are presented. Since it is mandatory for Norwegian hospitals to report all new cancer cases to the CRN, and since the CRN receives information about emigration from the National Population Register, a negligible number of patients were lost to follow up, and the CRN data are considered close to complete. Hence, it was possible to analyse different aspects of an unselected cohort of lung cancer patients registered in a high quality register. Further, due to the unique personal identification number, linkage of high quality individual-level SES and comorbidity data was possible. The ability to link individual-level data for income and education is possibly the greatest strength of this study when compared to other studies that have use area-specific measurements as a proxy for a patient’s SES.

On the other hand, there are still limitations and areas where the registry can improve. First, the registry has limited cancer treatment information available. Additional clinical information would be useful in studies examining prognostic factors and survival. Two examples of useful information are surgery procedure and resection margin. Treatment procedure information would be important to know as it would be possible for researchers to evaluate the differences in survival by procedure in a population-based material. Information about the resection margin would enable researchers to not only examine if certain subgroups of patients have better resection outcomes than others, but also to evaluate the quality of resections performed in different regions. Second, stage information for a patient is important when determining what kind of treatment the patient should receive, as well as, his prognosis. At the CRN the tumour stage information is coded and classified in-house according to the condensed TNM system, i.e. localised, regional and metastatic spread of disease. However, the TNM system is the internationally agreed-upon standard in describing and categorising cancer stages and it is being maintained by the Union for International Cancer Control (143). Both national and
international guidelines use the TNM system to guide their recommendations. Further, it was recently reported that comparing stage at diagnosis in six high-income nations using national population-based cancer registry data was challenging due to the use of different measures of stage (144). Having information about cTNM and pTNM would make the epidemiological studies more comparable internationally and more interpretable clinically. Hence, it is a clear limitation of the CRN data that only condensed TNM is used instead of TNM.

Discussion of the results

An improvement in lung cancer survival was found in Norway, which was larger than what was observed in most other countries and registries worldwide (100). Also, the results showed that survival improved among both non-resected and resected patients, indicating that the explanation is multifactorial and that the improvement possibly started as a consequence of increased attention and focus on this group of patients. This included the establishment of the Norwegian Lung Cancer Group in 1987 who prepared, revised and later updated national guidelines, which are published by the Norwegian Directorate of Health (29, 79, 145-148). The historical changes that occurred can be divided into two groups; diagnosis- and treatment-related changes. Diagnosis-related changes include increased ad hoc screening (i.e. more widespread use of image diagnostics when lung cancer is suspected), improved immunohistochemistry, introduction of positron emission tomography (PET) scan and better methods for analysing molecular genetic factors. The start of multidisciplinary team meetings, the introduction and updated national guidelines, and the centralisation of examination and surgical treatment contributed to both a more precise diagnosis and improved treatment of patients. The introductions of SBRT and adjuvant chemotherapy are examples of treatment-related changes (149-151).
The results of Paper I showed that significantly fewer diagnoses were based on a plain histological examination at the end of the study period compared to earlier. This was offset by an increase in the use of more advanced methods like immunohistochemistry and molecular genetics as these became available and improved. With the advances in immunohistochemistry, the pathologists have been able to more accurately classify tumour cells, which is of paramount importance for choosing the correct treatment. In addition, an increased use of computed tomography (CT) scanning – leading to more ad hoc screening – and a more widespread use of PET scanning – revealing possible distant metastases – are both contributing to an earlier and more precise diagnosis (152). Molecular genetics is used for personalised medicine, which had its entry in lung cancer care in Norway around 2009 (153). Lung cancer patients are tested for an EGFR-gene mutation, and research has shown that advanced and recurrent NSCLC EGFR-positive patients that were treated with gefitinib instead of traditional chemotherapy experienced an extended progression-free survival, i.e. the median time from date of administration of treatment until the date of disease progression or death (151, 154, 155). The median progression-free survival times were reported to be 9.2 and 6.3 months among EGFR-positive patients receiving gefitinib and those receiving a combination of cisplatin and docetaxel, respectively (156). In addition, testing for the ALK-gene mutation can be used as another example of how molecular genetics testing has affected survival. A randomised study from 2013 showed that there was a significant reduced risk of progression or death among ALK-positive advanced lung cancer patients receiving crizotinib compared to chemotherapy as a second-line treatment (HR = 0.49 95% CI 0.37–0.64) (157). In addition, they showed that the progression-free survival time was 7.7 months in the crizotinib group and 3.0 months in the chemotherapy group.

It was also observed in Paper I that stage distribution was approximately constant over time. An increase in ad hoc screening was observed during this period in Norway, and this may have led to an earlier diagnosis and hence a shift towards a distribution with more patients with localised stage
disease (158). However, this has probably been offset by an increase in, and better examination of possible metastases.

There was gradual centralisation of lung cancer treatment in Norway during the 2000s. While, less than 20% of the resected patients received surgery at a high-volume hospital in 2003, more than 80% did in 2011 (23). After the study period, this trend continued and 87% were resected at a high-volume hospital in 2013. This centralisation, in addition to the introduction of multidisciplinary team meetings and more diagnostic work-up, may have contributed to the increase in waiting time for surgery (149). The date of diagnosis registered in the CRN is the first of the following: date on the clinical report, date of first histological verification, or date of death. It can be argued that the increased time to surgery is affected by the extended diagnostics examination period. This is likely as more diagnostic work was done in the last years of the study period. Hence, one cannot conclude that the true waiting time for surgery has increased. A study from 2006 reported that the increased waiting time for resection was a consequence of better diagnostics (159). Despite the increased time to surgery, survival has still improved significantly. This can be seen as an indication that treatment and/or the selection of patients for surgery have become better. It was indicated in a Norwegian lung cancer study from 2007 that patients resected at a high-volume hospital (>19 resections/year) may have a lower 30-day post-operative mortality (OR: 0.76, p-value = 0.076), even if the result did not achieve statistical significance (9).

It can by hypothesised that centralisation of lung cancer care may lead to an increased case load for the clinicians and surgeons, which would provide addition experience and further improve the resection quality, as well as, the anaesthetic and pre-, and post-operative care. The introductions of cisplatin-based adjuvant chemotherapy in 2005, and SBRT in 2008 were the two major treatment-specific changes that happened in the 2000s (151, 160-164). Both are likely to have contributed to
the observed improved survival as other studies have shown significant effects of these interventions for selected groups (165).

The improvement in survival has been explained by advances in diagnostics and treatment, however, some caution is required when interpreting these results. Due to the development of new diagnostic methods, patients at the end of the study period may have been diagnosed earlier compared to patients diagnosed at the beginning of the period. This difference between the times of diagnoses is called a lead time, and the artificial improvement one may observe in survival, if the patients die at the same point in time, is called lead time bias. In Paper I, it is possible that lead time bias was introduced and that it may have influenced the improvement in survival. While it cannot be ruled out that some of the improvement observed was a result of lead time bias, it is unlikely that it can explain all improvement as there have been several treatment-related changes that have positively affected survival. One could question whether the observed improvement in survival could be the result of an earlier diagnosis, causing previously undetected cancers to be categorised as localised, previously defined localised patients with the most advanced spread of disease to be categorised as regional, and the previously defined regional patients with the most advanced spread to be categorised as metastatic. This would lead to an improved stage-specific survival, as patients with the worst prognosis shift to a more advanced group, where they would be the patients with the best prognosis. However, there would not be an overall observed improvement for the whole group. This is known as the Will Rogers phenomenon (166). In Paper I, since an improvement of survival was observed for the whole group of lung cancer patients, the results seem valid (166). The increasing difference between incidence and mortality among both men and women further supports that there has been a true improvement in survival (6).
Paper II found that patients classified with high education and high income were around 30% and 60% more likely to receive surgery, respectively, compared to patients in the corresponding low categories. Patients with high income were around 50% more likely to receive radiotherapy compared to low income patients. A possible reason for this difference is that patients with high education and income may be more health literate and able to seek information individually compared to those in the low education and income groups. This may make them more critical to the information they are given and they may be better able to discuss their options with their doctor. It can be further hypothesised that this group of patients may have higher self-confidence, stronger assertiveness, and interest in their own healthcare. These can all contribute to the way they interact with the health system, so that in situations where the doctor is ambivalent, the patient can push for a resection. Consequently, unnecessary medical care may be used on patients where the benefit might be minimal, i.e. there is no survival benefit from the surgery. However, the results from Paper I showed that survival was significantly better among patients with high compared to low SES in the same period. This superior prognosis was observed among resected and non-resected patients, and was thus an indication that the suspicion about a high rate of futile resections among high SES groups was less likely.

Another important consideration is the lack of information on which patients are active workers and which are pensioners at the time of diagnosis. This information could be important as there may be patients with high income that were grouped as low income because they were registered with their pension income the year before diagnosis. This can cause the observed effect for income in relation to receiving surgical treatment to be biased, i.e. one might observe a weakened effect (closer to 1). A way to adjust for this possible issue would be to include the patient’s annual income in a time period prior to diagnosis, or to have precise information about which patients are pensioners and which are workers. In Norway, the earliest age when a person can retire and receive his pension is 62 years, but
there are many people who choose to keep working and do not retire until they reach the age of 70, even if the normal pension age in Norway is 67 years. To examine the potential problem of having a combination of workers and pensioners in the same income group, a stratified analysis was performed on the following three age bands: patients under 63 years (the majority of patients assumed to be workers), 63–69 years (likely to be a mixture of pensioners and workers) and 70+ (the majority of patients assumed to be pensioners). For patients under 63 years, odds ratios for receiving surgery were 1.27 (95% CI 1.03–1.55) and 1.41 (95% CI 1.13–1.78), comparing intermediate and high income against low income, respectively. The comparable numbers for 63–69 years were 1.56 (95% CI 1.26–1.94) and 1.82 (95 CI 1.42–2.33) and for patients aged 70+ years the comparable numbers were 1.20 (95% CI 1.07–1.35) and 1.25 (95% CI 1.02–1.52). These results show a clear effect of having high income, however, it seems that the effect is smaller in the 70+ years group compared to those under 63 years. This indicates that as patients get older, the positive effect of having a high income fades. Since the effect of having high income is larger in the mixed group compared to the other two groups, the heterogeneity of patients aged 63–69 does not seem to be problematic. Hence the reported results appeared to be valid and not significantly biased by the misclassification of income among pensioners.

Paper II showed that there were significant differences in the likelihood of receiving surgery and radiotherapy based on where the patient lives. It was observed that patients living in a health trust that had a reduced chance of receiving surgery had an increased chance of getting radical radiotherapy. Hence, it seems like doctors and institutions in different health trusts have different views on what they consider to be the best pattern of care for the patient. International literature has also found that large within-country variations in the likelihood of getting treatment exist (29, 43, 46, 87, 167). In Paper III the relationship between survival and resection was explored further, and the results found significant geographical variation in survival. This is similar to what was found in a
Swedish and English study, but they also found that survival differences disappeared after adjusting for whether the patient received treatment or not (43, 88). In this study, the variation in survival could not be explained by the differences in resection rates between health trusts. The English study reported that the proportion of patients undergoing surgery and their survival depended on patient- (e.g. patient’s decision making preferences and processes), physician- (e.g. propensity to operate on a patient where the appropriateness of resection is uncertain), and institutional factors (e.g. availability of specialist thoracic surgical expertise). Since information about the physician and institutional factors is not available for this study, one cannot rule out the possibility that the observed trends reflect differences in pattern of care between the regions. Hence, some hospitals and regions may have better quality of care and better surgical practice for all lung cancer patients. The results of this study were adjusted for case-mix to rule out the possibility that differences in the cohort of patients could explain the results.

In the third paper where the existence of an optimal resection rate was analysed, it was established that among patients with localised disease, a higher resection rate were associated with better survival. However, among patients with regional spread, there was an upper limit for when performing more resections seemed futile. Using this information it appears that some hospitals may be more aggressive in performing resections on patients for whom the benefit is questionable. In practice, stage III-patients are divided into two groups where those diagnosed with clinical stage IIIA are rarely considered candidates for surgery and those diagnosed with stage IIIB are not considered at all. Paper III reported that 51.9% and 27.2% of resected patients in regions with high (>52%) and low (≤52%) resection rates, respectively, were diagnosed with pathological stage IIIA. This shows that an aggressive selection of patients for surgery occurred in regions with high resection rates.
Other possibilities that may explain regional variation in survival are the changes in lung cancer care that have occurred at different times in the different regions of Norway. One study that examined the utilisation and effectiveness of third-generation chemotherapy among advanced NSCLC patients in Norway observed significant survival improvement comparing the period before and after introduction of vinorelbine, and the study noted that there were substantial geographical variation in the uptake of chemotherapy use (168). While this study was conducted using data from 1994 to 2005, one can suspect that this time lag between regions can also be applied to more recent medical advances.
Conclusion

Paper I:
Between 1997 and 2011, the overall survival for both resected and non-resected lung cancer patients in Norway has improved. The reason for this is multifactorial and contributing factors include an increased attention on the group of lung cancer patients in general, more accurate diagnosis tools and improvements in treatment.

Paper II:
The overall resection rate for all lung cancer patients remained approximately constant at around 18%, while radical radiotherapy treatment increased from 8.6% in 2002 to 14.1% in 2011. Although Norway is an egalitarian country with a GINI index (i.e. a measure of the extent to which the distribution of income among individuals or households within an economy deviates from a perfectly equal distribution) in the upper quartile and has a free, universal healthcare system, lung cancer patients with low SES and increasing age are less likely to receive any treatment (169). There were also regional differences in the likelihood of receiving treatment. These results pose the question whether equal health care is provided, independent of place of residence, age and socioeconomic status.

Paper III:
The survival for non-small cell lung cancer patients differs significantly between the different regions in Norway during the period from 2002 to 2011, both among patients with localised and regional spread of disease. Neither case-mix nor differences in resection rates could explain these survival differences. For patients with localised disease, an optimal resection rate was not apparent, but for patients with regional spread of disease, an inflection point in the resection rate was found, beyond which there was no improvement in survival.
Future perspectives

Even though its survival has improved in recent years, lung cancer is still one of the most lethal cancer types. Around half of the diagnosed patients have metastatic spread at diagnosis and only 20% of all patients are resected. The focus in the following years should mainly (continue to) be on primary prevention. This can be done through more and/or better restrictions on tobacco smoking and stronger encouragement with better accompanying help for current smokers to quit smoking.

Screening is a secondary prevention for lung cancer and it has been analysed in a number of international studies. A systematic review showed that there was no reduction in the number of deaths from lung cancer after using chest radiography for screening (170). A randomised controlled trial from the US showed a significant relative reduction of 20.0% and 6.7% in lung cancer mortality and all-cause mortality, respectively, however, the rate of false positives was above 96% (171). It will be interesting to follow the discussion and the progress in the field of lung cancer screening.

The results of this study identified that there are differences in how likely patients are to receive treatment depending on their SES. It should be an aim to equate these groups by adjusting the treatment aggressiveness among patients in the low SES group. The greater likelihood of treatment in high SES patients was observed when focusing on surgery and radiotherapy, but the results of a study examining predictors for receiving chemotherapy among advanced stage lung cancer patients would be valuable to gain a complete picture of the treatment pattern in Norway.

Due to its gradual introduction in Norway after 2008, the proportion of patients in this study who received SBRT was low. Since the clinical register for lung cancer collects more specific stage information (cTNM), it will be interesting to analyse and compare lung cancer survival after the
introduction of SBRT, and to be able to examine surgery in more detail at different tumour-stages using a national population-based material.

Through the work of this thesis it is apparent that while the data quality at the CRN is considered to be high, there is room for significant improvement, especially when it comes to clinical variables such as stage. The continuity of how this is registered and keeping the level of missing data as low as possible should be a clear focus. With the establishment of the National Clinical Register for Lung Cancer in 2013, additional and relevant clinical information will become available and this data will be important in future epidemiological studies.
Errors in published papers

Paper I:
Table 2: The multivariable confidence intervals regarding the variable smoking for all patients and the resected patients should have been the following: Current: 1.18 (95% CI 1.09–1.28) instead of 1.18 (95% CI 0.94–1.03) and Former: 1.06 (95% CI 0.98–1.16) instead of 1.06 (95% CI 0.97–1.13) for all diagnosed patients. For resected patients the multivariable estimates should have been Current: 1.25 (95% CI 0.87–1.80) instead of 1.25 (95% CI 0.87–1.01) and Former: 1.07 (95% CI 0.73–1.57) instead of 1.07 (95% CI 0.73–1.26).

Paper II:
Table 2: The last category of the variable Comorbidity was supposed to be CCI and not PRI as published.
References


111. Statistics Norway. Dette er Statistisk sentralbyrå - statistikk og analyser til nytte for samfunnet [This is Statistics Norway - statistics and analysis for the benefit of the society]. Oslo: Statistics Norway; 2014.


