



UiT The Arctic University of Norway

[Faculty of Health Sciences

Department of Community Medicine]

[Aspects of Lung Cancer by sex]

[Merethe Selnes Hansen]

A dissertation for the degree of Philosophiae Doctor...[June 2020]

UiT Norges arktiske universitet
[Course title]

Acknowledgements

The work presented in this thesis was carried out at the Institute of Community Medicine, Faculty of Health Sciences, UiT - The Arctic University of Norway. My PhD-period has been funded by an independent research grant from the Northern Norway Regional Health Authority (Helse Nord).

I am sincerely grateful to my main supervisor, Professor Inger Torhild Gram, for your vast knowledge of science, and your dedicated supervision during this research project. Thank you for your encouragements, critics and comments due to which I have been able to accomplish my work.

Associate Professor Tonje Braaten has been my excellent co-supervisor. Thank you for your constructive ideas, statistical expertise, and your support and calm when I felt my whole work was a “sinking ship”.

I am thankful to Ildir Licaj, for your great patience when trying to teach me statistics. Thank you for giving valuable input, suggesting and sharing your statistical and epidemiological expertise. Your help and encouragement have been of great value. Thank you also for all non-scientific topics that have been enlightened.

I would also like to thank Professor Arnulf Langhammer, Professor Loic Le Marchand and Professor Emeritus Eiliv Lund for your valuable input, which always helped in improving the manuscript.

Being part of the Systems epidemiological group has been educational and rewarding. Thanks to all the other colleagues at ISM.

Thank to Trudy Perdrix-Thoma for invaluable help during preparation of the manuscripts and the thesis.

To my colleagues working at Division of Pulmonary Diseases at UNN. I am grateful that you made it possible to combine research and hospital work during the same period.

Dear friends and family. Thank you for your support.

Finally, I want to express my gratitude to my wife Berit for always being there for me. To my son Snorre, and to my mother for unconditional support.

Merethe

Table of Contents

Acknowledgements	1
Summary	4
List of papers	6
Abbreviations	7
1 Introduction	8
1.1 Lung cancer incidence, prevalence, and mortality	8
1.2 Smoking and risk of lung cancer	9
1.2.1 The four-stage model of the smoking epidemic	10
1.2.2 Smoking in Norway: a birth cohort perspective	12
1.2.3 Sex-specific cigarette consumption	13
1.2.4 Passive smoking	13
1.3 Other risk factors for lung cancer	14
1.3.1 Socioeconomic status	14
1.3.2 Diet	15
1.3.3 Alcohol consumption	15
1.3.4 Physical activity	15
1.3.5 Body mass index	16
1.4 Sex and lung cancer	16
1.5 Histology	17
2 Aims of the thesis	18
3 Materials and methods	19
3.1 Study population	19
3.1.1 The Norwegian Counties Study	19
3.1.2 The 40 years cohort	20
3.1.3 Cohort of Norway	20
3.1.4 The Norwegian Women and Cancer study	20
3.2 Exposure information in the Norwegian Counties Study, the 40 years cohort, and the Cohort of Norway	23
3.3 Exposure information in the Norwegian Women and Cancer study	24
3.4 Follow-up and endpoints	24
3.5 Statistical analysis	25
3.6 Ethical aspects	28
4 Results - summary of papers	29
4.1 Paper 1 - Sex differences in risk of smoking-associated lung cancer: Results from a cohort of 600,000 Norwegians	29

4.2	Paper 2 - Smoking related lung cancer mortality by education and sex in Norway.....	30
4.3	The fraction of lung cancer attributable to smoking in the Norwegian Women and Cancer study	31
5	Discussion of methodological considerations	32
5.1	Study design	32
5.2	Validity (external and internal)	33
5.2.1	Internal validity	33
5.2.2	Selection bias.....	33
5.2.3	Information bias.....	36
5.2.4	Information bias according to measures of smoking exposure in the NHSS	36
5.2.5	Information bias according smoking exposure in NOWAC.....	39
5.2.6	Information bias according to outcomes: lung cancer and lung cancer mortality	40
5.2.7	Confounding in the Norwegian Health Screening Surveys and the Norwegian Women and Cancer study	41
5.2.8	Residual confounding in the Norwegian Health Screening Surveys and the Norwegian Women and Cancer study.....	42
5.3	Information about education in the Norwegian Health Screening Surveys and the Norwegian Women and Cancer study.....	44
5.4	Time variable in the model.....	45
6	Discussion of main results.....	45
6.1	Sex differences in smoking-associated risk of lung cancer and lung cancer death	46
6.2	Social inequalities in smoking-associated lung cancer mortality	49
6.3	Active smoking and risk of lung cancer in women	50
6.4	Passive smoking and risk of lung cancer in women.....	51
7	Conclusions	53
7.1	Paper 1.....	53
7.2	Paper 2.....	53
7.3	Paper 3.....	53
	References	54

Summary

According to the 2014 World Cancer Report, lung cancer is the most common cancer in men and the third most frequent in women across the globe, and the leading cause of cancer death worldwide. In Norway, lung cancer was the second and third most frequent cancer among men and women, respectively in 2013, as well as being first in the country terms of cancer mortality in both sexes. The rising worldwide incidence of lung cancer among women has raised the possibility of a sex difference in the association between smoking and lung cancer.

It has been estimated that approximately 90% of all lung cancer is related to smoking. Other known and potential risk factors for lung cancer include exposure to passive smoke, and occupational and environmental carcinogens. Trends in lung cancer incidence among men and women tend to follow changes in cigarette smoking trends from decades earlier. The smoking prevalence among women peaked almost 20 years later than that in men in most countries, and men have consumed more than 70% of the cigarettes smoked in Norway.

The aims of this thesis were to study the association between smoking and risk of lung cancer (Paper 1) and lung cancer death (Paper 2) by sex, as well as to determine whether the association between smoking and lung cancer mortality was similar across education levels for men and women (Paper 2). In Paper 3, we wanted to examine active and passive smoking and risk of lung cancer in a national random sample of Norwegian women and estimate the population attributable fraction (PAF). We utilized the Norwegian Health Screening Surveys, a pooled cohort of three studies that included almost 600,000 Norwegian men and women, in Papers 1 and 2. In Paper 3, we used data from 142,508 participants of the Norwegian Women and Cancer study.

When we analyzed measures of smoking exposure in categorical groups, we detected no difference in the risk of lung cancer or lung cancer death between men and women. When we analyzed categories of measures of smoking exposures as continuous variables for current smokers, females had a significantly higher risk of lung cancer (Paper 1) and lung cancer death (Paper 2) than males. Among women, low education level was associated with an increased risk of lung cancer death (Paper 2). In Paper 3, we found that current and former smokers had a 14- and 4-fold increased risk of lung cancer, respectively, when compared with never smokers. Passive smokers had a 1.3-fold, non-significant, increased risk of lung cancer when compared with never smokers. Furthermore, we found that eight in 10 lung cancer cases among women could have been avoided if the women had never smoked.

List of papers

Paper 1

Hansen MS, Licaj I, Braaten T, Langhammer A, Le Marchand L, Gram IT.

Sex differences in risk of smoking-associated lung cancer: Results from a cohort of 600,000 Norwegians.

Am J Epidemiol 2018 May 1;187(5):971-981

Paper 2

Hansen MS, Licaj I, Braaten T, Langhammer A, Le Marchand L, Gram IT.

Smoking related lung cancer mortality by education and sex in Norway.

BMC Cancer 2019 Nov 21;19(1):1132.

Paper 3

Hansen MS, Licaj I, Braaten T, Lund E, Gram IT.

The fraction of lung cancer attributable to smoking in The Norwegian Women and Cancer (NOWAC) study.

Br J Cancer (submitted).

Abbreviations

BMI – body mass index

CI – confidence interval

CONOR – Cohort of Norway

EPIC – European Prospective Investigation into Cancer and Nutrition

HR – hazard ratio

IARC – International Agency for Research on Cancer

ICD – International Classification of Diseases

NHSS – Norwegian Health Screening Surveys

NOWAC study – Norwegian Women and Cancer study

PAF – population attributable fraction

RR – relative risk

1 Introduction

Cigarette smoking is by far the number one risk factor for lung cancer (1). Trends in lung cancer incidence among men and women tend to follow changes in cigarette smoking trends from decades earlier. In most countries, smoking prevalence among women peaked almost 20 years later than it did in men (2). The rising worldwide incidence of lung cancer among women has raised the possibility of a sex difference in the association between smoking and lung cancer.

1.1 Lung cancer incidence, prevalence, and mortality

According to the 2014 World Cancer Report, lung cancer is the most common cancer in men and the third most frequent in women across the globe, and the leading cause of cancer death worldwide. Due to a high and rather stable fatality rate, trends for lung cancer mortality rates are similar to those for incidence rates. More people die of lung cancer than of colon, breast, and prostate cancer combined (1). In 2013, 1,555 lung cancer cases were diagnosed in men and 1,301 in women in Norway (3).

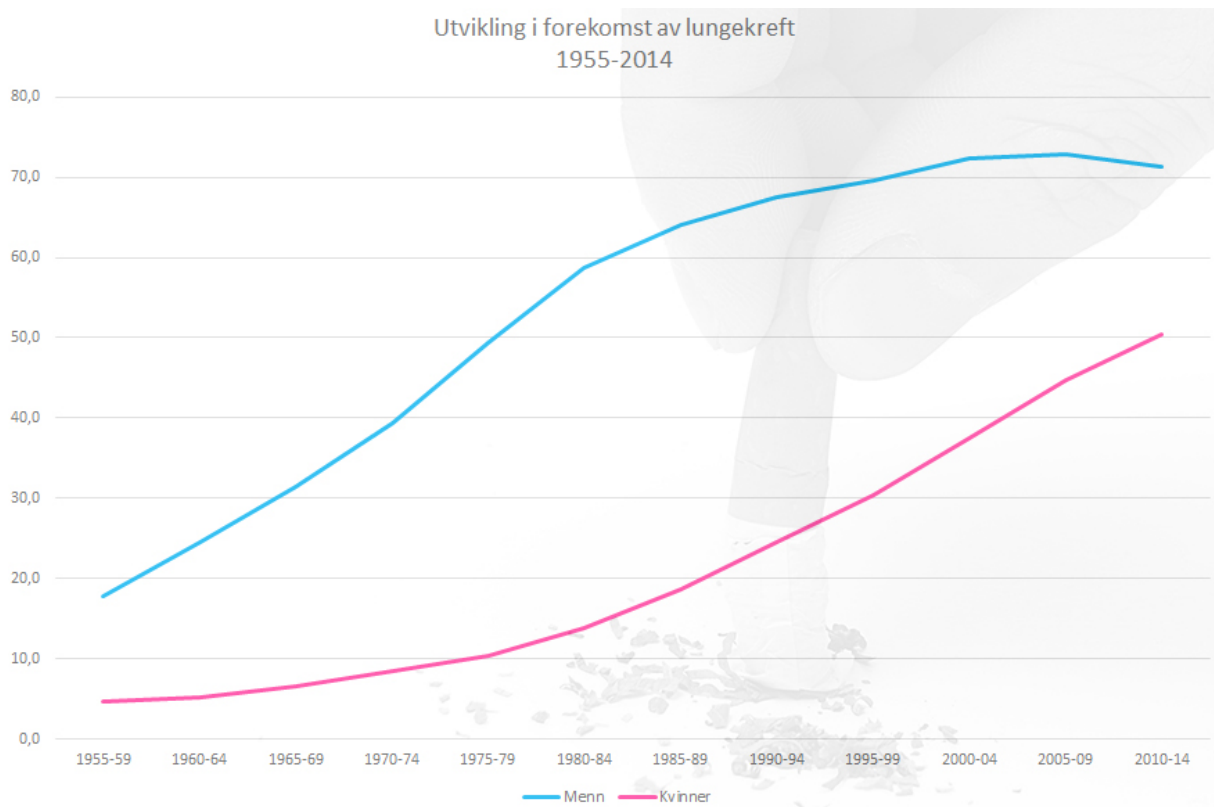


Figure 1: Lung cancer in Norwegian men and women 1955-2014. From the Norwegian Cancer Registry, 2016.

1.2 Smoking and risk of lung cancer

Lung cancer is one of the few cancer types for which the etiology is known in the majority of cases (1). In 1950, the landmark study by Richard Doll and colleagues investigated lung cancer incidence among British medical doctors. This study led the scientific community to categorize tobacco as a carcinogenic substance (4). The increased risk of lung cancer associated with smoking was first observed with convincing statistical proof in 1956, when The British Doctors' Study was published by Doll and Hill (5). It is estimated that approximately 90% of all lung cancer cases are related to smoking (6).

The risk of lung cancer increases with the number of years one has smoked (i.e., smoking duration), number of cigarettes smoked (i.e., smoking intensity), and early age at smoking

initiation (7). In 1978, Doll and Peto observed that smoking duration is much more important than smoking intensity in causing lung cancer (8). The close relationship between smoking and lung cancer can be observed when comparing historical lung cancer incidence and smoking prevalence, which follow a similarly shaped curve, with a time lag.

1.2.1 The four-stage model of the smoking epidemic

The rise and fall in smoking prevalence over time is remarkably similar across developed countries. Cigarette consumption was almost non-existent in the late 1800s, then increased exponentially and peaked around the middle of the 20th century. This increase unfolded much like the pattern produced by an epidemic, with four distinct stages (9).

In the first stage (1900-1920) was the increased popularity of cigarettes, particularly in higher social classes, which marked the start of mass smoking among men, while for women, smoking remained marginal. In the second stage (1920-1950), smoking prevalence rose sharply, with smokers representing more than half of men and around 20% of women by the end of the period. During the third stage (1950-1980), smoking among men started to decline, especially in groups with higher socioeconomic status, while smoking among women reached a maximum of approximately 40%. In the last stage, which is ongoing (1980-), smoking prevalence has decreased, and is currently at around 20% for both sexes. The four-stage model shows that the health consequences of smoking depend on smoking prevalence in the population, and illustrates the substantial time lag between smoking initiation and smoking-associated death, as the health consequences of smoking appear many decades after smoking cessation. Thus, the four-stage model predicts that smoking-associated health consequences for women will increase in the same way smoking prevalence did in this sex – i.e., consequences will occur later in women as they entered the smoking epidemic later than men

– and both smoking prevalence and smoking-associated mortality will peak at a lower level among women than men. These facts are crucial when studying the consequences of smoking among women, and when trying to compare health disparities between the sexes.

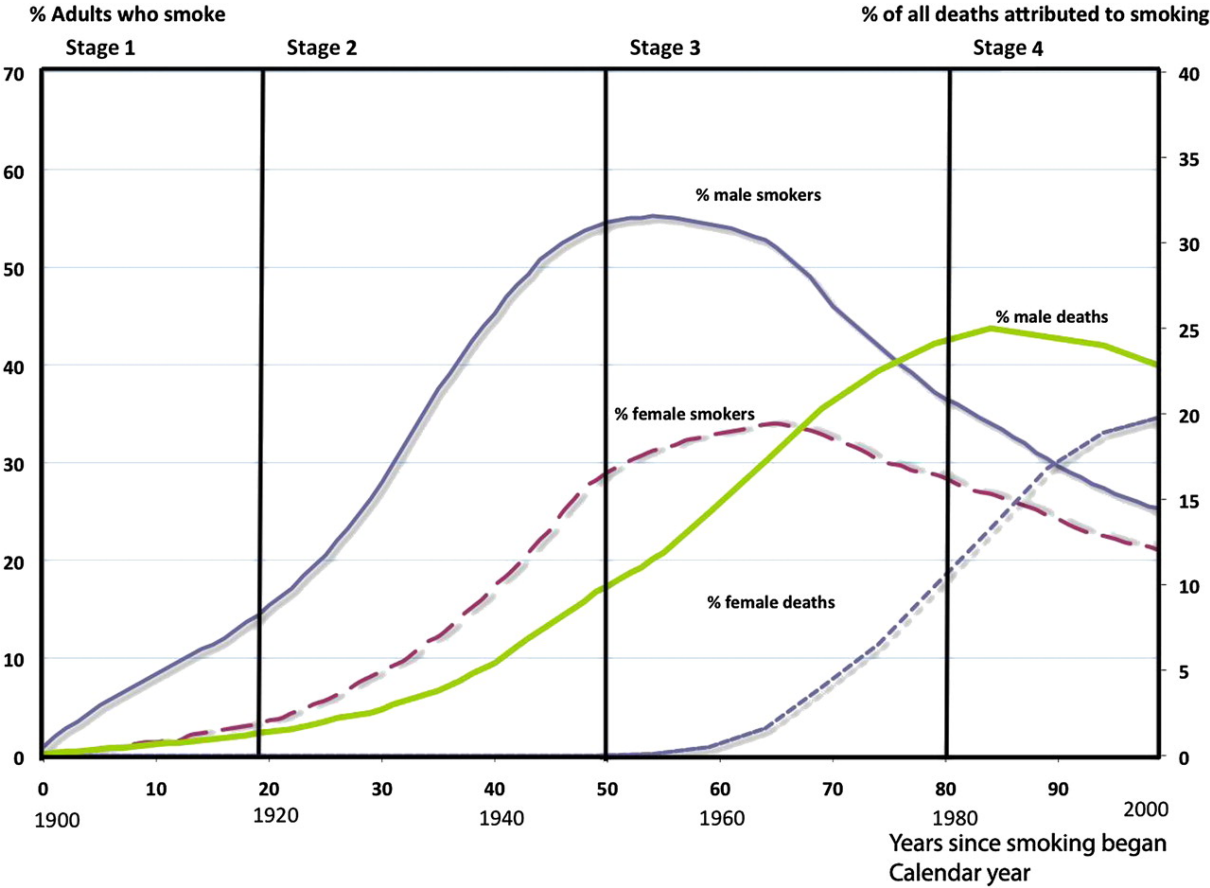


Figure 2. The four-stage model of the smoking epidemic. From Lopez et al. (1994). Reprinted with permission from Lopez.

1.2.2 Smoking in Norway: a birth cohort perspective

Over the last century, there has been large variability in the proportion of smokers within and across birth cohorts. Within male cohorts, the typical development has been a steady increase in the proportion of smokers from the youngest age groups to the age of 30 years, after which the situation tended to stabilize for a number of years, followed by a decrease in smoking prevalence in the older age groups. In the earliest birth cohorts, this decrease usually started around the age of 60 years, but in later cohorts it has tended to start earlier. The highest proportions of male ever smokers were observed in cohorts born between 1905 and 1935, with a peak prevalence above 70% occurring when the cohorts were between 20 and 45 years old. Among women, smoking prevalence has changed in a similar manner, but with a 20-year time lag and a lower prevalence. The highest smoking prevalence among women was 52% among 25-29-year-olds from cohorts born in 1940-1944, and 20-24-year-olds from cohorts born in 1945-1949. Even for female cohorts, there was a clear tendency for the proportion of smokers to increase in younger age groups and decrease in older age groups (2). In 2013, the percentage of daily smokers in Norway was 15% for both men and women (10).

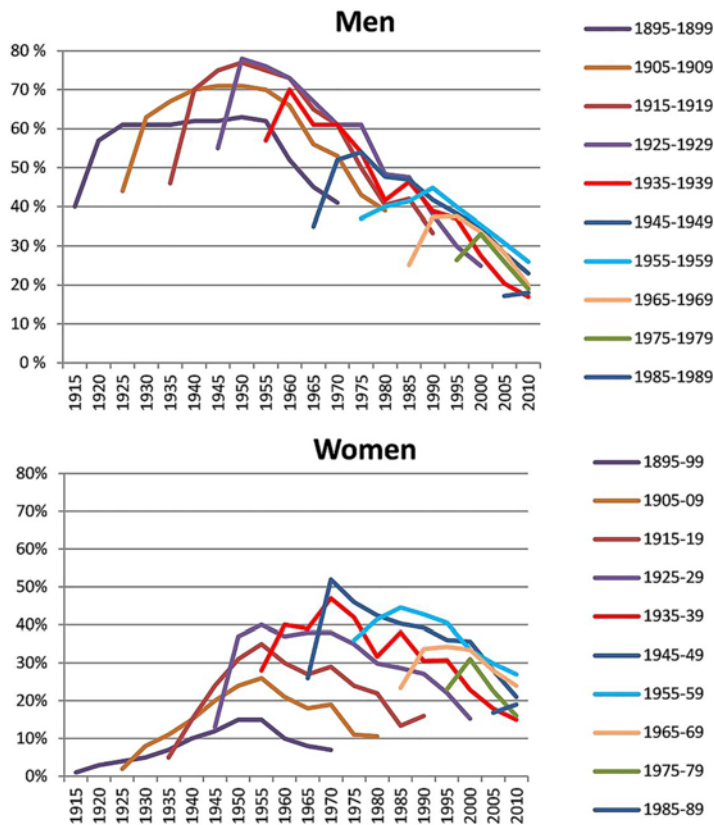


Figure 3: The proportion of current smokers in 1915-2010 among Norwegian men and women by birth cohort. From Lund I, et al (2014). Reprinted with permission from Lund.

1.2.3 Sex-specific cigarette consumption

The results of the combined effect of higher smoking prevalence and higher cigarette consumption indicate that, from an epidemiological perspective, men have had a longer and more intense smoking exposure than women (2).

1.2.4 Passive smoking

The inhalation of tobacco smoke by non-smokers has been referred to as “passive smoking” or “involuntary smoking”. The first epidemiological studies on passive smoking and the risk of lung cancer in non-smokers were published in 1981. In 1986, there was sufficient evidence to conclude that passive smoking causes lung cancer in non-smokers (1). The International

Agency For Research on Cancer (IARC) Monograph from 2004, the US Surgeon General's Report from 2006, and the World Cancer Report from 2014, state that approximately 20-30% of all lung cancer cases are related to passive smoking (1, 11, 12). However, data on the risk of lung cancer in passive smokers are limited. Among the 34 potential risk factors for lung cancer, including passive smoking, investigated among never smokers in the UK Million Women Study, only asthma, taller stature, and non-white vs. white ethnicity, were significantly associated with increased lung cancer incidence. There was little association with other sociodemographic, anthropometric, or hormonal factors, or with dietary intakes of meat, fish, fruit, vegetables, and fiber (13).

1.3 Other risk factors for lung cancer

Other known and potential risk factors for lung cancer include exposure to occupational and environmental carcinogens such as asbestos, radon, polycyclic aromatic hydrocarbons, and outdoor pollution (1, 14).

1.3.1 Socioeconomic status

Lung cancer is more likely to occur in poorer and less-educated populations, reflecting the increase in smoking prevalence that is observed in the presence of worse outcomes for indicators of socioeconomic status, like income, education, and occupation (15). The 2006 US Surgeon General's Report pointed out that exposure to secondhand smoke tends to be greater for persons with lower incomes than those with higher incomes. (12).

1.3.2 Diet

The 2014 World Cancer Report does not state anything about diet and the risk of lung cancer (1). A report from the European Prospective Investigation into Cancer and Nutrition (EPIC) study, which includes women from the Norwegian Women and Cancer (NOWAC) study, did not find that meat or fish consumption were significantly related to an increased risk of lung cancer (16). Another report from the EPIC study observed inverse associations between the consumption of vegetables and fruits and risk of lung cancer in never smokers, but a weak, significant association was found in current smokers (17).

1.3.3 Alcohol consumption

Alcohol consumption is known to increase the risk of certain cancers, but it is not mentioned as an independent risk factor for lung cancer in the 2014 World Cancer Report (1). A publication from the EPIC study reported a non-significant, increased risk of lung cancer among those with a high lifelong alcohol consumption. They also observed a lower risk of lung cancer among moderate consumers compared with light consumers of alcohol (18). A pooled analysis of data from seven prospective studies observed an increased risk of lung cancer among never smokers, and a slightly higher risk associated with a very high daily consumption of alcohol (19). However, no association between lung cancer mortality and any level of alcohol consumption was observed in men or women in a large cohort study of lifelong non-smokers (20).

1.3.4 Physical activity

The 2014 World Cancer Report does not discuss eventual associations between physical activity and lung cancer (1), and the 1996 Surgeon General's Report on physical activity did

not mention the risk of lung cancer (21). In 2002, the IARC concluded that the association between physical activity and risk of lung cancer was inconclusive (22). A study based on the EPIC cohort observed no inverse association between occupational or household physical activity and the risk of lung cancer, but saw some reduction in risk associated with sports (23). Results from a large prospective cohort study did not support any association between physical activity and the risk of lung cancer (24).

1.3.5 Body mass index

The 2014 World Cancer Report does not discuss eventual associations between body mass index (BMI) and lung cancer (1), and a report from the IARC concluded that the available studies are not of sufficient quality to permit a conclusion regarding the presence or absence of a preventive effect of the absence of excess body fatness on lung cancer (25). Results from the EPIC study showed a significant inverse association between BMI and the risk of lung cancer after adjustment for smoking and other confounders. Given the decline in the inverse association between BMI and lung cancer over time, the association is likely to be at least partly due to weight loss resulting from preclinical lung cancer that was present at baseline. They concluded that residual confounding from smoking could have influenced their results (26).

1.4 Sex and lung cancer

During the last decade in Norway, the prevalence of lung cancer in women has surpassed that of colon cancer. Lung cancer mortality has also surpassed breast cancer, and is now the most common cause of cancer death in women. For both men and women, lung cancer is now the second most common cancer type, as well as being the cancer type that causes the highest number of deaths (27). The question of whether women are more susceptible to smoking-

associated lung cancer has been raised as a result of the rising incidence of the disease in women globally. Neither the 2014 World Cancer Report nor the 2014 US Surgeon General's Report discussed a possible sex difference in the risk of smoking-associated lung cancer (1, 28). The 2004 IARC Monograph concluded that there was inconsistent and inadequate epidemiological evidence to support the proposal that women are more susceptible than men to developing lung cancer as a result of smoking (29). The 2012 IARC Monograph referred to cohort studies which observed that the risk of smoking-associated lung cancer was generally similar among men and women, but also referred to a meta-analysis that observed a slightly higher risk in women than in men (7).

1.5 Histology

Lung cancer can be divided into two broad categories: small-cell lung cancer, which accounts for 15% of lung cancer cases, and non-small cell lung cancer, which accounts for the remaining 85% of cases. Non-small cell lung cancer is further divided into three major histological subtypes: adenocarcinoma, squamous cell carcinoma, and large cell carcinoma (1). In the past decade, the incidence of adenocarcinoma has increased greatly, replacing adenocarcinoma as the most prevalent lung cancer subtype. Today adenocarcinoma is the most prevalent subtype for both men and women in Norway (30).

2 Aims of the thesis

The aims of this thesis were to study by sex the association between smoking and risk of lung cancer and lung cancer death. Moreover, this thesis examines whether the association between smoking and lung cancer mortality was similar for different education levels for men and women. In addition, we wanted to examine active and passive smoking and risk of lung cancer in a national random sample of Norwegian women, and to estimate the population attributable fraction (PAF). We utilized the Norwegian Health Screening Surveys (NHSS), a pooled cohort of three studies, in Papers 1 and 2. In Paper 3, we used data from the NOWAC study.

Thus, the specific aims were:

1. To examine if the association between smoking and the risk of lung cancer differed by sex.
2. To explore if smoking-associated lung cancer mortality differed by sex and education level.
3. To examine the association of active and passive smoking with risk of lung cancer, and calculate the PAF of lung cancer due to smoking.

3 Materials and methods

3.1 Study population

The study population in Papers 1 and 2 comprised Norwegian men and women born between 1899 and 1975, participating in three large prospective cohort studies conducted by the National Health Screening Service (now included in the Norwegian Institute of Public Health): the Norwegian Counties Study (inclusion period: 1974-1978), the 40 Years Study (inclusion period: 1985-1999), and the Cohort of Norway (CONOR, 1994-2003), hereafter referred to as the NHSS.

Invitees to the NHSS were selected based on year of birth and place of residence. They then received an invitation and baseline questionnaire. Trained nurses measured height and weight at the screening facility. The response rates varied between 56% and 88% in the NHSS (31).

3.1.1 The Norwegian Counties Study

The purpose of the Norwegian Counties Study was to investigate cardiovascular diseases. This survey was carried out in the three rural Norwegian counties (Finnmark, Sogn og Fjordane, and Oppland), and consisted of three rounds of health screenings carried out during the periods 1974-1978, 1977-1983, and 1985-1988. In the first round, all residents aged 35-49 years were invited, as well as a random sample of 10% of the general population aged 20-34 years. In the second and third rounds, a combination of previous participants and new cohorts were invited, and similar protocols and questionnaires were used. The participation rates were 88%, 88%, and 84% at the three screening rounds, respectively (31-33).

3.1.2 The 40 years cohort

The purpose of the 40 years cohort was to study cardiovascular diseases. This survey was carried out between 1985 and 1999, and included 420,000 Norwegian men and women from all counties of Norway. Mostly men and women aged 40-42 years were invited, though individuals aged 65-67 years were invited in some counties in the first of four phases of this study. The overall participation rate was 69% (31, 34, 35).

3.1.3 Cohort of Norway

The purpose of CONOR was to investigate causes of disease on a broad basis. In this survey, regional data from 10 epidemiological surveys conducted between 1994 and 2003 were merged into a national database. In all the CONOR surveys, standardized protocols, procedures, and questionnaires were used. The average response rate for the 10 epidemiological surveys included in CONOR was 56% (31, 36, 37). A further description of these 10 surveys is included in the appendix.

3.1.4 The Norwegian Women and Cancer study

In Paper 3, we used data from the NOWAC (kvinner og kreft) study. The NOWAC study is a national, population-based cohort study of women which was initiated in 1991 (38). The study was originally designed to investigate the association between oral contraceptive use and risk of breast cancer, and it has gradually expanded to cover other outcomes and risk factors.

NOWAC participants have been sampled randomly from the Norwegian Central Person Register (38). Participants born in 1927-1965 were enrolled in three main waves in a stepwise manner, mainly due to practical administrative reasons, limited financial support, and the

performance of methodological substudies. In this study, we used data only from questionnaires in the red boxes in figure 4. Figure 4 (red boxes) shows that the first wave of enrollment took place in 1991, the second wave in 1995-1997, and the third wave in 2003-2007.

A total of 179,388 women were invited to participate in the first and second wave of enrollment (1991, and 1995-1997), with an overall response rate of 57.5% (39). However, the response rate for the second wave of enrollment was 81% after correction for emigration and death. During the third wave of enrollment (2003-2007, Figure 4, red boxes), an additional 148,088 women were invited to participate, with a response rate of 48% after correction for emigration, death, and unknown addresses. After the exclusion of 29,970 women, our study sample consisted of 142,508 women.

The participants answered a detailed four-page questionnaire, which collected information on use of oral contraceptives and hormone replacement therapy, reproductive history, physical activity, alcohol consumption, anthropometric measures (height/weight), socioeconomic status, and smoking habits. Furthermore, a large proportion of the questionnaires contained a four-page food frequency questionnaire.

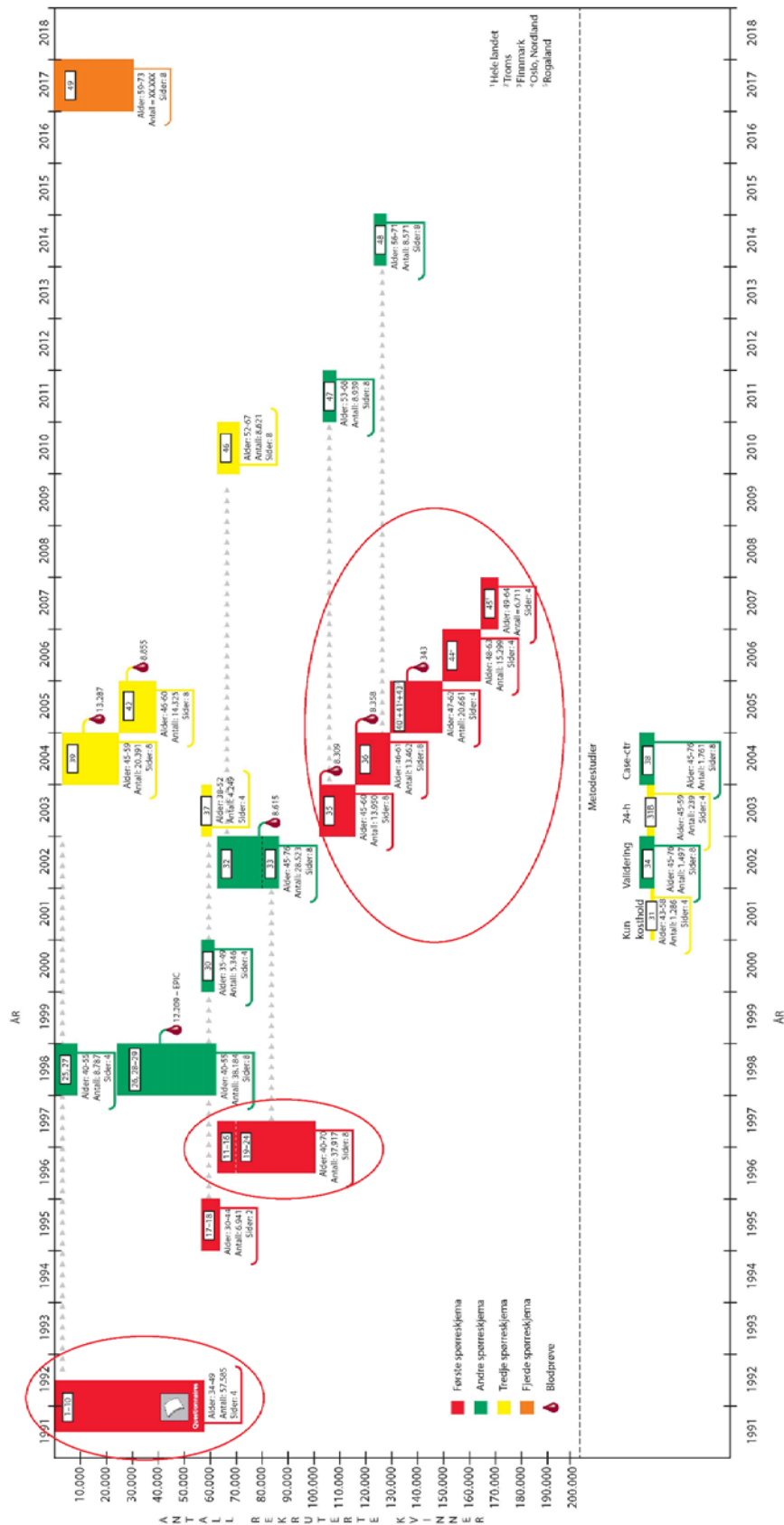


Figure 4: Enrollment in the Norwegian Women and Cancer Study

3.2 Exposure information in the Norwegian Counties Study, the 40 years cohort, and the Cohort of Norway

The NHSS had a baseline questionnaire, which collected detailed information on smoking habits, physical activity, and other lifestyle factors. The variables in CONOR were adequately structured; thus that questionnaire was used as reference for standardizing the other surveys' questionnaires. Questions on smoking habits were similar but not identical across all surveys, and questions asked about current and former daily smoking habits, smoking duration, average number of cigarettes smoked per day, and in a few surveys, former smokers were asked about time since smoking cessation. Only CONOR asked about age at smoking initiation. In the other surveys, this variable was estimated for both current (age at enrollment minus duration of smoking in years) and former (age at enrollment minus years since smoking cessation and duration of smoking) smokers. We also found common formats for other variables, such as alcohol consumption, which were available only in the latest surveys (the 40 years cohort III and IV and CONOR). After receiving specified exposure variables from the primary data of each survey, a standardized database for the pooled analysis was created. For participants who took part in more than one survey, only the earliest survey was included. Information about physical activity was obtained using a self-reported measure. The subjects were categorized into three groups based on level of physical activity reported at the time of enrollment: sedentary (reading, watching television, and sedentary activity), moderate (walking, bicycling, and/or similar activities ≥ 4 hours per week) and heavy (light sports or heavy gardening ≥ 4 hours per week, heavy exercise, or daily competitive sports). Education level is a reliable indicator of socioeconomic status because it is stable, established in early adulthood, and not modified by chronic disease (40). The most recent information from Statistics Norway on the number of completed years of education was used instead of the self-reported information in the questionnaires. We used the number of completed years of

education recorded in 1990 or 1980, and if this information was missing, we used data from 1970. Detailed information on how the files from the three surveys were merged into a single database is included in the appendix (with permission from R. Parajuli and E. Bjerkaas) (41, 42).

3.3 Exposure information in the Norwegian Women and Cancer study

All the different NOWAC questionnaires used in our study asked if the women had ever been smokers. Those answering “yes” were requested to fill in the number of cigarettes smoked per day at different age intervals, allowing us to calculate the average number of cigarettes smoked per day. From questionnaire 10 (Figure 4 red boxes) and later, women were asked if they presently smoked on a daily basis. We classified former smokers according to years since smoking cessation. We categorized women who had never smoked as passive smokers if they reported exposure to passive smoking as children or from their spouse as an adult. We categorized the remaining never smokers as never smokers. Some of the questionnaires asked about age at smoking initiation. We calculated average alcohol consumption in g/day among drinkers based on the content of pure alcohol in different beverages and usual portion sizes in Norway.

3.4 Follow-up and endpoints

Participants were followed through linkages to the Cancer Registry of Norway and the Central Population Register, using the unique 11-digit personal identification number, to identify all lung cancer cases, lung cancer deaths, emigrations, and other deaths. These national registries are both accurate and virtually complete (43). In Paper 1, we excluded 50,257 individuals: 647 who emigrated or died before the start of follow-up, 11,321 with prevalent cancer, 190 with missing information on vital status, 6,303 with missing measures of smoking exposure,

and 31,796 with other missing covariates that were included in the analyses, leaving 585,583 in the analytical cohort. In Paper 2, we excluded 40,091 individuals due to emigration or death before the start of follow-up, missing information on vital status, measures of smoking exposure, missing information on education, or missing information on any of the covariates included in the analysis, leaving 595,675 in the analytical cohort. In Paper 3, we excluded 29,970 individuals: 6,664 due to prevalent cancer, 74 who emigrated or died before start of follow-up, 32 with an age at exit below the age at recruitment, 590 with missing information on smoking exposure, 10,879 never smokers with missing information on passive smoking, and 11,731 with missing information on any of the covariates included in the analysis, leaving 142,508 in the analytical cohort. We classified lung cancer cases using the anatomical sites and histological codes in the International Classification of Diseases (ICD) for Oncology. We included all primary incident carcinomas of the trachea, bronchus, and lung (ICD codes 8, 9 and 10). We classified lung cancer into six histological subtypes (squamous cell, adenocarcinoma, large cell, other not specified non-small cell carcinoma, small-cell carcinoma and other carcinomas) (44). In Paper 2, we used the histological subtype registered at lung cancer diagnosis as the type that led to death if the patient died from lung cancer.

3.5 Statistical analysis

In Paper 1, we used a Cox proportional hazards model with attained age as the underlying time scale to estimate the multivariate-adjusted hazard ratios (HRs) with 95% confidence intervals (CIs) for the associations between different measures of smoking exposure, using both categorical variables and continuous variables, and lung cancer. The categorical smoking exposure variables we used were: smoking duration in years (1-9, 10-19, 20-29, ≥ 30), number of cigarettes smoked per day (1-10, 11-20, >20), pack-years (1-5, 6-15, ≥ 16) and age at smoking initiation (≥ 21 , 16-20, <16 years). The continuous variables were: smoking duration

in 10-year increments, tens of cigarettes smoked per day, tens of pack-years, and age at smoking initiation for former, current, and ever smokers. We evaluated the association between the categorical variable years since smoking cessation (0-4, 5-9, ≥ 10) and the continuous variable time since smoking cessation in 10-year increments and risk of lung cancer, for former smokers only. Never smokers were used as the reference group in all categorical smoking analyses, except for the association between years since smoking cessation and risk of lung cancer, where we used current smokers as the reference group. Never smokers were not included in analyses where smoking exposure was measured continuously. The *a priori*-selected covariates included in the final models were: physical activity level (sedentary, moderate, or heavy), BMI, and duration of education. We tested for linear trend across categories of measures of smoking exposure for former, current, and ever smokers, excluding never smokers, based on the median value in each smoking category, using the lowest category of each measure of smoking exposure as a reference. We used Cox models stratified by cohort study and birth cohort (≤ 1950 and >1950) to overcome any probable heterogeneity for these variables. We used the Wald test to assess heterogeneity by sex for the associations between different measures of smoking exposure and lung cancer.

In Paper 2, we used a Cox proportional hazards model with attained age as the underlying time scale to estimate the multivariate-adjusted HRs with the 95% CIs for the associations between different measures of smoking exposure using both categorical and continuous variables, and lung cancer mortality. The categorical smoking exposure variables used were: smoking duration in years (1-19, 20-29, >30), number of cigarettes smoked per day (1-10, 11-20, >20), and pack-years (1-9, 10-19, ≥ 20]. The continuous variables were: smoking duration in 10-year increments, tens of cigarettes smoked per day, and tens of pack-years, for current smokers. Never smokers were used as the reference group in all categorical smoking analyses, but they were not included in the continuous smoking analyses. *A priori*, we considered

alcohol, physical activity, BMI, and education level as possible confounders. We tested for interaction between smoking status and sex, and between smoking status and education level, and decided to stratify by sex and by education level. We decided to adjust on BMI and physical activity, but did not include alcohol as a covariate due to the large amount of missing data. We tested for linear trend across categories of measures of smoking exposure for current smokers, excluding never smokers, based on the median value in each smoking category, using the lowest category of each measure of smoking exposure as a reference. We used Cox models stratified by cohort study and birth cohort (≤ 1950 and >1950) to overcome any probable heterogeneity for these variables. We used the Wald test to assess heterogeneity by sex and by education level for the associations between different measures of smoking exposure and lung cancer mortality.

In Paper 3, we used a Cox proportional hazards model with age as the underlying time scale to estimate crude and multivariate-adjusted HRs and 95% CIs for the associations between lung cancer and measures of smoking exposure. The reference group was composed of never active and never passive smokers, unless otherwise noted. We included the covariates that changed the HR estimate in the crude model by at least 5%, that could potentially confound the association between smoking and lung cancer. We included the following variables in the final multivariate models: age at enrollment, duration of education (<10 , $10-12$, ≥ 13 years), and average alcohol consumption (0 , ≤ 4 , $5-9$, ≥ 10 g/day). We stratified all models by recruitment sub-cohort (1991-1992, 1996-1997 and 2003 -2007) to control for potential differences in recruitment waves. We tested for linear trend for smoking exposure (smoking duration, number of cigarettes smoked per day, pack-years, and age at smoking initiation) for ever smokers, based on the median value in each category, using the lowest category of each measure of smoking exposure as a reference, i.e., excluding never smokers. We tested for linear trend among former smokers and years since smoking cessation in three categories

based on the median value, using the lowest category of years since smoking cessation as a reference, i.e., excluding current smokers, which was the reference group. We estimated PAF (%) to indicate the proportion of lung cancer cases that could have been prevented in the female population in the absence of smoking.

In all 3 papers, we tested and found that the criteria for the proportional hazards assumption were met using Schoenfeld residuals. Analyses in all 3 papers were done in STATA, version 14.0 or 15.0 (StataCorp LP, College Station, Texas, USA). Two-sided p-values <0.05 were considered statistically significant.

3.6 Ethical aspects

For NHSS participants recruited before 1994, returning the completed questionnaire was considered sufficient as acceptance to participate. As from 1994, participants gave written informed consent to participate in the surveys. We also obtained approval from the respective steering committees to all the health surveys included. Our study was approved by the Regional Ethical Committee and the Norwegian Data Inspectorate. The women included in the NOWAC study were sent an invitation letter along with the questionnaire. The women were also informed about later linkages to the Cancer Registry of Norway and the Cause of Death Register. The Regional Ethical Committee and the Norwegian Data Inspectorate has approved the NOWAC study.

4 Results - summary of papers

4.1 Paper 1 - Sex differences in risk of smoking-associated lung cancer: Results from a cohort of 600,000 Norwegians

In this paper, we examined the association between smoking and the risk of lung cancer by sex. During nearly 12 million years of follow-up, 6,534 participants (43% women) were diagnosed with lung cancer. The age-standardized incidence rate of lung cancer in current smokers was more than 30-fold greater in men and more than 10-fold greater in women, compared with sex-specific never smokers. Compared with sex-specific never smokers, male and female current smokers who smoked for ≥ 16 pack-years had HRs for lung cancer of 27.24 (95% CI: 22.42-33.09) and 23.90 (95% CI: 20.57-27.76), respectively. For current smokers, in a model with pack-years measured continuously, men and women had HRs of 1.43 (95% CI: 1.39-1.48) and 1.64 (95% CI: 1.57-1.71), respectively, for each 10-pack-year increment of smoking (p for heterogeneity < 0.01). When we analyzed different measures of smoking exposure according to categorical groups (smoking duration, cigarettes smoked per day, and pack-years) and the risk of lung cancer in current and former smokers, we did not detect a difference between men and women for most of the measures of smoking exposure. Among current smokers, we observed a higher risk of lung cancer in men compared with women only in the three lowest categories (1-9, 10-19, 20-29 years) of smoking duration, and in the lowest category of pack-years (1-5 pack-years). For all the other categorical measures of smoking exposure in current smokers (≥ 30 years of smoking, all categories of cigarettes smoked per day, 1-5 pack-years, and 6-15 pack-years), we did not detect a difference in the risk of lung cancer in men compared with women. For former smokers, we did observe a sex difference, but only among those with a smoking duration of ≥ 30 years. When we analyzed the three different measures of smoking

exposure as continuous variables for current smokers, women had a significantly higher risk of lung cancer compared with men. We did not detect a sex difference when looking at measures of smoking exposure as continuous variables and the risk of lung cancer for former smokers.

4.2 Paper 2 - Smoking related lung cancer mortality by education and sex in Norway

The objectives of the study were to explore a potential heterogeneity in smoking-related lung cancer mortality by sex and education level. Compared with sex-specific never smokers, current smokers had a lung cancer mortality HR of 20.05 (95% CI: 16.25-24.74) for men, and 13.97 (95% CI: 11.98-16.29) for women (p for heterogeneity = 0.01). For each 10-year increase in smoking duration, women had a 65% higher HR (1.65, 95% CI: 1.54-1.78) and men a 36% higher HR (1.36, 95% CI: 1.28-1.44) for lung cancer mortality (p for heterogeneity <0.01). For female current smokers, there was a significant difference between those with <10 years (HR: 15.85, 95% CI: 12.32-20.38) compared with those with ≥ 13 years of education (HR: 9.41, 95% CI 6.49-13.68) (p for heterogeneity <0.01). For male current smokers, the HR did not vary for the different categories of smoking exposure when we compared those with the lowest and highest education level (all p for heterogeneity >0.05). Our results suggest that women have an increased risk of dying from lung cancer compared with men, given the same level of smoking exposure. Among women, low education level was associated with increased lung cancer mortality, compared with those a middle and high education level.

4.3 The fraction of lung cancer attributable to smoking in the Norwegian Women and Cancer study

In Paper 3, we examined the risk of lung cancer associated with active and passive smoking, and estimated the PAF of lung cancer due to smoking. During the more than 2.3 million person-years of observation and 15.9 years of follow-up, we identified 1,507 lung cancer cases in the Cancer Registry of Norway. Current smokers had a lung cancer HR that was almost 14-fold (HR: 13.88, 95% CI: 10.18-18.91), and former smokers a HR that was almost 4-fold that of never smokers (HR: 3.69, 95% CI: 2.68-5.09). Compared with never smokers, passive smokers had a 1.3-fold (HR 1.34, 95% CI: 0.89-20.01), non-significant increased risk of lung cancer. The PAF of lung cancer due to smoking was 85.3% (95% CI: 80.1-89.2). More than eight in ten lung cancer cases could have been prevented among women if all had been never smokers.

5 Discussion of methodological considerations

Epidemiological studies present considerable opportunities for errors, which could take place in any step of the research process. The conclusions drawn from any analysis could, therefore, be limited (45).

5.1 Study design

There are three basic types of observational study designs in which individuals are the units of observation: the cohort or prospective study, the case-control study, and the cross-sectional study. Our results are based on cohort studies. Cohort data, if unbiased, reflect the “real-life” cause-effect temporal sequence of events. This is the only necessary criteria to establish causality (46). A well-designed cohort study can give powerful results. In a cohort study, an outcome or disease-free study population is first established by the exposure or event of interest and followed in time until the disease or outcome of interest occurs. Because exposure is identified before the outcome, cohort studies have the structure to produce strong scientific evidence. A main disadvantage of cohort studies is the need for a large sample size and potentially long follow-up duration due to relatively rare outcomes such as cancer, which results in high costs. Also, if the data is based on a single measurement, as it is in our study, the researcher has no information about eventual changes in the main exposure (e.g., smoking), or covariates over the study period. This thesis is based on cohorts with large sample sizes. Given this opportunity, we wanted to examine potential sex differences in the exposure-disease association between smoking and risk of lung cancer/lung cancer mortality (Papers 1 and 2). In Paper 3, we estimated the risk of lung cancer in active and passive smokers, and the number of lung cancer cases that could have been avoided in the absence of smoking.

5.2 Validity (external and internal)

Validity is an expression of the degree to which a test or a study is capable of measuring what it is intended to measure, and consists of two parts: internal validity and external validity (46). External validity, or generalizability, applies to whether one can generalize beyond the source population to the target population (45). Our study sample from the NHSS included surveys that have been validated separately (35-37, 47). NHSS participants were drawn from the Norwegian population, based on year of birth and residence, and represent a selection of the Norwegian population, both rural and urban. NOWAC participants were selected randomly from the general female population of Norway according to age (38, 39). Therefore, the external validity has been secured to a large extent in the NHSS and the NOWAC study. Together with the relatively high response rates, this suggest that our results from the NHSS can be generalized to the Norwegian population and Western countries. Our results from the NOWAC study can be generalized to middle-aged women in Norway and Western countries.

5.2.1 Internal validity

Internal validity is the degree to which the results of an observation are correct for the particular group of people studied. Any bias is usually placed into one of three categories: selection bias, information bias, and confounding (45, 46).

5.2.2 Selection bias

Selection bias occurs when a systematic error in the recruitment of study subjects, exposed subjects, or unexposed subjects, results in a tendency to distort the association between the exposure and the outcome. Because participants (exposed and unexposed) in cohort studies are selected before the outcome occurs, differential selection according to disease status is

less likely to occur (46). However, selection bias could be caused by a “healthy volunteers effect”, as volunteers are often characterized as healthier than the general population (48). Both the NHSS and the NOWAC study have high numbers of ever smokers, reducing the concern that a large number of smokers did not attend the surveys.

NHSS participants represent a selection of the Norwegian population, both rural and urban, as the participants were selected based on age and county of residence. The participation rate was higher in the earliest surveys, ranging from 88% in the Norwegian Counties Study to 56% in CONOR. However, the overall participation rate in CONOR was influenced by the low participation rate in those aged ≤ 30 years (37). As the enrollment period in CONOR was 1994-2003, the low proportion of young participants likely does not disturb our results, since mean age at lung cancer diagnosis in our study was 64 years. Previous reports showed that individuals who choose to participate in research studies have either a high or a low education level (49). Recent studies have found an increasing over-representation of highly educated women as the age of study samples increases (39). In the HUNT study, non-participants had lower socioeconomic status and higher mortality than participants (50). Thus, as lung cancer is more common in men and women with a low education level, possible selection bias should be taken into consideration. For the NHSS, selection bias could lead to a reduced estimate of the effect of smoking on lung cancer. Furthermore, we had a similar proportion of male and female participants in the NHSS. A total of 50,257 (55% women) participants were excluded in Paper 1, and of the 11,321 excluded participants with prevalent cancer, 63% were women. In Paper 2, the proportion of excluded participants among men and women were quite similar to those in Paper 1. As our study had the same proportion of men and women excluded due to missing data, it is unlikely that our analytical sample suffers from severe selection bias according to the proportion of men and women.

Since its beginnings, the NOWAC study has been designed to create a representative, population-based, prospective study cohort. The response rate in the NOWAC study was investigated at the time the study was initiated in order to describe the responders versus the non-responders. The results showed that recruitment decreased with age, and that non-responders were less likely than responders to reside in northern part of Norway. Validation studies of the population within the NOWAC cohort have shown that the distribution of exposures is independent of response rate (51). Furthermore, the incidence rates of total cancer correspond with national figures (39). As the response rate decreased with age, our observed risk of lung cancer and our PAF could be underestimated, as the mean age at lung cancer diagnosis was above 60 years, and the proportion of daily smokers declined more in 2005-2016 among those <45 year compared to older age groups. If the smoking prevalence in the NOWAC cohort is lower than in the Norwegian, population of middle-aged women because of the “healthy volunteer effect”, our PAF could be underestimated. Within the NOWAC cohort, we assume that the possibility of selection bias in our study is limited.

In cohort studies, differential loss to follow-up is an analogue of selection bias. That is, whether participants who are lost to follow-up over the course of the study are different from those who remain under observation up to the event of interest or the end of the study.

Participants who leave the study due to mortality from other causes instead of to the outcome of interest, called competing risk, tend to have different probabilities of the outcome than the participants who remain in the cohort until the end of the study (46). In Norway, men have a higher mortality from cardiovascular diseases than women; indeed until 2017, cardiovascular diseases were the most common cause of death in Norway (52). As a result, men are less likely than women to survive long enough to develop lung cancer. This scenario would mean that our observed risk of lung cancer in men is underestimated. Thus, the observation of a higher risk of lung cancer (Paper 1) and risk of lung cancer death (Paper 2) in women

compared with men could be biased toward a reduced risk in men. Likewise, in the NOWAC study, the observed risk of lung cancer in smokers could be underestimated, but not to the same extent in women as in men, because cardiovascular diseases are not as common in women as in men (53). One of the main strengths of both the NHSS and the NOWAC study is the lack of loss to follow-up (37, 39).

5.2.3 Information bias

Information bias can occur if either the measurements of risk factors, like smoking, the outcome (diagnosed lung cancer or lung cancer mortality), or other covariates, are systematically distorted. Information bias leads to misclassification of the exposure and the outcome. Misclassification bias can be classified as either differential (dependent on the outcome) or non-differential (not dependent on the outcome). In cohort studies, information bias tends to be non-differential, meaning that the bias does not affect any one group more than another (46). In the NHSS and in the NOWAC study, standard protocols were used to minimize such errors. As measurements of exposure data were based on self-administered questionnaires in both the NHSS and the NOWAC study, inaccurate exposure measurements may be a main source of information bias, as discussed in the next section.

5.2.4 Information bias according to measures of smoking exposure in the NHSS

Self-reported smoking status has been widely used to assess the detrimental health effects of smoking. However, self-reporting can be unreliable if the subject feels under pressure because of social or medical disapproval. A systematic review found trends of underestimation when smoking prevalence was based on self-reported information (54). Furthermore, smokers may adopt a healthier lifestyle when participating in health studies, making it more difficult to detect associations.

All measures of smoking exposure were done at the time of enrollment. The relevance of an exposure depends on the stability of the variable. A variable that is unstable over time will usually veil the association in the direction of a zero association (55). As we only measured smoking exposure at enrollment, we do not know if smoking exposure changed during follow-up. Likewise, we do not know if some of those who reported that they did not smoke at enrollment started to smoke during follow-up. In Norway, smoking prevalence has decreased among men since the late 1950's, and among women since 1970 (2). We believe that our information regarding all measures of smoking exposure could be biased as a consequence of the decreased smoking prevalence in Norwegian men and women during follow-up. In other words, smoking exposure has been an unstable variable throughout follow-up for both men and women. In Norway, the decline in the smoking prevalence has been greatest among men. This could veil the association between smoking and lung cancer in the direction of zero, increasingly in men compared with women. Our results of an increased risk of lung cancer (Paper 1) and lung cancer death (Paper 2) in women compared with men, could be biased because of the unstable smoking variable. To increase the accuracy of measures of smoking exposure in our study, differences in smoking behavior should have been measured throughout follow-up, instead of only at baseline.

Never smokers were the reference in all Cox regression analyses in Papers 1 and 2, but we did not explain how never smokers were defined in the methods section of either paper.

Participants who reported to have smoked less than 100 cigarettes in their lifetime were defined as never smokers. As most men and women in Norway initiate smoking before age 25 years (56), we consider it unlikely that a significant number of men or women who reported being never smokers at study enrollment started smoking during follow-up. Passive and occasional smoking was not considered, as no data was available; therefore passive and occasional smokers were included in the reference group (among never smokers). During our

follow-up period, around 10% of the Norwegian population reported to be occasional smokers (57). Norwegian occasional smokers often define themselves as never smokers (58). Some of these occasional smokers may have been excluded from our analytical sample due to insufficient smoking information, whereas others may have been included as never smokers. If some of the occasional smokers in our cohort are in the reference group as never smokers, the risk estimates between smoking and lung cancer/lung cancer mortality may be underestimated. If there is a different proportion of occasional smokers among never smokers by sex, this could have distorted our results of an increased risk of lung cancer and lung cancer mortality in women compared with men.

In the NHSS, we divided participants into never, former, and current smokers. Participants who reported being current or former smokers were classified as ever smokers in Paper 1. By dividing smokers into current and former smokers, we made more homogenous smoking groups, although there are differences within the respective groups. Current smokers had an increased risk of lung cancer compared with former smokers in both men and women in Paper 1. The increased risk in current smokers compared with former smokers shows that the risk is dependent on time. Also in former smokers, we observed an increased risk of lung cancer by duration of smoking in years, number of cigarettes smoked per day, and pack-years. Likewise, for male and female former smokers, we observed a reduced risk of lung cancer (Paper 1) by time since smoking cessation. According to the smoking variables used to assess dose-response (smoking duration in years, number of cigarettes smoked per day, pack-years), all showed an increased risk of lung cancer (Paper 1) and lung cancer death (Paper 2) in both men and women.

In the NHSS, we also used continuous variables for smoking duration in 10-year increments, tens of cigarettes smoked per day, and tens of pack-years in addition to the categorical analyses to estimate dose-response associations for lung cancer. Using continuous variables, we managed to bypass the fact that in each category of smoking duration (e.g., 1-9, 10-19, 20-29 years of smoking), men were in the upper threshold of each category. When we used measures of smoking exposure as categorical variables, we observed an increased risk of lung cancer in men compared with women. In contrast, when we used the measures of smoking exposure as continuous variables, we observed an increased risk of lung cancer and lung cancer death in women compared with men. We believe that using smoking exposure as a continuous variable improved our risk estimates for the purpose of studying eventual sex differences in the risk of smoking-associated lung cancer and lung cancer death.

5.3 Information bias according smoking exposure in NOWAC

In the NOWAC study, we measured smoking exposure only from time of enrollment. At enrollment, women reported whether they had ever smoked, the average number of cigarettes smoked per day, if they currently smoked daily, if their parents smoked during childhood, and if they lived with a smoker as adults. Based on the answers to these questions, we computed total years of smoking, number of cigarettes smoked per day, pack-years, age at smoking initiation, and, for former smokers, years since smoking cessation. We used never smokers not exposed to passive smoking as the reference group. As we had information about passive smoking, we could calculate the risk of lung cancer in passive smokers compared with never smokers. Participants who reported to have smoked less than 100 cigarettes in their lifetime were defined as never smokers. Among passive smokers, we did not find any significant, increased risk of lung cancer compared with never smokers. There are several explanations as to why we did not find any clear association between passive smoking and risk of lung cancer. There may be inaccuracies in self-reported passive exposure, which is likely most pronounced

for childhood exposure. There was no information regarding smoking exposure at the workplace in our study, thus we can't exclude any increased risk of lung cancer associated with passive smoking at workplace. Additionally, we observed few lung cancer cases among never smokers, resulting in a lack of statistical power. By dividing smokers into current and former smokers, we created more homogenous smoking groups, but there were still differences. Current smokers had an increased risk of lung cancer compared with former smokers. Additionally, in Paper 3, any decreased smoking prevalence in Norwegian women during follow-up would have reduced our PAF estimates.

5.3.1 Information bias according to outcomes: lung cancer and lung cancer mortality

The outcomes of interest were lung cancer (Papers 1 and 3) and lung cancer mortality (Paper 2). In a cohort study, information about endpoints should be obtained in the same way (59). In Norway, all medical doctors are instructed by law to report new cancer cases to the Cancer Registry of Norway. The report is done by physicians based on clinical evidence, and by a pathologist based on the histological report. The Cancer Registry of Norway is regarded as one of the most complete in the world; in evaluations it has shown a high degree of comparability, accuracy, and timeliness (43).

Doctors, who are required to complete a death certificate, report information about cause of death in Norway. The cause of death is based on the doctor's clinical evaluation, previous knowledge of the deceased, previous radiological examinations, and other relevant information. Lack of clinical experience, lack of time, and lack of knowledge about the patient may lead to erroneous conclusions (60).

To correct for errors and incorrect conclusion regarding cause of death, rules from the World Health Organization are used to ensure that the correct cause of death is recorded on the death

certificate (61). In 2012, a Norwegian report compared the underlying cause of death in death certificates with the results from all medical autopsies; it revealed a change in the underlying cause of death in 61% of cases (61). Further, the researchers observed a change in the assigned ICD code in 32% of the reports, illustrating considerable uncertainty when cause of death is taken from death certificates only. Overall, the validity of the mortality data from the Norwegian Death Registry should be considered with some reservation.

5.3.2 Confounding in the Norwegian Health Screening Surveys and the Norwegian Women and Cancer study

Confounding refers to a situation in which a non-causal association between a given exposure and an outcome is observed as a result of the influence of a third variable, usually called a confounder. The confounding variable must be related to both the exposure and the outcome of interest. In an observational cohort study, a confounding variable would differ between exposed and unexposed participants (46). In contrast to bias, it is possible to control for confounders by statistical techniques like adjustment and stratification in multivariate models (46, 62). The comparison of unadjusted and adjusted associations is the best approach to support the presence of confounding if the estimates differ (46). The identification of potential confounders is usually based on *a priori* knowledge of the dual association of the possible confounder with the exposure and outcome. In the multivariate models of Paper 1, we *a priori* selected physical activity, BMI, and duration of education as covariates, and they were all available, in an attempt to exclude the possibility that these factors confounded our results regarding smoking and lung cancer. In Paper 2, we considered alcohol, physical activity, BMI, and duration of education as possible confounders. As we tested and found an interaction between smoking and sex, and a borderline interaction between smoking and duration of education, we decided to stratify by sex and duration of education. We decided to

adjust for BMI and physical activity, but did not include alcohol as a covariate due to a large amount of missing data. A recent cohort study based on data from the Tromsø study examined the association between occasional smoking and total mortality; it observed that adjustment for alcohol consumption changed total mortality only marginally (63). As lung cancer has been one of the most frequent causes of death in Norway recent years (64), lung cancer mortality constituted a high proportion of deaths in the cohort study from Tromsø. Therefore, we did not do a subanalysis with alcohol as a covariate in Paper 2. In Paper 3, we included covariates that changed the HR estimates in the crude model by at least 5%, that could potentially confound the association between smoking and lung cancer. We included the following variables in the final multivariate models; age at enrollment, duration of education (<10, 10-12, \geq 13 years), and average alcohol consumption. We observed a tendency of a higher HR in the crude models compared with the multivariate-adjusted HR. This could be due to confounding by education level and alcohol consumption. As we did not include alcohol consumption as a covariate in Papers 1 or 2, this could distort the HR toward higher estimates. As men tend to consume more alcohol than women, the risk estimates could potentially be distorted toward an increased risk in men (65). Alcohol consumption is not an independent risk factor for lung cancer according to the 2014 World Cancer Report (1). Cohort studies are inconsistent regarding the association between alcohol consumption and lung cancer (19, 20), and we do not consider missing information on alcohol consumption to be a big limitation of Papers 1 or 2.

5.3.3 Residual confounding in the Norwegian Health Screening Surveys and the Norwegian Women and Cancer study

Residual confounding occurs when adjustment does not completely remove the confounding effect of a variable or a set of variables (46). We adjusted for smoking using the categorical

definitions never, former, current, and ever smokers. The variability in cumulative dose within former, current, and ever smokers may be large, and could have resulted in important residual confounding when we evaluated the relationship between smoking and lung cancer/lung cancer mortality. As we studied possible sex differences in smoking-associated lung cancer/lung cancer mortality, broad categories like former, current, and ever smokers could have led to an increased risk in men compared with that in women, because men were at the upper threshold of each smoking category. Thus, the increased risk of lung cancer/lung cancer death in men compared with women, when using such broad categories of smoking is likely a result of residual confounding. To reduce the confounding effect of smoking, we used other measures of smoking exposure, such as duration and intensity. In addition, we estimated dose-response associations between lung cancer/lung cancer mortality and continuous variables for smoking duration in 10-year increments, tens of cigarettes smoked per day, and tens of pack-years. If some other important confounding variables were not included in our models, our adjusted estimates would still be confounded. In the NHSS, we lack information about passive smoking. Our reference group (never smokers) is therefore most likely contaminated with passive smokers. Since more men than women among our participants were smokers, it is likely that more female never smokers than male never smokers were exposed to passive smoke. For women, this would have attenuated our observed risk of lung cancer among ever smokers. In Paper 3, we did not find any significantly increased risk of lung cancer in passive smokers compared with never smokers, which could be due to few cases in never smokers. Causes of lung cancer other than smoking include some occupational exposures, like radon, certain metals, asbestos, and diesel engine exhaust (1). Although occupational exposure might be a confounding variable, we could not adjust for it in any of the three papers, as we did not have this information. The types of occupational exposures

associated with lung cancer are most frequent in men, and as a confounder, could lead to a higher risk of lung cancer in men compared with women.

5.4 Information about education in the Norwegian Health Screening Surveys and the Norwegian Women and Cancer study

We used the most recent information regarding duration of education obtained from Statistics Norway to classify each participant by education level. In 1965, duration of compulsory school attendance in Norway changed from 7 to 9 years. Therefore, 10 years of education means primary school with at most 2 years of additional education. Similarly, those with 10-12 years of education have completed secondary school, or at most 5 years of professional training. Education lasting ≥ 13 years corresponds to university level in terms of education, or a lower level of education with several years of professional training. Information about education in Statistics Norway comes from three population censuses conducted in 1970, 1980, and 1990 (each census year is November 1st) (66, 67). In the 1970 census (as in the 1960 census), education was coded according to information from personal visits to each household. Register data for highest duration of education was used to determine education level in the 1980 and 1990 censuses. From the 1970 census onwards, education was coded according to the Norwegian Standard Classification of Education, which is compatible with the International Standard Classification of Education (68). We assigned participants to one of three categories of education: low (< 10 years), moderate (10-12 years), and high (≥ 13 years). We do not think that there is much information bias according to education level in the NHSS or the NOWAC study.

Education is an indicator of socioeconomic status and has been shown to be inversely associated with cancer mortality and survival (69, 70). Both the NHSS and the NOWAC

study contained information about income. We found it problematic to use this information in our longitudinal study, as it is difficult to compare income levels across groups recruited at different time periods. Therefore, we used education as a measure of socioeconomic status.

5.5 Time variable in the model

We used the Cox proportional hazards model to calculate our risk estimates. The time-independent model was used, with age (at enrollment) as the time scale. In the Cox model, the assumption underlying the model was that the risk factor is associated with the fixed relative increase in the instantaneous risk of the outcome of interest, compared with the reference hazard (46), i.e., the hazard among those exposed is constant at any given point in time. The two papers from the NHSS had a long follow-up period that varied from 10 to 39 years; thus the hazard will fluctuate with time (“calendar time”). As recommended, we stratified by birth cohort to account for some of these changes (71). One problem with using age as the time scale (underlying time variable) is that the three surveys included in NHSS were conducted decades apart. The model considers a woman who was 40 years of age at enrollment in 1975 in exactly the same manner as a woman who was 40 years of age at enrollment in 1995. A model using calendar year of birth would have accounted for this possible bias. However, when stratifying by birth cohort, our model with age should have accounted sufficiently for this problem.

6 Discussion of main results

The main results have been discussed in detail in the enclosed papers (Papers 1, 2, and 3). In this section, we present a more general discussion of the main findings, in the context of what was known previously.

6.1 Sex differences in smoking-associated risk of lung cancer and lung cancer death

In Papers 1 and 2, we found that the proportion of heavier smokers and former smokers were higher among men than among women. This is in accordance with the knowledge of smoking habits in Norway, where men have consumed more than 70% of the cigarettes smoked (2). During the nearly 12 million person-years of follow-up in each of the studies (Papers 1 and 2), we identified 6,534 lung cancer cases [43% women (Paper 1)] and 5,702 (42% women) lung cancer deaths (paper 2). The higher lung cancer incidence and mortality observed in men compared with women reflects the differences in smoking prevalence by sex, as men and women entered the stages of the smoking epidemic at different calendar times (72). The 2017 report, *Cancer in Norway*, showed optimistic signs of stabilization of the lung cancer rate in women. However, the following report in 2018 showed the rate for lung cancer in women at an all-time-high, with 1,674 new cases (73). Since lung cancer mortality rates for Norwegian women have not yet peaked, they may go even higher than previously expected, as noted by Tverdal in a paper published in 2001 (74). We observed a small difference in the number of lung cancer cases and lung cancer deaths by sex, but this is due to the high fatality rate for lung cancer. In Norway in 2013, the 5-year survival rate for lung cancer was 13% for men and 19% for women (3). When we analyzed smoking exposure according to categorical groups (smoking duration in years, cigarettes smoked per day, and pack-years of smoking), we did not detect any difference in the risk of lung cancer or lung cancer death between men and women. Similar observations have been reported from other recent cohort studies on smoking and the risk of lung cancer. Bain and colleagues (75) found a HR of 1.11 (95% CI: 0.87-1.42) in female current smokers compared with male ones, adjusted for number of cigarettes smoked per day and age at smoking initiation. In a 2008 study from the US by Freedman et al. (76), current smokers who smoked 1-10 cigarettes per day showed a HR of 0.7 (95% CI: 0.6-

0.9) for women compared with men in regard to the risk of lung cancer. In a study published in 2016 (77), Freedman et al., found that the relative risk of lung cancer for current smokers who smoked 1-10 cigarettes per day compared with never smokers, 17.3 (95 CI: 14.3-20.9) men and 12.2 (95% CI: 10.4-14.2) for women. However, when we analyzed smoking exposure as a continuous variable, female current smokers had a significantly higher risk of lung cancer and lung cancer death than male current smokers for 10-year increments of smoking, 10 cigarettes/day, and 10 pack-years. Results from analyses where measures of smoking exposure were included as continuous variables were not in line with those of other cohort studies. The Freedman and Bain reports analyzed the risk of lung cancer according to fixed categories of smoking exposure. The increased risk of lung cancer among women that we observed when we analyzed the data continuously was most likely concealed when the smoking exposure data were categorized as in the other three studies. Since men are generally heavier smokers than women, they are more likely to be in the upper threshold of each category.

Other cohort studies also observed no difference in the smoking-associated risk of lung cancer death by sex (78-83). A Norwegian study published in 2006 (78) found that the risk of lung cancer death among current smokers who smoked 1-9 cigarettes per day was 2.24 (95% CI: 1.99-2.53) for men and 1.80 (95% CI: 1.59-2.04) for women. They also estimated the women-men mortality HR, adjusted for age at smoking initiation and number of cigarettes smoked per day among current smokers (1.00, 95% CI: 0.80-1.26), and did not observe any significant differences between men and women for deaths due to lung cancer. The study “50-year trends in smoking-related mortality in the United States” by Thun and colleagues (82) covered the contemporary period (2000-2010). The authors observed that the risk of lung cancer death among current smokers compared with never smokers was virtually identical for men and women, at 24.97 (95% CI: 22.20-28.09) and 25.66 (95% CI: 23.17-28.40), respectively. Also,

studies by Prescott et al. (81), Nilsson et al. (79), and Jamrozik et al. (80) found that the risk of smoking-associated lung cancer death was similar in men and women with comparable smoking exposures. The study by Marang-van de Mheen et al. (83) included 15,393 men and women and reported that rate ratios for lung cancer mortality among light smokers (1-14 cigarettes per day) tended to be higher for men than for women. For the other categories of smoking exposure (15-24 cigarettes and 25+ cigarettes), they reported similar rate ratios for men and women. The mortality studies did not use continuous measures of smoking exposure, and thus men may have been in the upper and women in the lower part of a specific category, but they were still classified as similarly exposed.

An alternative explanation for the higher risk of lung cancer and lung cancer death in women compared with men may be competing risk of death. Since men smoke more than women, they have an increased risk of dying from other smoking-associated diseases before they experience the outcome of interest (lung cancer). In our follow-up period in Paper 2, a total of 30,585 (9.9%) women died, and the number of deaths among men was 40,974 (14.3%). The number of lung cancer deaths in our study was 3,323 in men and 2,379 in women. Thus, competing risk of death might be a problem in our study, attenuating the risk of lung cancer and lung cancer death among men.

A major limitation of our study (Papers 1 and 2) is that we have information on smoking only at study enrollment. Indeed, smoking prevalence has changed significantly in Norway since the start of study enrollment in 1974, with the prevalence of daily smokers decreasing from 51% to 15% in men and from 32% to 14% in women during our follow-up period (84). As a result of the steeper decline in smoking prevalence in men compared with women, our risk estimates of lung cancer in men could be overestimated. Repeated assessments of smoking during follow-up could have been one way to avoid this kind of systematic error.

Another limitation is that 10% of the Norwegian general population reported to be occasional smokers during our follow-up period (57). Thus some of these occasional smokers may have been included as never smokers in our study sample. This would have attenuated the risk of smoking-associated lung cancer and lung cancer death, but we do not think that occasional smokers have distorted our results. Our results from Papers 1 and 2 support a hypothesis that women are more susceptible than men to lung cancer and lung cancer mortality, given the same smoking exposure. Our findings are exciting, and future studies should be conducted to confirm them. In addition, research should focus on possible risk factors other than smoking in the risk of lung cancer in women. It is of great importance to determine whether women might be more vulnerable than men to smoking-associated lung cancer.

6.2 Social inequalities in smoking-associated lung cancer mortality

In Norway, there is a marked social gradient for active as well as passive smoking. The lower one's education, the more likely it is that one may smoke (85). As expected, in Paper 2, we observed that the age standardized rates of lung cancer mortality were highest among male and female current smokers with the lowest education level compared with those with middle and high education. Among male current smokers, the risk of lung cancer death did not vary significantly when we compared those with the lowest and highest education levels. For female current smokers, there was a significant difference in the risk of lung cancer death between those with low and high education levels. As a consequence of the large proportion of heavy smoking men and women with a low education level, the differences we observed by education level should be interpreted with caution, as they could be due to residual confounding by smoking in the less educated. The less educated are more likely to be exposed to passive smoking from spouses, air pollution, and occupational exposures, which are also risk factors for lung cancer, and could potentially increase their risk of lung cancer diagnosis,

and thereby lung cancer death (14). Our results of a higher risk of lung cancer death in the less educated as a probable consequence of heavier smoking exposure is supported by two other cohort studies. Hart et al. (69) compared the risk of lung cancer death among manual workers and non-manual workers, and observed significantly higher risks for both male and female manual workers. Adjustment for smoking exposure reduced the risks substantially, but they remained statistically significant for men, although they became non-significant for women. Hart et al., suggested that smoking only provides a partial explanation for the higher risk of lung cancer death in manual workers compared with non-manual workers, although the possibility of residual confounding by smoking cannot be ruled out. A report from the NOWAC study found an overall negative socioeconomic gradient in lung cancer mortality when socioeconomic status was measured as years of education or income (70). After adjustment for stage and smoking status, mortality differences according to both education level and income became non-significant. In a study from Sweden, they observed that the risk of lung cancer was significantly elevated among participants with low compared to high socioeconomic status, despite accounting for smoking habits (86). They suggested that the association between low socioeconomic status and lung cancer could be a result of residual confounding by smoking, or other lifestyle factors, and occupational exposure. A study from the EPIC cohort observed that adjustment for smoking decreased educational differences in the risk of lung cancer by 50-70% in both men and women in Northern European countries (87). The EPIC study also observed a socioeconomic gradient in the risk of lung cancer among never smokers, and said that, in future studies, risk factors other than smoking should be considered. Our study adds to the information that healthcare policy should be aimed at changing risky behaviors, especially in the less educated, but it should also include measures to reduce inequalities in health.

6.3 Active smoking and risk of lung cancer in women

In Paper 3, current smokers had a higher risk of lung cancer than former smokers, which confirms prior findings. Our study showed a consistent, strong dose-response relationship between smoking and risk of lung cancer. Compared with never smokers, current smokers in our study had a 13-fold increased risk of lung cancer. The strength of the associations in our study is similar to those from a prospective study of one million women in the UK, published in 2013, and a prospective study from the US published in 2015 with 93,000 women. These two prospective studies reported a 21-fold and a 13-fold increased risk of lung cancer, respectively, among current smokers compared with never smokers (88, 89). In our study, those who quit smoking within the last 9 years had a 63% lower risk of lung cancer compared with current smokers. These results are consistent with a Norwegian cohort study which reported a 75% lower risk in female former smokers who quit smoking 5-9 years prior (90). Moreover, more than eight in ten lung cancer cases could have been avoided if the women had been never smokers. The PAF in our study was larger than that reported for US white women (78%) in a 2014 biracial prospective cohort study from the US, with 173 lung cancer cases among the 5487 white women in the study (91). This high PAF is what we found in a nationally representative cohort study. Norway was one of the first countries to introduce restrictive tobacco control (92). Our study shows that tobacco control campaigns and restrictions should continue to be a high priority in Western countries. In developing countries, more restrictive tobacco campaigns should be implemented so that men and women in these countries can avoid the experience of populations in developed countries.

6.4 Passive smoking and risk of lung cancer in women

Lung cancer is also a significant health problem in those with no history of active smoking (1, 11, 12, 93). In Paper 3, we found that women with exposure to passive smoking either in childhood or from a spouse had a 34% (HR: 1.34, 95% CI: 0.89-2.01), non-significant

increased risk of lung cancer, compared with never smokers. There are several possible explanations as to why we did not find any clear association between passive smoking and the risk of lung cancer. The main reason is lack of statistical power, as the number of never and passive smokers is low in the NOWAC study. There may be inaccuracies in the self-reporting of passive exposures, which is likely most pronounced for childhood exposures. However, as exposure information was collected at baseline before lung cancer diagnosis, true recall bias is unlikely. There was no information regarding smoking exposure at the workplace in our study, and we cannot rule out that some of the risk among never smokers could be due to this, and thus attenuate the association between passive smoking and lung cancer. We must also consider that passive smoking may have a weaker than expected association with lung cancer for middle-aged women, which has been suggested by some previous prospective cohort studies. A 2005 study based on EPIC data included 95,947 women, among whom 70 were diagnosed with lung cancer; it did not find any significantly increased risk of lung cancer in women exposed to passive smoke at home (94). A 2015 cohort study from the US included 152 lung cancer cases among never smokers and did not find any significant association between passive smoking and the risk of lung cancer (89). Likewise, in the UK Million Women Study, which included 1,469 cases among the 634,039 never smokers, no significant association was found between lung cancer and passive smoking as a child and/or adult. They did not have information on exposure duration among passive smokers (13). Our results suggest that future studies should confirm whether passive smoking is associated with lung cancer.

7 Conclusions

7.1 Paper 1

- Women have an increased risk of lung cancer compared with men, given the same lifetime smoking exposure

7.2 Paper 2

- Women are more susceptible than men to the effect of smoking in regard to lung cancer mortality, given the same smoking history
- Low education level confers an increased risk of dying of lung cancer, which could be due to residual confounding by smoking or other factors unknown to us among both men and women

7.3 Paper 3

- Current smokers showed a consistently higher risk of lung cancer than former smokers
- More than eight in ten lung cancer cases could have been avoided in Norwegian women in the absence of smoking
- We did not find any significantly increased risk of lung cancer in passive smokers compared with never smokers

References

1. Steward BW, Wild CP. World Cancer Report 2014.: International Agency for Research on Cancer World Cancer Report; 2014.
2. Lund I, Lund KE. Lifetime smoking habits among Norwegian men and women born between 1890 and 1994: a cohort analysis using cross-sectional data. *BMJ Open*. 2014;4(10):e005539.
3. Cancer Registry of Norway. Cancer in Norway 2013. Cancer incidence, mortality, survival and prevalence. Oslo, Norway; 2013.
4. Doll R, Hill AB. Smoking and carcinoma of the lung; preliminary report. *Br Med J*. 1950;2(4682):739-748.
5. Doll R, Hill AB. Lung cancer and other causes of death in relation to smoking; a second report on the mortality of British doctors. *Br Med J*. 1956;2(5001):1071-1081.
6. Tyczynski JE, Bray F, Parkin DM. Lung cancer in Europe in 2000: epidemiology, prevention, and early detection. *Lancet Oncol*. 2003;4(1):45-55.
7. IARC. Monogr Eval Carcinog Risks Hum. Personal habits and indoor combustions. Volume 100 E. A review of human carcinogens. Lyon; 2012.
8. Doll R, Peto R. Cigarette smoking and bronchial carcinoma: dose and time relationships among regular smokers and lifelong non-smokers. *J Epidemiol Community Health*. 1978;32(4):303-313.
9. Lopez A, D, Collishaw, N. E, Piha, T, A,. A descriptive model of the cigarette epidemiv in developed countries. *Tob Control*. 1994;3:242-247.
10. Statistics Norway. Percentage daily smokers 2013 [updated 18.01.2018; cited 2018 20.02]. Available from: <https://www.ssb.no/en/royk>.
11. IARC Monogr Eval Carcinog Risks Hum. Tobacco smoke and involuntary smoking. International Agency for Research on Cancer, Lyon; 2004.
12. Surgeon General. US Department of Health and Human Services, Centers for Disease Control and Prevention, Coordinating Center for Health Promotion, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, Atlanta.; 2006.
13. Pirie K, Peto R, Green J, Reeves GK, Beral V. Lung cancer in never smokers in the UK Million Women Study. *Int J Cancer*. 2016;139(2):347-354.
14. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin*. 2015;65(2):87-108.
15. Devesa SS, Diamond EL. Socioeconomic and racial differences in lung cancer incidence. *Am J Epidemiol*. 1983;118(6):818-831.
16. Linseisen J, Rohrmann S, Bueno-de-Mesquita B, Buchner FL, Boshuizen HC, Agudo A, et al. Consumption of meat and fish and risk of lung cancer: results from the European Prospective Investigation into Cancer and Nutrition. *Cancer Causes Control*. 2011;22(6):909-918.
17. Buchner FL, Bueno-de-Mesquita HB, Linseisen J, Boshuizen HC, Kiemenev LA, Ros MM, et al. Fruits and vegetables consumption and the risk of histological subtypes of lung cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC). *Cancer Causes Control*. 2010;21(3):357-371.
18. Rohrmann S, Linseisen J, Boshuizen HC, Whittaker J, Agudo A, Vineis P, et al. Ethanol intake and risk of lung cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC). *Am J Epidemiol*. 2006;164(11):1103-1114.
19. Freudenheim JL, Ritz J, Smith-Warner SA, Albanes D, Bandera EV, van den Brandt PA, et al. Alcohol consumption and risk of lung cancer: a pooled analysis of cohort studies. *Am J Clin Nutr*. 2005;82(3):657-667.
20. Thun MJ, Hannan LM, DeLancey JO. Alcohol consumption not associated with lung cancer mortality in lifelong nonsmokers. *Cancer Epidemiol Biomarkers Prev*. 2009;18(8):2269-2272.

21. Surgeon General's report on physical activity and health. From the Centers for Disease Control and Prevention. *JAMA*. 1996;276(7):522.
22. Vainio H, Bianchini F. IARC Handbooks of cancer prevention. Volume 6: weight control and physical activity. IARC Handbooks of Cancer Prevention. France: IARC Press; 2002.
23. Steindorf K, Friedenreich C, Linseisen J, Rohrmann S, Rundle A, Veglia F, et al. Physical activity and lung cancer risk in the European Prospective Investigation into Cancer and Nutrition Cohort. *Int J Cancer*. 2006;119(10):2389-2397.
24. Patel AV, Carter BD, Stevens VL, Gaudet MM, Campbell PT, Gapstur SM. The relationship between physical activity, obesity, and lung cancer risk by smoking status in a large prospective cohort of US adults. *Cancer Causes Control*. 2017;28(12):1357-1368.
25. Lauby-Secretan B, Scoccianti C, Loomis D, Grosse Y, Bianchini F, Straif K. Body Fatness and Cancer--Viewpoint of the IARC Working Group. *N Engl J Med*. 2016;375(8):794-798.
26. Dewi NU, Boshuizen HC, Johansson M, Vineis P, Kampman E, Steffen A, et al. Anthropometry and the Risk of Lung Cancer in EPIC. *Am J Epidemiol*. 2016;184(2):129-139.
27. Grimsrud TK, Skaug HK, Larsen IK. [Lung cancer--changes in incidence by gender, age and county of residence 1984-2013]. *Tidsskr Nor Laegeforen*. 2015;135(23-24):2131.
28. US Department of Health and Human Services. The Health Consequences of Smoking-50 Years of Progress. A Report of The Surgeon General. Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and Chronic Disease prevention and Health Promotion, Office on Smoking and Health; 2014.
29. Tobacco smoke and involuntary smoking. *IARC Monogr Eval Carcinog Risks Hum*. 2004;83:1-1438.
30. Cancer Registry of Norway. Nasjonalt kvalitetsregister for lungekreft. Oslo, Norway; 2013-2014.
31. Stocks T, Borena W, Strohmaier S, Bjorge T, Manjer J, Engeland A, et al. Cohort profile: The metabolic syndrome and cancer project (Me-Can). *Int J Epidemiol*. 2010;39(3):660-667.
32. Bjartveit K, Foss OP, Gjervig T. The cardiovascular disease study in Norwegian counties. Results from first screening. *Acta Med Scand*. 1983;214(SUPPL. 675):184P.
33. Tverdal A, Bjartveit K. Health consequences of reduced daily cigarette consumption. *Tob Control*. 2006;15(6):472-480.
34. Aires N, Selmer R, Thelle D. The validity of self-reported leisure time physical activity, and its relationship to serum cholesterol, blood pressure and body mass index. A population based study of 332,182 men and women aged 40-42 years. *Eur J Epidemiol*. 2003;18(6):479-485.
35. Bjartveit K, Stensvold I, Lund-Larsen PG, Graff-Iversen S, Urdal P. Cardiovascular screenings in Norwegian counties. Trends in risk pattern during the period 1985-90 among persons aged 40-42 in 4 counties. [Norwegian]. *Tidsskr Nor Laegeforen*. 1991;111(17):2072-2076.
36. Aamodt G, Sogaard AJ, Naess O, Beckstrom AC, Samuelsen SO. [The CONOR database--a little piece of Norway]. *Tidsskr Nor Laegeforen*. 2010;130(3):264-265.
37. Naess O, Sogaard AJ, Arnesen E, Beckstrom AC, Bjertness E, Engeland A, et al. Cohort profile: cohort of Norway (CONOR). *Int J Epidemiol*. 2008;37(3):481-485.
38. Lund E, Dumeaux, V, Braaten, T, Hjartaker, A, Engeset, D, Skeie, G, et al. Cohort profile: The Norwegian Women and Cancer Study--NOWAC--Kvinner og kreft. *Int J Epidemiol*. 2008;37(1):36-41.
39. Lund E, Kumle M, Braaten T, Hjartaker A, Bakken K, Eggen E, et al. External validity in a population-based national prospective study--the Norwegian Women and Cancer Study (NOWAC). *Cancer Causes Control*. 2003;14(10):1001-1008.
40. Bobak M, Hertzman C, Skodova Z, Marmot M. Own education, current conditions, parental material circumstances, and risk of myocardial infarction in a former communist country. *J Epidemiol Community Health*. 2000;54(2):91-96.
41. Ranjan Parajuli. Smoking and incidence and mortality of colorectal cancer [A dissertation for the degree of Philosophiae Doctor]: UiT-The Arctic University of Norway; 2014.
42. Eivind Bjerkaas. Aspects of Active Smoking and Breast Cancer [A dissertation for the degree of Philosophiae Doctor]: UiT-The Arctic University of Norway; 2014.

43. Larsen IK, Smastuen M, Johannesen TB, Langmark F, Parkin DM, Bray F, et al. Data quality at the Cancer Registry of Norway: an overview of comparability, completeness, validity and timeliness. *Eur J Cancer*. 2009;45(7):1218-1231.
44. World Health Organization. International Classification of Diseases for Oncology ICD-O-3 online. Geneva: World Health Organization; 2000.
45. Grimes DA, Schulz KF. Bias and causal associations in observational research. *The Lancet*. 2002;359(9302):248-252.
46. Szklo M, Nieto F.J. *Epidemiology: Beyond the Basics*. USA: Jones & Bartlett, LLC, an Ascend Learning Company; 2014. 515 p.
47. Bjartveit K, Foss OP, Gjervig T. The cardiovascular disease study in Norwegian counties. Results from first screening. *Acta Med Scand Suppl*. 1983;675:1-184.
48. Laake P, Benestad H, Olsen B. *Research methodology in the medical and biological sciences*. Academic Press. Elsevier: Amsterdam; 2007.
49. Holme I, Helgeland A, Hjermann I, Leren P, Lund-Larsen PG. Four and two-thirds years incidence of coronary heart disease in middle-aged men: the Oslo study. *Am J Epidemiol*. 1980;112(1):149-160.
50. Krokstad S, Langhammer A, Hveem K, Holmen TL, Midthjell K, Stene TR, et al. Cohort Profile: the HUNT Study, Norway. *Int J Epidemiol*. 2013;42(4):968-977.
51. Lund E, Gram IT. Response rate according to title and length of questionnaire. *Scand J Soc Med*. 1998;26(2):154-160.
52. Norwegian Institute of Public Health. Kreft er nå hyppigste dødsårsak i Norge 2018 [updated 20.12.2018; cited 2019 19.11]. Available from: <https://www.fhi.no/nyheter/2018/dodsarsakene-2017/>.
53. Norwegian Institute of Public Health. *Cardiovascular disease in Norway Oslo, Norway*: Norwegian Institute of Public Health; 2014.
54. Connor Gorber S, Schofield-Hurwitz S, Hardt J, Levasseur G, Tremblay M. The accuracy of self-reported smoking: a systematic review of the relationship between self-reported and cotinine-assessed smoking status. *Nicotine & tobacco research : official journal of the Society for Research on Nicotine and Tobacco*. 2009;11(1):12-24.
55. Rosner B, Willett WC, Spiegelman D. Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Stat Med*. 1989;8(9):1051-1069; discussion 1071-1053.
56. Lund M, Lindbak R. *Norwegian tobacco statistics 1973-2006*. Oslo; 2007.
57. Norwegian Institute of Public Health. *Public Health Report 2014* [updated 12.04.2016; cited 2016 03.10]. Available from: <https://www.fhi.no/en/>.
58. Kvaavik E, Scheffels J, Lund M. [Occasional smoking in Norway]. *Tidsskr Nor Laegeforen*. 2014;134(2):163-167.
59. Grimes DA, Schulz KF. Cohort studies: marching towards outcomes. *Lancet*. 2002;359(9303):341-345.
60. Aase S. Autopsy--still the gold standard? *Tidsskr Nor Laegeforen*. 2013;133(7):730-731.
61. Alfsen GC, Maehlen J. The value of autopsies for determining the cause of death. *Tidsskr Nor Laegeforen*. 2012;132(2):147-151.
62. Pourhoseingholi MA, Baghestani AR, Vahedi M. How to control confounding effects by statistical analysis. *Gastroenterology and hepatology from bed to bench*. 2012;5(2):79-83.
63. Lochen ML, Gram IT, Mannsverk J, Mathiesen EB, Njolstad I, Schirmer H, et al. Association of occasional smoking with total mortality in the population-based Tromso study, 2001-2015. *BMJ Open*. 2017;7(12):e019107.
64. Norwegian Institute of Public Health. Kreft er nå hyppigste dødsårsak i Norge [updated 20.12.2018].
65. Norwegian Institute of Public Health. *Alcohol and other psychoactive substances in Norway*. 2019.

66. Aasland OG, Hem E, Haldorsen T, Ekeberg O. Mortality among Norwegian doctors 1960-2000. *BMC Public Health*. 2011;11:173.
67. Vassenden K. Folke- og boligtellingsene 1960, 1970 og 1980 (Norwegian). Oslo-Kongsvinger: Central bureau of statistics Norway; 1987.
68. UNESCO. International Standard Classification of education ISED. Paris: United Nations Educational, Scientific and Cultural Organization; 1997.
69. Hart CL, Hole DJ, Gillis CR, Smith GD, Watt GC, Hawthorne VM. Social class differences in lung cancer mortality: risk factor explanations using two Scottish cohort studies. *Int J Epidemiol*. 2001;30(2):268-274.
70. Braaten T, Weiderpass E, Lund E. Socioeconomic differences in cancer survival: the Norwegian Women and Cancer Study. *BMC Public Health*. 2009;9:178.
71. Korn EL, Graubard BI, Midthune D. Time-to-event analysis of longitudinal follow-up of a survey: choice of the time-scale. *Am J Epidemiol*. 1997;145(1):72-80.
72. Thun M, Peto R, Boreham J, Lopez AD. Stages of the cigarette epidemic on entering its second century. *Tob Control*. 2012;21(2):96-101.
73. Cancer Registry of Norway. Cancer in Norway 2018. Cancer incidence, mortality, survival and prevalence in Norway. Oslo, Norway; 2018.
74. Tverdal A. [Lung cancer mortality--now higher in women than in men under 50 years]. *Tidsskr Nor Laegeforen*. 2001;121(21):2487-2488.
75. Bain C, Feskanich D, Speizer FE, Thun M, Hertzmark E, Rosner BA, et al. Lung cancer rates in men and women with comparable histories of smoking. *J Natl Cancer Inst*. 2004;96(11):826-834.
76. Freedman ND, Leitzmann MF, Hollenbeck AR, Schatzkin A, Abnet CC. Cigarette smoking and subsequent risk of lung cancer in men and women: analysis of a prospective cohort study. *Lancet Oncol*. 2008;9(7):649-656.
77. Freedman ND, Abnet CC, Caporaso NE, Fraumeni JF, Jr., Murphy G, Hartge P, et al. Impact of changing US cigarette smoking patterns on incident cancer: risks of 20 smoking-related cancers among the women and men of the NIH-AARP cohort. *Int J Epidemiol*. 2016;45(3):846-856.
78. Vollset SE, Tverdal A, Gjessing HK. Smoking and deaths between 40 and 70 years of age in women and men. *Ann Intern Med*. 2006;144(6):381-389.
79. Nilsson S, Carstensen JM, Pershagen G. Mortality among male and female smokers in Sweden: a 33 year follow up. *J Epidemiol Community Health*. 2001;55(11):825-830.
80. Jamrozik K, McLaughlin D, McCaul K, Almeida OP, Wong KY, Vagenas D, et al. Women who smoke like men die like men who smoke: findings from two Australian cohort studies. *Tob Control*. 2011;20(4):258-265.
81. Prescott E, Osler M, Andersen PK, Hein HO, Borch-Johnsen K, Lange P, et al. Mortality in women and men in relation to smoking. *Int J Epidemiol*. 1998;27(1):27-32.
82. Thun MJ, Carter BD, Feskanich D, Freedman ND, Prentice R, Lopez AD, et al. 50-year trends in smoking-related mortality in the United States. *N Engl J Med*. 2013;368(4):351-364.
83. Marang-van de Mheen PJ, Smith GD, Hart CL, Hole DJ. Are women more sensitive to smoking than men? Findings from the Renfrew and Paisley study. *Int J Epidemiol*. 2001;30(4):787-792.
84. Norway S. Tobacco, alcohol and other drugs Oslo, Norway: Statistics Norway; 2019 [updated 18.01.2019; cited 2019 14.12]. Available from: <https://www.ssb.no/statbank/table/05307/>.
85. Næss Ø, Rognerud M, Strand B.H. Sosial ulikhet i helse. Folkehelseinstituttet; 2007.
86. Ekberg-Aronsson M, Nilsson PM, Nilsson JA, Pehrsson K, Lofdahl CG. Socio-economic status and lung cancer risk including histologic subtyping--a longitudinal study. *Lung Cancer*. 2006;51(1):21-29.
87. Menvielle G, Boshuizen H, Kunst AE, Dalton SO, Vineis P, Bergmann MM, et al. The role of smoking and diet in explaining educational inequalities in lung cancer incidence. *J Natl Cancer Inst*. 2009;101(5):321-330.
88. Pirie K, Peto R, Reeves GK, Green J, Beral V, Million Women Study C. The 21st century hazards of smoking and benefits of stopping: a prospective study of one million women in the UK. *Lancet*. 2013;381(9861):133-141.

89. Wang A, Kubo J, Luo J, Desai M, Hedlin H, Henderson M, et al. Active and passive smoking in relation to lung cancer incidence in the Women's Health Initiative Observational Study prospective cohort. *Ann Oncol.* 2015;26(1):221-230.
90. Hansen MS, Licaj I, Braaten T, Langhammer A, Le Marchand L, Gram IT. Sex Differences in Risk of Smoking- Associated Lung Cancer: Results From a Cohort of 600,000 Norwegians. *Am J Epidemiol.* 2017.
91. Prizment AE, Yatsuya H, Lutsey PL, Lubin JH, Woodward M, Folsom AR, et al. Smoking behavior and lung cancer in a biracial cohort: the Atherosclerosis Risk in Communities study. *Am J Prev Med.* 2014;46(6):624-632.
92. Norwegian Institute of Public Health. Tobacco Control in Norway [updated 10.01.2019; cited 2019 19.11]. Available from: <https://www.helsedirektoratet.no/english/tobacco-control-in-norway>.
93. Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. *CA Cancer J Clin.* 2005;55(2):74-108.
94. Vineis P, Airoidi L, Veglia F, Olgiati L, Pastorelli R, Autrup H, et al. Environmental tobacco smoke and risk of respiratory cancer and chronic obstructive pulmonary disease in former smokers and never smokers in the EPIC prospective study. *BMJ.* 2005;330(7486):277.

PAPER 1

Merethe S. Hansen, Ildir Licaj, Tonje Braaten, Arnulf Langhammer, Loic Le Marchand and Inger T. Gram

Sex Differences in Risk of Smoking-Associated Lung Cancer: Results from a Cohort of 600,000 Norwegians

Am J Epidemiol. 2018;187(5):971-981

Merethe S. Hansen, Ildir Licaj, Tonje Braaten, Arnulf Langhammer, Loic le Marchand and Inger Torhild Gram

Smoking related lung cancer mortality by education and sex in Norway

BMC Cancer (2019) 19:1132

RESEARCH ARTICLE

Open Access



Smoking related lung cancer mortality by education and sex in Norway

Merethe S. Hansen^{1*} , Ildir Licaj^{1,2}, Tonje Braaten¹, Arnulf Langhammer³, Loic Le Marchand⁴ and Inger Torhild Gram¹

Keywords: Cohort study, Lung cancer mortality, Smoking, Education, Sex differences

Background

Lung cancer is one of the most common forms of cancer and the leading cause of cancer death worldwide, with tobacco smoking as the main cause [1]. In Norway, as in other western countries, smoking was more prevalent among men and in the highest social classes six to seven decades ago [2]. The proportion of male smokers increased until the 1960s, when it was around 65%. Among women, the peak (35%) occurred in the late 70's [2]. From 1930 until the turn of the century, men have consumed more than 70% of the cigarettes smoked in Norway [3]. The decline in smoking prevalence occurred first and proceeded fastest among those with long education [4]. In Norway, lung cancer mortality for men has been declining since 2011, whereas as of 2013 it is still increasing for women [5]. Due to the lag period between start of smoking and lung cancer death, current mortality rates reflect smoking trends two to three decades earlier [6].

Neither the most recent World Cancer Report [1] nor the United States Surgeon General Report [7] discuss a possible sex difference in the risk of smoking associated lung cancer mortality. In 2001, Tverdal reported that among Norwegians under 50 years of age, lung cancer mortality was higher in women than in men [8]. Later Jha et al. reported from a US cohort, that among current compared with never smokers, women had a higher lung cancer mortality compared with men [9]. Since men and women have entered the stages of the smoking epidemic at different calendar times [10], a possible sex difference for smoking and lung cancer mortality may just have started to emerge. Education, an indicator

of socioeconomic status is inversely associated with cancer mortality [11, 12].

Studies from Europe have reported an increased risk of lung cancer in participants of low socioeconomic status despite accounting for smoking habits [13, 14]. To our knowledge, no other prospective cohort studies have examined lung cancer mortality by sex and education.

The objectives of the study were to explore a potential heterogeneity in smoking associated lung cancer mortality by sex and education.

Methods

Study population

The study population has been previously described [15] and comprises three national Norwegian health studies conducted between 1974 and 2003 by the Norwegian National Health Screening Service. Selection of participants was based on year of birth and residence (municipality or county). The response rate in the three studies varied from 56 to 88% [16]. Briefly, the three surveys used a similar protocol and study design, but there were some modifications made during different time periods, mainly due to questionnaires regarding smoking, physical activity and other lifestyle factors. Altogether 595,675 participants remained in the analytical cohort after exclusion of 40,091 participants due to emigration or death before the start of follow-up, missing information on vital status, measures of smoking exposure, education, or missing of any of the covariates included in the analyses.

The present study was approved by the Regional Committee for Medical Research Ethics South-East, Norway, and the National Data Inspectorate.

Exposure information

The questionnaires elicited information on current and former daily smoking, smoking duration in years

* Correspondence: merethe.s.hansen@uit.no

¹Department of Community Medicine, Faculty of Health Sciences, The UiT Arctic University of Norway, Tromsø, Norway

Full list of author information is available at the end of the article



(continuous), and average number (continuous) of cigarettes smoked per day.

Among the 373,283 ever smokers in our sample, the proportion of missing values was 5% ($n = 18,886$) for smoking duration, number of cigarettes per day, and pack-years (i.e., number of cigarettes smoked per day, divided by 20, multiplied by the smoking duration in years).

We categorized current smokers according to smoking duration in years (1–19, 20–29, ≥ 30), number of cigarettes smoked per day (1–10, 11–20, > 20), and pack-years (1–9, 10–19, ≥ 20).

We classified participants by level of education into three categories: < 10 , 10–12, and ≥ 13 years by using the most recent information regarding duration of education obtained from Statistics Norway. We classified for physical activity in three: [sedentary (reading, watching television, and sedentary activity), moderate (walking, bicycling, and/or similar activities ≥ 4 h per week), and heavy (light sports or heavy gardening ≥ 4 h per week, heavy exercise, or daily competitive sports)] categories. We calculated BMI as weight in kg divided by height in m^2 and classified in three and classified in three (< 18.5 kg/m^2 , 18.5–24.9 kg/m^2 , ≥ 25.0 kg/m^2) categories. All variables were obtained at study enrollment. As questions on alcohol consumption were only included from 1994 onwards, information on alcohol consumption was missing in 73% of the participants in the analytical cohort.

Follow-up and endpoints

The data were linked to the Cancer Registry of Norway, the Norwegian Cause of Death Registry, and the Central Population Register by the national, unique 11-digit personal identification number. Lung cancer mortality was classified according to the eighth, ninth and tenth revisions of The International Classification of Diseases (ICD-8, ICD-9, ICD-10). Follow-up ended at the time of death from primary lung cancer, death from any other causes, emigration, or the end of follow-up (December 31, 2013), whichever occurred first.

All deaths connected to primary incident carcinomas of the trachea, bronchus, and lung (ICD-8 code 162 or corresponding codes from ICD-9 and ICD-10) were included as endpoint, i.e. death from lung cancer.

Statistical analysis

We calculated the age-standardized (European Standard Population) overall lung cancer mortality rate by smoking status, and categories of education [17].

We used Cox proportional hazards model with attained age between cohort entry and exit as the underlying time scale to estimate the multivariable-adjusted hazard ratios (HRs) with 95% confidence intervals (CI),

for the associations between different measures of smoking exposure and lung cancer mortality. We used stratified Cox models by cohort study and birth cohort (≤ 1950 and > 1950) to overcome any probable heterogeneity for these variables. A priori we considered alcohol, physical activity, BMI and education as possible confounders. We tested for interaction between smoking status and sex, and between smoking and education, and decided to stratify by sex and by education. We decided to adjust on BMI and physical activity, but did not include alcohol as a covariate because of a lot of missing data. We estimated dose-response associations among current smokers for the following variables measured continuously: smoking duration in 10 years, number of 10 cigarettes smoked per day, number of 10-pack-years, and lung cancer mortality overall. Never smokers were not included in these analyses.

Subsequently, we tested for linear trend for smoking exposure (smoking duration, cigarettes smoked per day and pack-years) among current smokers based on the median value in each category, using the lowest category of each measure of smoking exposure as reference.

We used the Wald test to assess heterogeneity by sex and by education for the associations between different measures of smoking exposure and lung cancer mortality. We tested and found that the criteria for the proportional hazard assumption were met using Schoenfeld residuals (data not shown).

Subsequently, we performed the same analyses after excluding individuals who died from lung cancer within < 2 years of follow-up, and we also performed the same analyses after excluding participants with prevalent cancer.

We conducted all analyses using STATA version 14.0 (Stata Corp.). We considered two-sided p -values of < 0.05 as statistically significant.

Results

During the nearly 12 million (48% men) person-years of observation and an average of 19 years of follow-up, we identified 5702 (58% men) lung cancer deaths. Altogether 39% were current, 24% former and 37% never smokers at enrollment. The majority (55%) of participants had from 10 to 12 years of education, 23% had less than 10 years, and 22% had 13 years or more. The overall crude LC mortality rate was 6.1 per 100,000 among never, 23.9 per 100,000 among former and 99.2 per 100,000 among current smokers. The corresponding rates for those at the lowest, middle and highest level of education was 87.6 per 100,000, 38.7 per 100,000 and 20.4 per 100,000, respectively. There was an interaction between smoking and sex ($P < 0.0001$), and a borderline interaction between smoking and education ($P = 0.06$).

Table 1 shows that compared with women, men were more likely to be ever (current or former) smokers, and to have smoked more pack-years for all three levels of education. The proportion of never smokers were 41% in women and 33% in men. More men (23%) than women (20%) were in the highest level of education. Women with the longest education had the highest (57%) proportion of never smokers. Among both men and women the number of lung cancer deaths was highest in the less educated (Table 1).

Additional file 1: Table S1 shows that the mean age at enrollment was 40, 43 and 48 in the Norwegian Counties Study, the 40 years Study and Cohort of Norway (CONOR) respectively. The Norwegian Counties Study was characterized by lower level of education and higher proportion of current smokers than the 40 years study and CONOR (Additional file 1: Table S1).

Table 2 shows that compared with sex-specific never smokers, current smokers had a lung cancer mortality hazard ratio of 20.05 (95% CI 16.25–24.74) for men, and

13.97 (95% CI 11.98–16.29) for women ($P_{\text{heterogeneity}} = 0.01$). For each 10-years increase in smoking duration women had a 65% higher hazard ratio [HR: 1.65 (95% CI 1.54–1.78)] and men a 36% higher HR [HR: 1.36 (95% CI 1.28–1.44)] ($P_{\text{heterogeneity}} < 0.01$). For women compared with men, current smokers had a greater increase in lung cancer mortality per unit of number of cigarettes per day and number of pack-years (Both $P_{\text{heterogeneity}} < 0.01$) (Table 2).

Additional file 2: Table S2 shows the multivariable HR for lung cancer mortality according to categorical measures of smoking exposure for current smokers by sex compared with sex specific never smokers. The estimates did not vary much by sex, except that men who had smoked < 20 years, had a higher HR [HR: 11.78 (95% CI 9.26–14.98)] compared with women [HR: 7.29 (95% CI: 6.05–8.78)] ($P_{\text{heterogeneity}} < 0.01$). For those who had smoked less than 10 pack-years, men had a higher HR compared with women ($P_{\text{heterogeneity}} = 0.02$) (Additional file 2: Table S2).

Table 1 Characteristics of the study population by education, the Norwegian Health Screening Surveys, 1974–2003, ($N = 595,675$)

Characteristics	Education in years							
	< 10		10–12		≥ 13		All	
	Men	Women	Men	Women	Men	Women	Men	Women
Subjects (%)	64,024 (22)	76,455 (25)	155,905 (55)	169,949 (55)	66,332 (23)	63,010 (20)	286,261 (48)	309,414 (52)
Lung cancer cases ^a , n (%)	1646 (44)	1338 (47)	1759 (46)	1297 (45)	385 (10)	238 (8)	3790	2873
Lung cancer deaths, n (%)	1517 (46)	1138 (48)	1473 (44)	1056 (44)	333 (10)	185 (8)	3323	2379
Person-years of follow up	1,365,688	1,666,446	3,106,850	3,429,805	1,314,443	1,224,279	5,786,981	6,320,530
Body mass index (mean, SD)	26 (3)	25 (4)	26 (3)	25 (4)	25 (3)	24 (4)	26 (3)	25 (4)
Heavy physical activity ^b (%)	28	14	36	22	41	28	35	20
Never smokers (%)	20	33	32	38	50	57	33	41
Former smokers (%)	27	18	27	22	26	23	27	21
Current smokers (%)	53	49	41	40	24	20	40	38
Duration of smoking ^c , years, median (interquartile range)	22 (15–26)	20 (15–25)	20 (13–25)	19 (10–23)	18 (10–22)	15 (8–20)	20 (13–25)	20 (10–23)
Cigarettes smoked per day ^c , median (interquartile range)	15 (10–20)	10 (8–15)	15 (10–20)	10 (7–15)	12 (10–20)	10 (5–15)	15 (10–20)	10 (7–15)
Pack-years ^c , median (interquartile range)	14 (8–21)	10 (5–16)	13 (7–20)	9 (4–15)	10 (5–18)	6 (3–12)	13 (7–20)	9 (4–15)
Age at enrollment, median (interquartile range)	42 (40–45)	42 (41–45)	41 (40–42)	41 (40–42)	42 (41–43)	41 (40–42)	41 (40–42)	41 (40–43)
Age at lung cancer death, never smokers, median (interquartile range)	75 (62–80)	77 (68–84)	62 (52–73)	64 (57–76)	57 (55–69)	61 (55–66)	63 (54–76)	66 (59–80)
Age at lung cancer death, current smokers, median (interquartile range)	66 (60–74)	63 (57–71)	63 (57–70)	60 (55–66)	64 (58–70)	61 (56–68)	64 (58–72)	62 (56–69)

^aAt enrollment

^bHeavy physical activity: light sports or heavy gardening ≥ 4 h/week, heavy exercise or daily competitive sports

^cDuration of smoking, cigarettes smoked per day and pack-years in ever smokers

Table 2 Hazard ratios^a for lung cancer mortality according to smoking status and continuous measures of exposure

Smoking status	Cases	Men HR 95%CI	Cases	Women HR 95%CI	Heterogeneity test for men versus women P-values
Never	91	1.00 (ref)	188	1.00 (ref)	
Former	459	4.08 (3.25–5.11)	208	2.71 (2.22–3.30)	0.01
Current	2773	20.05 (16.25–24.74)	1983	13.97 (11.98–16.29)	0.01
Duration of smoking, 10-years					
^b	2761	1.36 (1.28–1.44)	1969	1.65 (1.54–1.78)	< 0.01
Cigarettes smoked per day, 10 per day					
^b	2676	1.48 (1.42–1.54)	1974	1.76 (1.66–1.86)	< 0.01
Pack-years (10 years)					
^b	2269	1.39 (1.35–1.44)	1965	1.61 (1.54–1.69)	< 0.01

^a Multivariable Hazard ratios (95% CI) adjusted for body mass index, physical activity level, all at enrollment, and level of education

^b Per 10-year increase in smoking duration, per 10-cigarettes increase number of cigarettes smoked per day, per 10 increase in pack-years, for current smokers

Table 3 shows that among never smokers, women with the lowest level of education had the highest age-adjusted lung cancer mortality rate which was (16.7 per 100,000 person-years). The highest rate was among the less educated current smokers for both men (319.0 per 100,000 person-years) and women (183.0 per 100,000 person-years). For all three levels of education, males had a higher lung cancer mortality rate compared with females for both former and current smokers (Table 3).

Table 4 shows that for male current smokers the HR did not vary for the different categories of smoking exposure when we compared those with the lowest and highest level of education (all $P_{\text{heterogeneity}} > 0.05$). For female current smokers there was a significant difference between those with < 10 years [HR: 15.85 (95% CI 12.32–20.38)] compared with those with ≥ 13 years of education [HR: 9.41 (95% CI 6.49–13.68)] ($P_{\text{heterogeneity}} < 0.01$). For female current smokers the HR in the lowest category for the three smoking exposures (duration of smoking, cigarettes smoked per day and pack-years) were significantly higher when we compared those with the lowest and highest level of education (all $P_{\text{heterogeneity}} < 0.02$) (Table 4).

The results did not change substantially when we excluded individuals who died from lung cancer within <

Table 3 Age adjusted^a lung cancer mortality rates per 100,000 person-years by education and smoking status

Smoking status	Men			Women		
	Education in years ^b					
	< 10	10–12	≥ 13	< 10	10–12	≥ 13
Never	8,6	9,7	6,8	16,7	8,9	8,3
Former	83,8	56,7	51,0	47,2	26,5	24,6
Current	319,0	208,8	194,2	183,0	133,1	102,6

^a Age adjusted according to the European Standard Population

^b Education: < 10 years, 10–12 years, ≥ 13 years

2 years of follow-up. The results stayed the same when we excluded those with prevalent cancer at enrollment (data not shown).

Discussion

In this large Norwegian cohort study, we found that more men were current or former smokers, more were heavy smokers and more smokers had died from lung cancer, regardless of level of education, compared with women. For both men and women, those with the lowest compared with the highest level of education, were more likely to die from lung cancer regardless of smoking status. However, when we analyzed the three smoking exposure measures for current smokers as continuous variables, female smokers seem to be more likely to die from lung cancer, for increments of 10 years of smoking, 10 cigarettes/day and 10 pack-years compared with male smokers.

Our results are in line with those of other prospective cohort studies [18–21] and a meta-analysis of three prospective cohort studies [22], which have found that compared with females, males are heavier smokers and die more from lung cancer. In the present study, we observed a difference in lung cancer mortality between male and female smokers, while several other cohorts did not [18–23]. These studies did not use continuous measures for smoking exposure as we did, but rather broad categories for number of cigarettes smoked per day. Thus men may be in the upper and women in the lower part of a specific category, but still be classified as being similarly exposed. The US cohort, with 17,670 cases, found a virtually identical lung cancer mortality rate for male and female current smoker in the most recent time periods, while for the earliest cohorts they observed a higher risk for men, reflecting the differences in smoking prevalence by sex [22], and the stages of the

Table 4 Hazard ratios for lung cancer mortality in current smokers, by smoking exposure and education

Men ^a HRs 95% CI							
Smoking status	Education in years						Heterogeneity test ^b P-values
	Cases	< 10 years HR ^a 95% CI	Cases	10–12 years HR ^a 95% CI	Cases	≥13 years HR ^a 95% CI	
Never smokers ^c	18	1.00 (ref)	54	1.00 (ref)	19	1.00 (ref)	
Current smokers	1303	28.96 (18.17–46.14)	1216	16.01 (12.19–21.05)	254	22.50 (14.09–35.92)	0.45
Duration of smoking (years)							
1–19	120	18.27 (11.08–30.12)	130	9.13 (6.61–12.61)	27	9.95 (5.49–18.03)	0.13
20–29	717	27.93 (17.43–44.74)	758	15.61 (11.80–20.66)	151	23.56 (14.51–38.25)	0.62
> 30	465	32.72 (20.35–52.62)	318	19.92 (14.65–27.09)	75	35.32 (19.59–63.69)	0.84
P for trend ^d		< 0.01		< 0.01		< 0.01	
Cigarettes smoked per day							
1–10	400	20.71 (12.91–33.23)	262	8.98 (7.00–12.05)	53	12.26 (7.25–20.74)	0.15
11–20	684	34.57 (21.62–55.28)	710	19.00 (14.38–25.10)	142	27.96 (17.27–45.29)	0.54
> 21	165	54.57 (33.46–89.00)	211	33.76 (24.96–45.65)	49	50.24 (29.37–85.93)	0.82
P for trend ^d		< 0.01		< 0.01		< 0.01	
Pack-years							
1–9	141	15.84 (9.68–25.92)	101	6.49 (4.65–9.05)	19	7.34 (3.88–13.91)	0.06
10–19	497	25.45 (15.88–40.78)	443	13.55 (10.20–18.00)	87	21.18 (12.85–34.89)	0.60
≥ 20	611	40.43 (25.28–64.65)	632	23.84 (18.04–31.54)	138	37.02 (22.78–60.17)	0.80
P for trend ^d		< 0.01		< 0.01		< 0.01	
Women ^a HRs 95% CI							
Smoking status	Cases	HR ^a 95% CI	Cases	HR ^a 95% CI	Cases	HR ^a 95% CI	
Never smokers	70	1.00 (ref)	81	1.00 (ref)	37	1.00 (ref)	
Current smokers	980	15.85 (12.32–20.38)	887	14.22 (11.28–17.92)	116	9.41 (6.49–13.68)	< 0.01
Duration of smoking (years)							
1–19	155	8.37 (6.21–11.29)	158	7.62 (5.78–10.06)	20	3.83 (2.20–6.65)	0.01
20–29	624	16.11 (12.32–21.08)	603	16.19 (12.70–20.63)	73	11.01 (7.35–16.48)	0.12
> 30	194	23.05 (17.16–30.97)	120	21.85 (15.79–30.25)	22	27.18 (13.30–55.52)	0.68
P for trend ^d		< 0.01		< 0.01		< 0.01	
Cigarettes smoked per day							
1–10	458	12.81 (9.87–16.62)	326	9.43 (7.37–12.07)	36	5.15 (3.26–8.20)	< 0.01
11–20	465	22.88 (17.50–29.92)	500	21.35 (16.75–27.22)	78	14.47 (9.66–21.69)	0.06
> 21	54	41.62 (28.75–60.25)	57	39.87 (28.22–56.34)	8	19.70 (9.04–42.97)	0.09
P for trend ^d		< 0.01		< 0.01		< 0.01	
Pack-years							
1–9	197	8.31 (6.27–11.02)	148	6.08 (4.61–8.01)	18	3.31 (1.88–5.84)	< 0.01
10–19	481	18.27 (14.05–23.77)	457	16.95 (13.31–21.60)	57	12.00 (7.89–18.25)	0.10
≥ 20	294	29.66 (22.60–38.93)	274	27.92 (21.66–35.98)	39	18.49 (11.68–29.26)	0.08
P for trend ^d		< 0.01		< 0.01		< 0.01	

^a Multivariable Hazard ratios (95% CI) adjusted for body mass index and physical activity, both at enrollment^b Heterogeneity test for those with < 10 years of education compared with ≥13 years education^c Never smokers^d Trend test without never smokers

smoking epidemic by sex described earlier [10]. Since lung cancer mortality rates for Norwegian women have not peaked yet, they may become higher than that for the US women, which already in 2001 was warned by Tverdal [8]. Jha et al. [24], have pointed out that the full effects of smoking can take 50 years to measure in individuals, and up to 100 years to measure in populations. The results from the present study and from that of Tverdal [8], both showing sex differences in Norway, may be early indicators of this long-term development of sex differences in smoking related lung cancer mortality. Other indicators that the sex difference in smoking related lung cancer mortality in the long-term effect of smoking are our [15], and those of the US cohort [9]. An alternative explanation for the higher lung cancer mortality in smoking females compared with men in our study may be competing risk of death. Since men smoke more than women, they have increased risk for dying of other smoking-related diseases before they get lung cancer.

In Norway, there is a marked social gradient for active as well as passive smoking. The lower the education, the more smoking [4]. As expected, the age standardized rates of lung cancer mortality were highest in the less educated male and female current smokers. For both men and women, our results indicate that the less educated had a higher lung cancer mortality compared with the highly educated. The difference by level of education for both men and women should be interpreted with caution, as this could be due to residual confounding by smoking as there was a large proportion of heavy smoking men and women, in the less educated. Another explanation for smoking related difference in lung cancer mortality by both sex and education could be related to measures of socioeconomic status like passive smoking from spouses, radon, occupational exposure and air pollution. Similarly, studies from the EPIC (European Prospective Investigation into Cancer and Nutrition) and Sweden, respectively, observed a higher risk of lung cancer in the lower social class despite accounting for smoking habits [13, 14].

Among never smokers, we observed that both men and women in the lowest level of education died more from lung cancer compared with their counterparts in the highest level of education. A possible explanation may be residual confounding by smoking as well as exposure to occupational and passive smoking exposure.

Our study has several major strengths. It is based on a large, prospective Norwegian cohort, comprising a high proportion of male and female ever smokers, with long, virtually complete follow-up. The questions on smoking duration and number of cigarettes per day allowed respondents to give open-ended answers which allowed us to utilize continuous measures of smoking exposure.

Moreover, we have more than 5500 lung cancer deaths, yielding higher precision of the estimates and power to discover a true difference.

One limitation is that we only have information on smoking and other potential confounders at study enrollment. Another limitation is that we lack information on passive and occupational smoking.

Around 10% of the Norwegian population reported to be occasional smokers in our follow-up period [25]. Some of them may have been included as never smokers, which most likely will have attenuated the observed associations between smoking and lung cancer death. We do not believe that these limitations would distort the smoking related sex difference in lung cancer mortality revealed in our study.

Conclusion

Our findings, in this large cohort study, suggest that women have increased risk of dying from lung cancer compared with men, given the same smoking history. In addition, low education confers an increased risk of dying from lung cancer, which could be due to residual confounding by active and passive smoking.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12885-019-6330-9>.

Additional file 1: Table S1. Selected characteristics of the study population at enrollment, stratified by cohort, (N = 595,675).

Additional file 2: Table S2. Hazard ratios^a (95% CIs) for lung cancer mortality according to categorical measures, for current smokers.

Acknowledgements

The authors acknowledge the services of CONOR, the contributing research centers delivering data to CONOR, and all the study participants.

Authors' contributions

Conception and design: MSH, IL, ITG. Development of methodology: MSH, IL, ITG. Statistical analysis and interpretation of data: MSH, IL, TB, AL, LLM, ITG. Writing, review, and revision of the manuscript: MSH, IL, TB, AL, LLM, ITG. All authors have read and approved the manuscript.

Funding

The cost of publishing this manuscript was paid by Northern Norway Health Authority. MSH was supported by Northern Norway Health Authority (SFP1227–15). IL was supported by Norwegian Cancer Society (Grant agreement number: 4510766–2013). Neither of the funding bodies Northern Norway Health Authority (SFP1227–15) and Norwegian Cancer Society (Grant agreement number: 4510766–2013), have any role in the design of the study, collection, analysis, nor interpretation of data nor in writing the manuscript.

Availability of data and materials

The dataset used during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

The present study was approved by the Regional Committee for Medical Research Ethics South-East, Norway.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interest.

Author details

¹Department of Community Medicine, Faculty of Health Sciences, The UiT Arctic University of Norway, Tromsø, Norway. ²Clinical Research Department, Centre François Baclesse, Caen, France. ³Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Trondheim, Norway. ⁴Cancer Epidemiology Program, University of Hawai'i Cancer Center-University of Hawai'i at Manoa, Honolulu, HI, USA.

Received: 11 March 2019 Accepted: 4 November 2019

Published online: 21 November 2019

References

- Steward BW, Wild CP. World Cancer Report 2014: International Agency for Research on Cancer World Cancer Report; 2014.
- Lund KE, Lund M, Bryhni A. Tobacco consumption among men and women 1927-2007. *Tidsskr Nor Laegeforen*. 2009;129(18):1871-4.
- Lund I, Lund KE. Lifetime smoking habits among Norwegian men and women born between 1890 and 1994: a cohort analysis using cross-sectional data. *BMJ Open*. 2014;4(10):e005539.
- Næss Ø, Rognerud M, Strand BH. Sosial ulikhet i helse. En faktarapport FHI. Oslo Folkehelseinstituttet. 2007.
- Cancer Registry of Norway. Cancer in Norway 2013. Oslo, Norway: Cancer Registry of Norway 2013 [updated March, 2015; cited 2016 03.10]. Available from: <https://www.ssb.no/en/helse/statistikker/royk/aar/2014-02-05>.
- Islami F, Torre LA, Jemal A. Global trends of lung cancer mortality and smoking prevalence. *Translational lung cancer research*. 2015;4(4):327-38.
- National Center for Chronic Disease P, Health Promotion Office on S, Health. Reports of the Surgeon General. The Health Consequences of Smoking-50 Years of Progress: A Report of the Surgeon General. Atlanta (GA): Centers for Disease Control and Prevention (US); 2014.
- Tverdal A. Lung cancer mortality—now higher in women than in men under 50 years. *Tidsskr Nor Laegeforen*. 2001;121(21):2487-8.
- Jha P, Ramasundarahettige C, Landsman V, Rostron B, Thun M, Anderson RN, et al. 21st-century hazards of smoking and benefits of cessation in the United States. *N Engl J Med*. 2013;368(4):341-50.
- Thun M, Peto R, Boreham J, Lopez AD. Stages of the cigarette epidemic on entering its second century. *Tob Control*. 2012;21(2):96-101.
- Hart CL, Hole DJ, Gillis CR, Smith GD, Watt GC, Hawthorne VM. Social class differences in lung cancer mortality: risk factor explanations using two Scottish cohort studies. *Int J Epidemiol*. 2001;30(2):268-74.
- Braaten T, Weiderpass E, Lund E. Socioeconomic differences in cancer survival: the Norwegian women and Cancer study. *BMC Public Health*. 2009;9:178.
- Ekberg-Aronsson M, Nilsson PM, Nilsson JA, Pehrsson K, Lofdahl CG. Socio-economic status and lung cancer risk including histologic subtyping—a longitudinal study. *Lung Cancer*. 2006;51(1):21-9.
- Menvielle G, Boshuizen H, Kunst AE, Dalton SO, Vineis P, Bergmann MM, et al. The role of smoking and diet in explaining educational inequalities in lung cancer incidence. *J Natl Cancer Inst*. 2009;101(5):321-30.
- Hansen MS, Licaj I, Braaten T, Langhammer A, Le Marchand L, Gram IT. Sex differences in risk of smoking-associated lung Cancer: results from a cohort of 600,000 Norwegians. *Am J Epidemiol*. 2017.
- Stocks T, Borena W, Strohmaier S, Bjorge T, Manjer J, Engeland A, et al. Cohort profile: the metabolic syndrome and cancer project (me-can). *Int J Epidemiol*. 2010;39(3):660-7.
- Geography PDT. Standard populations. In: Scotland NS, editor. *ISD Scotland: Scotland*; 2018.
- Vollset SE, Tverdal A, Gjessing HK. Smoking and deaths between 40 and 70 years of age in women and men. *Ann Intern Med*. 2006;144(6):381-9.
- Nilsson S, Carstensen JM, Pershagen G. Mortality among male and female smokers in Sweden: a 33 year follow up. *J Epidemiol Community Health*. 2001;55(11):825-30.
- Jamrozik K, McLaughlin D, McCaul K, Almeida OP, Wong KY, Vagenas D, et al. Women who smoke like men die like men who smoke: findings from two Australian cohort studies. *Tob Control*. 2011;20(4):258-65.
- Prescott E, Osler M, Andersen PK, Hein HO, Borch-Johnsen K, Lange P, et al. Mortality in women and men in relation to smoking. *Int J Epidemiol*. 1998; 27(1):27-32.
- Thun MJ, Carter BD, Feskanich D, Freedman ND, Prentice R, Lopez AD, et al. 50-year trends in smoking-related mortality in the United States. *N Engl J Med*. 2013;368(4):351-64.
- Marang-van de Mheen PJ, Smith GD, Hart CL, Hole DJ. Are women more sensitive to smoking than men? Findings from the Renfrew and Paisley study. *Int J Epidemiol*. 2001;30(4):787-92.
- Jha P. Avoidable global cancer deaths and total deaths from smoking. *Nat Rev Cancer*. 2009;9(9):655-64.
- Norwegian Institute of Public Health. Public Health Report 2014 Oslo2014 [updated 12.04.2016; cited 2016 03.10]. Available from: <https://www.fhi.no/en/>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions



Merethe S. Hansen, Ildir Licaj, Tonje Braaten, Eiliv Lund, Inger Torhild Gram

The fraction of lung cancer attributable to smoking in the Norwegian Women and Cancer (NOWAC) study

Br J Cancer (Submitted)

The fraction of lung cancer attributable to smoking in the Norwegian Women and Cancer (NOWAC) study

Authors

Merethe S. Hansen¹, Ildir Licaj¹, Tonje Braaten¹, Eiliv Lund¹, Inger Torhild Gram¹

Institutions

¹Department of Community Medicine, Faculty of Health Sciences, The UiT Arctic University of Norway, Tromsø, Norway;

Key words

Lung cancer, smoking, passive smoking, PAF, NOWAC study

Words count abstract

194

Main text (without tables and refs)

2980

References

32

Abstract

Background

Few prospective cohort studies have been able to study both active and passive smoking, and lung cancer risk among women. We examined the association between active and passive smoking and lung cancer risk. Subsequently, we estimated the population attributable fraction (PAF) of lung cancer due to active smoking.

Method

We followed 142 508 women through linkages to national registries through December 2015. We used Cox proportional hazards models, to estimate hazard ratios (HRs) with 95% confidence intervals (CIs). We calculated PAF to indicate what proportion of lung cancer cases could have been prevented in the absence of smoking.

Results

We identified 1507 lung cancer cases during 15.9 years of follow-up. Compared with never smokers, current (HR 13.88, 95% CI 10.18-18.91) and former (HR 3.69, 95% CI 2.68-5.09) smokers had significantly increased risk of lung cancer. Female never smokers exposed to passive smoking had a 1.3-fold (HR 1.34, 95% CI 0.89-20.01) non-significantly increased risk of lung cancer, compared with never smokers. The PAF of lung cancer was 85.3% (95% CI 80.0-89.2).

Conclusion

More than eight in ten lung cancer cases could have been avoided in Norway, if the women did not smoke.

BACKGROUND

Lung cancer is the third most common cancer in women worldwide ¹. Active smoking is the main cause of lung cancer. The incidence of lung cancer among women in Western Europe has stagnated in recent years ². In Norway, the incidence rate for lung cancer in women decreased in 2017. If this reduction continues, 2015 will be the top year with the highest incidence of lung cancer among Norwegian women, with a rate of 55.4 per 100 000 ³.

Despite declines in overall smoking prevalence, especially in high-income countries, there are alarming trends in tobacco use among youth, especially females, in low and middle income countries ⁴. Although, in developing countries it is not easy to predict whether or when women will begin smoking in large numbers ⁵. A potentially scenario, is that women in these countries will experience the same trends in lung cancer incidence that have been reported in Norway.

Among Norwegian women, the smoking prevalence increased sharply from 1920 and towards 1950 when 20% of women were smokers ⁶. The prevalence of daily smoking in Norwegian women peaked at 37% around 1970 ⁶. Today, 12% of Norwegian women are daily smokers ⁷. However, lung cancer is also a significant health problem among those exposed to passive smoking ^{8,9}. In Northern Europe, approximately 26% of lung cancer cases occur in women who have never smoked ¹⁰. The IARC Monograph from 2004, The US Surgeon General and the latest World Cancer Report state that exposure to involuntary smoking increases lung cancer risk by about ^{1, 11, 12}. Other known risk factors for lung cancer are exposure to asbestos, radon, polycyclic aromatic hydrocarbons, and emissions from household combustion of coal ².

Few prospective cohort studies have been able to study both active and passive smoking, and the lung cancer risk they confer among women.

We utilized the Norwegian and Cancer Study, a nationally representative prospective cohort, to estimate the fraction of lung cancer attributable to active and passive smoking and the number of lung cancer cases that could have been avoided in the absence of smoking in Norway in 2015.

MATERIALS AND METHODS

Study population

The Norwegian Women and Cancer Study cohort profile has been previously described in detail ¹³. Briefly, the Central Population Register selected a random sample of women according to year of birth. Subsequently, an invitation to participate in the study together with a baseline questionnaire and a pre-stamped return envelope enclosed was mailed to each woman. All women gave informed consent (<https://site.uit.no/nowac/>). The National Data Inspectorate and the Regional Committee for Medical Research Ethics approved the study.

Women who completed a questionnaire during three waves of data collection: 1991-1992, 1996-1997, and 2003-2007 (172 478), were included. The overall response rate was 52.7%. We excluded women with prevalent cancer (n= 6664), those who emigrated (n=64), or died before the start of follow-up (n=10), those with an age at exit that was below age at recruitment (n=32), those with missing information on smoking status (n=590), and never smokers with missing information on passive smoking (n=10 879). Finally, we excluded those with missing information the covariates as education and alcohol consumption (n= 11 731). Altogether, 349 women with lung cancer were excluded in this process. The cohort comprised 142 508 women.

Exposure information

The questionnaire included detailed assessment of smoking habits, if their parents smoked during childhood, if they lived with a smoker as adults, physical activity, alcohol consumption as well as height and current weight (which were used to compute body mass index (BMI) as weight in kilograms divided by the square of height in meters). The questionnaires asked if the women had ever been smoking, and those answering “yes” were asked for the number of cigarettes smoked daily at different age intervals. Subsequently, they were asked if they smoked on a daily basis at present. We categorized ever smokers according to current and former smoking status, age at smoking initiation, smoking duration, average number of cigarettes smoked daily, pack-years of smoking (i.e., number of cigarettes smoked per day, divided by 20, multiplied by the number of years smoked), all at enrolment. Former smokers were classified according to years since quitting smoking. All women who were neither current nor former smokers were classified as never smokers. Among never smokers, those who reported that their parents smoked during childhood or they lived with a smoker as adults were classified as passive smokers. We calculated average alcohol consumption in g/day among drinkers based on the content of pure alcohol in different beverages and usual portion sizes in Norway.

Follow up

We followed the women through linkages to the Cancer Registry of Norway and the Norwegian Central Population Register to identify all cancer cases, and emigrations and deaths, respectively, using the unique 11-digit national personal identification number. We calculated person-years from the start of follow-up to any incident cancer diagnosis (except basal cell carcinoma), emigration, death or the end of follow-up (31 December 2015), whichever came first. We classified lung cancer cases according to the original codes in the International Classification of Diseases, Seventh Revision.

Statistical analysis

We calculated the age standardized (WHO 2000-2025) incidence rate of lung cancer overall by smoking status ¹⁴.

We stratified all models by recruitment sub-cohort (1991-1992, 1996-1997 and 2003-2007) to control for potential differences at the three recruitment waves. We used the Cox proportional hazards model, with age as the underlying time scale, to estimate crude and multivariate-adjusted hazard ratios (HRs) and 95% confidence intervals (CIs) for the associations between lung cancer and different measures of smoking exposure. Smoking exposure was defined using smoking status at cohort entry (never, passive, former, current or ever); for women who had ever smoked, further exposures were defined using smoking duration (1-9, 10-19, 20-29, ≥ 30 years), number of cigarettes smoked per day (1-9, 10-19, ≥ 20), number of pack-years of smoking (1-5, 6-15, > 15) age at smoking initiation (≤ 17 , > 17 -20, > 20) years), and for former smokers: years since quit smoking (1-9, 10-20, > 20 years). The reference group is never active, never passive smokers throughout the manuscript, unless otherwise noted.

We included covariates that changed the HR estimate by at least 5% as confounders of the association between smoking and lung cancer. We included the following variables in the final multivariable models; age at enrollment, years of education (< 10 , 10-12, ≥ 13 years) and average alcohol consumption, in grams of alcohol per day (0, ≤ 4 , 5-9, ≥ 10). Women who reported to be teetotalers and those answering “seldom or never” had their alcohol consumption set to 0.

We tested for linear trends across categories of smoking exposure variables, excluding never smokers, by assigning a median score in order to account for the distance between categories, treating the variable as continuous in the analysis. We tested and found that the criteria for the proportional hazards assumption were met using Schoenfeld residuals (data not shown).

We calculated PAFs (%) to indicate what proportion of lung cancer cases could have been prevented in the population in absence of smoking. We used the formula $PAF = P_e (RR_e - 1) \div (P_e \times RR_e + [1 - P_e])$, where the notation P_e = the proportion of persons in the population exposed to the risk factor (i.e., ever smokers) and RR_e = the relative risk in the exposed compared with the unexposed group estimated through the HRs (ever compared with never active never passive smokers)¹⁵. We calculated the two-sided 95% CIs for the PAFs using the PUNAF Stata module¹⁶.

We performed analyses using STATA version 15.0 (Stata Corp.) and considered two-sided P-values ≤ 0.05 as statistically significant.

RESULTS

During the more than 2.3 million person-years of observation (mean follow-up time was 15.9 (± 6.5) years, we ascertained 1507 lung cancer cases. At enrollment, the mean age of the participants was 48.9 (± 8.3 s.d) years. Of these women, 69.3% (n = 98 795) reported to be ever (current 32.2%, former 37.1%) smokers, with 14.8 (± 12.9 s.d) years of smoking. Former smokers quit smoking 16.9 (± 9.8 s.d) years ago. Among never smokers 17.6% (n = 25 082) reported to be passive smokers. Altogether 77.9% (n=114 191) of the women reported to be alcohol drinkers, with a mean alcohol consumption of 3.9 (± 6.1 s.d) grams per day. Mean years of education was 12.4 (± 3.5) among the women. The age-standardized incidence rate for lung cancer among never, passive, and ever smokers was 13.4, 20.0, and 87.1 per 100 000 person-years respectively.

Table 1 shows that current compared with passive and never smokers were younger at enrolment and at lung cancer diagnosis. Current smokers were also less likely to have higher education, more likely to have lower BMI as well as more likely to consume alcohol.

Adenocarcinoma was the most common histological lung cancer type in both never and ever

smokers, while squamous cell- and small cell carcinoma were most common in current smokers and almost absent in never smokers (Table 1).

Table 2 shows that when compared with never smokers, current smokers had a lung cancer hazard ratio that was 14-fold (HR 13.88, 95% CI 10.18-18.91) higher and former smokers a hazard ratio that was 4-fold (HR 3.69, 95% CI 2.68-5.09) higher, with current also having a significantly higher risk of lung cancer than former smokers. For ever smokers there was a dose response for the different smoking exposures (smoking duration, cigarettes smoked per day and pack-years, all p-trend <0.001). The highest hazard ratios for lung cancer among ever compared with never smokers, was a 21 fold increase in risk of lung cancer among those who smoked > 20 cigarettes per day (HR 21.66, 95% CI 16.54-28.37) and for those who smoked > 15 pack-years (HR 21.24, 95% CI 15.52-29.06). Female never smokers exposed to passive smoking had a hazard ratio of 1.34 (95% CI, 0.89-2.01) compared with never smokers without any passive exposure.

Table 3 shows that compared with current smokers, former smokers who had quit smoking 1-9 years ago had a 63% (HR 0.37, 95% CI 0.30-0.45) reduced risk, while those who had quit smoking > 20 years ago had a 89% (HR 0.11, 95% CI 0.08-0.15) reduced risk of lung cancer (all p-trend <0.001).

The PAF of lung cancer was 85.3% (95% CI 80.0-89.2) for ever smoking.

DISCUSSION

In this cohort of middle-aged Norwegian women, we found that passive, former, and current smokers had a higher lung cancer risk than never smokers. Current smokers showed consistently higher risk estimates than former smokers. Moreover, more than eight in 10 lung cancers could have been prevented if the women did not smoke.

The age-standardized incidence rates for lung cancer that we found for never smokers not exposed to passive smoking is in accordance with the 14.3 per 100 000 reported among never smokers not exposed to passive smoking in “The UK Million Women Cancer Study”¹⁷.

We found adenocarcinoma to be the most frequent histological subtype in both ever and never smokers, in accordance with other prospective cohort studies¹⁷⁻¹⁹.

Few prospective cohort studies have been able to study both active and passive smoking, and the lung cancer risk they confer in women. Our study confirms a consistent, strong dose-response relationship between smokers and lung cancer risk, as has been previously shown in other cohort studies^{18,20}. As our former smokers had lower risk of lung cancer than current smokers our results confirms prior findings¹⁸. Those who quit smoking within the last 9 years had a 63% lower risk of lung cancer compared with current smokers, which are consistent with results from a Norwegian cohort study reporting a 75% lower risk in female former compared with female current smokers¹⁹. The risk of lung cancer for former smokers decreased with number of years since quit smoking. The excess risk of lung cancer in the middle aged women in our study decreased almost to the level of a never smoker 20 years after quitting. Another cohort study, with 89 000 female participants of whom 144 former smokers were diagnosed with lung cancer, observed that the risk of lung cancer decreased to almost the level of a never smoker only 15 years after quit smoking²⁰. The inconsistency between the two studies, regarding the reduced risk of lung cancer as years since quit smoking go by is most likely distorted because of few cases in former smokers in both studies.

The magnitude of our estimate for lung cancer risk due to passive smoking was higher than those reported in two expert reports^{11,12}. We had less than 100 lung cancer cases among never smokers, and of these 50 occurred among passive smokers. The lack of power in our study, and in most other studies that have examined the association between passive smoking and lung cancer risk, is obvious. A 2018 meta-analysis included seven cohort studies, each of

which reported between 11 and 136 lung cancer cases among passive smokers ²¹. Similarly, the estimates for lung cancer risk among passive smokers varied from 1.2 in the study with 70 lung cancer cases to 1.9 in the study with 79 cases, and the relative risk was statistically significant in only one study from Korea, published in 1999. It included close to 160 000 female participants, and 79 lung cancer cases among women exposed to passive smoke from their spouse ²². The 2005 European study included in the meta-analysis comprised 95 947 women, among whom 70 were diagnosed with lung cancer; it did not find any significantly increased risk of lung cancer in women exposed to passive smoke at home ²³. In the “UK Million Women Study”, with follow-up through 2011, there were 1469 cases among the 634 039 never smokers, but exposure to passive smoke as a child and/or as adult was not associated with any significantly increased risk of lung cancer ¹⁷. Likewise a cohort study from the US, with 76 304 participants and follow-up until 2009, and 152 lung cancer cases among never smokers, did neither find any significant association between any passive smoking and lung cancer risk ¹⁸. We found that, compared with active smoking, passive smoking has a much weaker association with lung cancer risk. The most likely explanation as to why we did not find any significant association between passive smoking and lung cancer risk is the lack of cases. However, we can not exclude that there are no association, which is supported by the “UK Million Women Study” ¹⁷.

More than eight in 10 lung cancer cases could have been avoided if the women did not smoke. The PAF in our study is higher than the corresponding value for US white women estimated to be 78% in a prospective cohort study, published in 2014, with 14 610 participants, and 173 lung cancer cases among the 5487 white women in the study ²⁴. Our PAF value is in accordance with the PAF for ever smoking in UK women, which was estimated to be 80% in 2010 ²⁵. To our knowledge, we are the first to report this high PAF for women based on

individual data from a random nationwide sample. It shows that former and current smoking is causing more lung cancer among women than previously anticipated.

1

Strengths and limitations of the study

The most important strength of our study is that it is a nationally representative cohort study allowing us to calculate the PAF of lung cancer due to smoking for middle-aged women. We know from our previous studies²⁶⁻²⁹ that the smoking exposure and the cancer incidence¹³ reflect known smoking patterns⁶ and cancer incidence³⁰ for Norwegian women. Thus, we are confident that our cohort is representative of the Norwegian female population, born between 1927 and 1965, both according to exposure and outcome of this study. Other strengths of our study are that we have a high proportion of both current and former smokers, virtually complete follow-up through the National population based registries. The 30-year lag period between smoking initiation and time of cohort enrolment for the majority of smokers resulted in a large number of cases, which gave us the ability to examine the dose response association and risk in former smokers. Another force is that we focus our PAF estimates on the comparison between ever versus never smokers. Thus, only never smokers could possibly change smoking status during follow-up. Since very few Norwegians start to smoke after the age of 30 and the mean age at enrolment of our study is more than 40 years, we are confident that the possible changes in smoking status among the never smokers during follow-up did not influence our PAF estimates.

Our main limitation is that we have few lung cancer cases among never smokers, resulting in a lack of statistical power.

In summary, more than eight in ten lung cancer cases among middle aged women in Norway could have been prevented if the women did not smoke. As pointed out by Jha³¹, the full effects of smoking can take up to 50 years to measure in individuals, and up to 100 years to measure in a population. This high PAF is what we found in the present study in a country which was one of the first countries to introduce restrictive rules for tobacco control when “The Norwegian Tobacco Act” entered into force in 1975³². Milestones in “The Norwegian Tobacco Act” are: Since 1988, there has been legal protection from exposure to tobacco smoke in workplaces, and since 2004, even a complete ban on smoking in bars and restaurants, and the legal age to buy tobacco was increased to 18 years in 1996. Our study shows that tobacco control campaigns and restrictions should continue to be a high priority in Western countries. In developing countries, tobacco control policies should be expanded and more restrictive measures on smoking should be implemented, so that women in these countries can avoid the experience of women in developed countries.

REFERENCES

1. Steward BW, Wild CP. World Cancer Report 2014.: International Agency for Research on Cancer World Cancer Report; 2014.
2. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin.* 2015;65(2):87-108.
3. Norway Cro. Have we passed the top of the curve for lung cancer in women? Norway: Cancer Registry of Norway; 2018 [Available from: <https://www.kreftregisteret.no/en/General/News/er-vi-forbi-lungekrefttoppen-blant-kvinner/>].
4. Drope J, Schluger N, Cahn Z, Drope J, Hamill S, Islami F, et al. The Tobacco Atlas: American Cancer Society and Vital strategies. Atlanta; 2018.
5. Thun M, Peto R, Boreham J, Lopez AD. Stages of the cigarette epidemic on entering its second century. *Tob Control.* 2012;21(2):96-101.
6. Lund I, Lund KE. Lifetime smoking habits among Norwegian men and women born between 1890 and 1994: a cohort analysis using cross-sectional data. *BMJ Open.* 2014;4(10):e005539.
7. Norway S. Tobacco, alcohol and other drugs Oslo, Norway: Statistics Norway; 2019 [updated 18.01.2019; cited 2019 29.01]. Available from: <https://www.ssb.no/en/helse/statistikker/royk>.
8. Sun S, Schiller JH, Gazdar AF. Lung cancer in never smokers--a different disease. *Nat Rev Cancer.* 2007;7(10):778-90.
9. Subramanian J, Govindan R. Lung cancer in never smokers: a review. *J Clin Oncol.* 2007;25(5):561-70.
10. Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. *CA Cancer J Clin.* 2005;55(2):74-108.
11. IARC Monogr Eval Carcinog Risks Hum. Tobacco smoke and involuntary smoking. International Agency for Research on Cancer, Lyon; 2004.
12. Surgeon General. US Department of Health and Human Services, Centers for Disease Control and Prevention, Coordinating Center for Health Promotion, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, Atlanta.; 2006.
13. Lund E, Dumeaux, V, Braaten, T, Hjartaker, A, Engeset, D, Skeie, G, et al. Cohort profile: The Norwegian Women and Cancer Study--NOWAC--Kvinner og kreft. *Int J Epidemiol.* 2008;37(1):36-41.
14. National Cancer Institute. Surveillance, Epidemiology, and End results Program: World (WHO 2000-2025) Standard. USA.
15. World Health Organization. Global report on mortality attributable to tobacco.: World Health Organization, Geneva; 2012.
16. Newson R. PUNAF: Stata Module to Compute population Attributable Fractions for Cohort Studies. Statistical Software Components S4576193. 2010.
17. Pirie K, Peto R, Green J, Reeves GK, Beral V. Lung cancer in never smokers in the UK Million Women Study. *Int J Cancer.* 2016;139(2):347-54.
18. Wang A, Kubo J, Luo J, Desai M, Hedlin H, Henderson M, et al. Active and passive smoking in relation to lung cancer incidence in the Women's Health Initiative Observational Study prospective cohort. *Ann Oncol.* 2015;26(1):221-30.
19. Hansen MS, Licaj I, Braaten T, Langhammer A, Le Marchand L, Gram IT. Sex Differences in Risk of Smoking- Associated Lung Cancer: Results From a Cohort of 600,000 Norwegians. *Am J Epidemiol.* 2017.
20. Speizer FE, Colditz GA, Hunter DJ, Rosner B, Hennekens C. Prospective study of smoking, antioxidant intake, and lung cancer in middle-aged women (USA). *Cancer Causes Control.* 1999;10(5):475-82.
21. Ni X, Xu N, Wang Q. Meta-Analysis and Systematic Review in Environmental Tobacco Smoke Risk of Female Lung Cancer by Research Type. *Int J Environ Res Public Health.* 2018;15(7).

22. Jee SH, Ohrr H, Kim IS. Effects of husbands' smoking on the incidence of lung cancer in Korean women. *Int J Epidemiol*. 1999;28(5):824-8.
23. Vineis P, Airoidi L, Veglia F, Olgiati L, Pastorelli R, Autrup H, et al. Environmental tobacco smoke and risk of respiratory cancer and chronic obstructive pulmonary disease in former smokers and never smokers in the EPIC prospective study. *BMJ*. 2005;330(7486):277.
24. Prizment AE, Yatsuya H, Lutsey PL, Lubin JH, Woodward M, Folsom AR, et al. Smoking behavior and lung cancer in a biracial cohort: the Atherosclerosis Risk in Communities study. *Am J Prev Med*. 2014;46(6):624-32.
25. Parkin DM. Tobacco-attributable cancer burden in the UK in 2010. *Br J Cancer*. 2011;105 Suppl 2:S6-s13.
26. Gram IT, Braaten T, Terry PD, Sasco AJ, Adami HO, Lund E, et al. Breast cancer risk among women who start smoking as teenagers. *Cancer Epidemiol Biomarkers Prev*. 2005;14(1):61-6.
27. Gram IT, Braaten T, Adami HO, Lund E, Weiderpass E. Cigarette smoking and risk of borderline and invasive epithelial ovarian cancer. *Int J Cancer*. 2008;122(3):647-52.
28. Gram IT, Braaten T, Lund E, Le Marchand L, Weiderpass E. Cigarette smoking and risk of colorectal cancer among Norwegian women. *Cancer Causes Control*. 2009;20(6):895-903.
29. Gram IT, Little MA, Lund E, Braaten T. The fraction of breast cancer attributable to smoking: The Norwegian women and cancer study 1991-2012. *Br J Cancer*. 2016;115(5):616-23.
30. Cancer Registry of Norway. Cancer in Norway 2013. Cancer incidence, mortality, survival and prevalence. Oslo, Norway; 2013.
31. Jha P. Avoidable global cancer deaths and total deaths from smoking. *Nat Rev Cancer*. 2009;9(9):655-64.
32. Norwegian Institute of Public Health. Tobacco Control in Norway Oslo, Norway 2019 [updated 10.01.2019. Available from: <https://www.helsedirektoratet.no/english/tobacco-control-in-norway>.

TABLES

Table 1. Selected characteristics of the study sample by smoking status, in the Norwegian Women and Cancer Study, 1991-2015, (N=142 508)

Characteristics	Never	Passive ^a	Former	Current
Subjects	18 631	25 082	52 887	45 908
Person-years of follow up	303 310	404 765	797 385	764 512
Age at enrolment, mean (±SD)	51 (9)	48 (8)	50 (8)	47 (8)
Age at diagnosis, mean (±SD)	66,6 (9,3)	63,0 (8,1)	63,3 (8,0)	63,0 (8,1)
Years since quit smoking, mean (±SD)	NA ^b	NA ^b	16.6 (9.8)	NA ^b
Education years, mean (±SD)	12,9 (3,8)	13,1 (3,6)	12,6 (3,6)	11,5 (3,1)
Physical activity level score	2,5 (2,0)	2,4 (1,8)	2,4 (1,8)	2,5 (2,19)
BMI ^c , mean (±SD)	24,4 (3,9)	24,3 (4,0)	24,6 (4,0)	23,6 (3,8)
Alcohol g/day, mean (±SD)	2,2 (3,8)	3,1 (4,6)	4,4 (6,1)	4,4 (7,2)
Lung cancer overall	43	53	329	1,082
Adenocarcinoma ^d (%)	26 (61)	34 (64)	173 (52)	459 (42)
Squamous cell carcinoma ^d (%)	1 (2)	2 (4)	35 (11)	159 (15)
Small cell lung cancer ^d (%)	2 (5)	2 (4)	45 (14)	246 (23)
Other nsclc ^{d,e} (%)	12 (28)	13 (24)	59 (18)	138 (13)
Large cell carcinoma ^d (%)	1 (2)	2 (4)	8 (2)	22 (2)
Excluded ^{d,f} (%)	1 (2)	0 (0)	9 (3)	58 (5)

^aPassive smokers include never smokers who lived with a smoker in their childhood and/or live with a smoker at enrollment of the study

^bNA (not applicable)

^cBody mass index; weight in kilograms divided by the square of the heights in metres

^dThe percentage of each histological subtype in each column is the percentage of the subtype of lung cancer overall in that column

^eOther nsclc; not specified non small cell lung carcinoma

^fExcluded; Other or not otherwise specified carcinoma

Table 2. Crude^a- and multivariate^b adjusted hazard ratio (HR) estimates for lung cancer with 95% confidence intervals (CIs) for different measures of smoking exposures

Smoking exposures	Cases (n)	HR ^a (95% CI)	HR ^b (95% CI)
Never ^c	43	1.00 (ref)	1.00 (ref)
Passive	53	1.33 (0.89-2.01)	1.34 (0.89-2.01)
Former	348	3.81 (2.77-5.23)	3.69 (2.68-5.09)
Current	1139	15.44 (11.37-20.98)	13.88 (10.18-18.91)
Ever	1464	7.67 (5.66-10.41)	6.80 (5.00-9.24)
Smoking duration in ever smokers ^d			
1-9	231	4.58 (3.29-6.39)	4.16 (2.98-5.82)
10-19	203	4.24 (3.05-5.91)	4.22 (3.02-5.90)
20-29	429	9.26 (6.75-12.71)	8.55 (6.20-11.77)
≥ 30	595	18.11 (13.25-24.74)	16.48 (12.02-22.60)
P-trend ^e		<0.001	<0.001
Cigarettes smoked per day in ever smokers ^d			
1-9	376	4.46 (3.57-5.58)	4.28 (3.42-5.37)
10-19	767	13.16 (10.65-16.28)	12.17 (9.81-15.10)
>20	125	23.07 (17.70-30.06)	21.66 (16.54-28.37)
P-trend ^e		<0.001	<0.001
Pack-years in ever smokers ^d			
1-5	199	4.01 (2.89-5.57)	3.81 (2.73-5.31)
6-15	429	8.40 (6.13-11.50)	7.90 (5.74-10.86)
>15	671	22.52 (16.53-30.69)	21.24 (15.52-29.06)
P-trend ^e		<0.001	<0.001
Age at smoking initiation in ever smokers ^d			
>20	334	6.33 (4.84-7.62)	5.76 (4.55-7.28)
17-20	739	9.95 (8.04-12.31)	9.08 (7.31-11.27)
≤ 17	315	6.07 (4.84-7.61)	5.70 (4.53-7.17)
P-trend ^e		<0.001	<0.001

^a Adjusted for age

^b Adjusted for age, duration of education and alcohol consumption, all at enrolment

^c Never-active, never-passive smokers as reference group

^d The sum of cases in each interval in smoking duration, cigarettes day, pack-years and age at smoking initiation does not sum up to the total number of cases in ever smokers because of missing values

^e Test for trend excluding never smokers

Table 3. Crude^a- and multivariate^b adjusted hazard ratio (HR) estimates for lung cancer with 95% confidence intervals (CIs) for former smokers according to years since quit smoking

	Cases	HR ^a (95% CI)	HR ^b (95% CI)
Current smokers		1.00 (ref)	1.00 (ref)
1-9 years	90	0.34 (0.28-0.42)	0.37 (0.30-0.45)
10-19 years	56	0.16 (0.12-0.21)	0.17 (0.13-0.23)
>20 years	46	0.11 (0.08-0.14)	0.11 (0.08-0.15)
P-trend		<0.001	<0.001

^a Adjusted for age

^b Adjusted for age, duration of education and alcohol consumption, all at enrolment

